Non-Crossing Quantile Regression with Shape Constraints

Haruki Kono *

MIT

October 23, 2025
Click here for the latest version.

Abstract

Quantile regression is widely used to study heterogeneous effects of covariates across the outcome distribution. However, standard estimators such as Koenker and Bassett's often violate fundamental shape restrictions implied by probability or economic theory, including the non-crossing property of conditional quantile functions, the monotonicity of production output with respect to inputs, and the monotonicity of equilibrium bidding strategies in structural auction models. Such violations can produce theoretically inconsistent estimates and undermine downstream analysis.

We develop a framework for shape-constrained quantile regression based on a variational formulation. This approach uses an infinite-dimensional linear program to characterize the quantile coefficients, enabling the imposition of restrictions—such as global non-crossing, derivative-based inequalities, and covariate monotonicity—across the entire continuum of quantiles. A computationally feasible estimator is obtained via finite-dimensional approximation, and its asymptotic properties are established.

Monte Carlo simulations show that the proposed estimator improves upon both classical and existing non-crossing approaches. In an application to U.S. timber auctions, it yields smooth, theory-consistent estimates of bid and valuation distributions and bidding strategies, in contrast to conventional methods that often violate basic economic restrictions.

*Email: hkono@mit.edu. I am deeply grateful to Alberto Abadie for his invaluable guidance, encouragement, and continuous support throughout this research. I have greatly benefited from discussions with Isaiah Andrews, Victor Chernozhukov, Anna Mikusheva, and Whitney Newey, whose insights have significantly improved the clarity and scope of this paper. I also thank Abhishek Ananth, Matias Cattaneo, Harold Chiang, Hiroaki Kaido, Andres Santos, and Stefan Wager for helpful conversations and comments. I appreciate the participants of seminars at MIT and University of Tokyo and Summer Workshop on Economic Theory for their thoughtful feedback and suggestions. I am grateful to Nathalie Gimenes and Emmanuel Guerre for sharing their code. This research was supported by the Jerry A. Hausman Dissertation Fellowship.

1 Introduction

Quantile regression provides a way to study how covariates affect the entire distribution of an outcome, not just its mean. This makes it a powerful tool for analyzing heterogeneous effects. In the case where conditional quantiles are assumed to be linear in covariates, Koenker and Bassett (1978) (KB) proposed an estimator that can be computed efficiently as a linear program. Their estimator has become a cornerstone of applied work in econometrics and statistics.

Despite its popularity, the KB estimator suffers from the quantile crossing problem. In theory, quantiles are ordered by construction: the 60th percentile should always exceed the 40th percentile. Yet in finite samples, the estimated conditional quantile curves from KB's method may violate this ordering. These crossings produce quantile functions that are not monotone, and thus inconsistent with any valid probability distribution. When this happens, the resulting estimates are hard to interpret and potentially misleading for distributional analysis.

The non-crossing property is just one example of a shape restriction. In many economic applications, theory dictates additional restrictions of this kind. For example, in production analysis, output should be increasing in inputs such as labor or capital, regardless of the level of production shocks. In first-price sealed-bid auction models with independent private values, it is well-known that the unique symmetric Bayesian-Nash equilibrium bidding strategy is increasing in bidders' valuations. These structural properties hold at the population level, but standard quantile regression estimates need not respect them. When they fail to do so—for example, producing a production function that decreases with labor—estimation results become difficult to interpret.

A large literature addresses these issues by incorporating shape restrictions into quantile regression estimation. For the quantile crossing problem, approaches include rearrangement of estimated quantiles (Dette and Volgushev (2008), Chernozhukov et al. (2010)), as well as constrained versions of the KB estimator that impose monotonicity across a grid of quantiles (Bondell et al. (2010)). For restrictions involving covariates, Koenker and Ng (2005) and Parker (2019) study linear quantile regression under affine constraints.

Still, enforcing non-crossing—a minimal requirement implied by probability theory—together with other structural restrictions remains challenging. Rearrangement methods can ensure non-crossing but may undo other imposed restrictions. Moreover, when the conditional quantile function is linearly specified, rearrangement destroys this structure, making the estimated coefficients difficult to interpret. Constrained estimators provide more flexibility but do not accommodate certain economically relevant conditions, such as the monotonicity of bidding strategies in auctions, which involves the derivative of the conditional quantile function with respect to the quantile index. These gaps motivate the search for a more general framework that can handle a broad class of shape restrictions within quantile regression.

To tackle this problem, we consider a variational characterization of quantile regression coefficients, which is originally introduced by Carlier et al. (2016), Carlier et al. (2017), and Carlier et al. (2020). Specifically, we formulate a linear program (LP) defined by the joint distribution of the observed variables. The LP is infinite-dimensional, involving functional variables, and it is known that a solution to the LP is given by an antiderivative of the quantile regression coefficients, viewed as a function of the quantile index.

Our first contribution is to show that this solution is (essentially) unique. Beyond its theoretical significance, this property yields a new identification of the quantile regression coefficients: they are given by the derivative of the LP solution. This framework identifies the coefficients jointly as a function of the quantile index, in sharp contrast to the classical approach that estimates them separately at each quantile by minimizing the check loss function.

It is natural in this framework to construct an estimator of the quantile regression coefficients by taking the derivative of the solution to the finite-sample counterpart of the LP. On its own, however, such an estimator does not necessarily respect economically or probabilistically motivated shape restrictions. These can be incorporated directly by adding suitable linear constraints to the LP. A key advantage of our formulation is that it treats the quantile regression coefficients as a function of the quantile index, rather than as separate quantities estimated point by point. This functional perspective makes it possible to impose restrictions that extend across the entire continuum of quantile levels. For example, the non-crossing condition can be enforced globally, not just on a finite grid as in Bondell et al. (2010) and Ando and Li (2025). By embedding this global restriction into the estimation procedure, we incorporate more information about the structure of the underlying distribution. As a result, our method is expected to deliver superior finite-sample performance compared to existing estimators. Numerical simulations in Section 5 confirm this expectation: our estimator recovers the quantile regression coefficients more accurately than both the original KB estimator and the non-crossing estimator of Bondell et al. (2010).

Likewise, because the coefficients are viewed as a function on the whole interval, it becomes straightforward to impose linear restrictions on their derivatives. Such restrictions are relevant to the auction example. The monotonicity of the equilibrium bidding strategy implies a functional inequality of the quantile regression coefficients and their first-order derivatives. If conditional quantiles are considered only on a grid, it is not easy to impose restrictions on derivatives, which highlights another advantage of using our estimator. It is also worth noting that rearrangement-based approaches cannot accommodate these derivative constraints.

In addition to restrictions across quantiles, our framework naturally allows for linear restrictions with respect to covariates. For example, in production analysis, monotonicity of the production function in input variables translates into monotonicity of conditional quantiles with respect to those covariates. By embedding covariate-based constraints directly into the LP, our method provides a unified way to enforce both cross-quantile and cross-covariate restrictions, thereby enhancing the interpretability and economic relevance of the resulting estimates.

A key challenge in computing our estimator is that the LP is infinite-dimensional and therefore cannot be solved directly. To overcome this, we approximate the problem by restricting attention to a finite-dimensional functional space that serves as a sieve for the set of quantile regression coefficient functions. Instead of minimizing the LP objective over the entire infinite-dimensional space, we minimize it over this approximating space, which reduces the problem to an LP with finitely many variables. While the resulting program still contains infinitely many constraints, it falls within the class of semi-infinite programming problems—LPs with finite-dimensional variables and infinite-dimensional constraints—for which well-developed solution methods exist (e.g., Shapiro (2009)). We implement the estimator using a discretization-based algorithm, as detailed in Section 3.

One of the main theoretical contributions of this paper is to establish the convergence of the finite-dimensional approximation scheme. We develop the duality theory for the infinite-dimensional LP that characterizes the quantile regression coefficients and show that the duality properties are preserved under approximation. This structure allows us to prove that the estimator obtained from the finite-dimensional LP is (locally) uniformly consistent for the true quantile regression coefficients. Furthermore, under the assumption that the approximation is exact, we establish the asymptotic normality of the estimator.

It is also possible to view our estimator as a sieve extremum estimator, for which large-sample properties are well established in the literature (Chen (2007)). In that framework, consistency typically requires conditions on the growth rate of the sieve dimension relative to the sample size. By contrast, our approach exploits the duality structure of the underlying LP, which allows us to establish uniform consistency without imposing such rate restrictions. In other words, the estimator remains consistent regardless of how quickly or slowly the approximation degree grows.

That said, the choice of approximation degree is not innocuous for convergence rates. An approximation space that is too small may introduce bias, while an excessively large one may lead to slower rates. We conjecture that there exists an optimal degree of approximation depending on the sample size, but characterizing this tradeoff—as well as deriving the corresponding asymptotic distribution in a fully nonparametric setting—is beyond the scope of this paper and left for future research.

In Section 6, we apply our estimator to a first-price sealed-bid auction model using data from timber auctions. While the literature on structural estimation of auction models is extensive, most existing methods ignore either the non-crossing condition of bid or valuation distributions, the monotonicity of the equilibrium bidding strategy, or both. As illustrated in Section 1.1 through numerical simulations, estimators that do not impose these shape restrictions frequently violate them in finite samples. By contrast, our estimator enforces these constraints by construction, ensuring that the estimated quantile regression coefficients are consistent with the economic model and facilitating reliable structural analysis.

Finally, we discuss relevant studies. A large body of work addresses the quantile crossing problem. For linear heteroscedastic models, He (1997) proposes a non-crossing estimator of quantile regression coefficients. For nonparametric conditional quantile functions, Dette and Volgushev (2008) and Chernozhukov et al. (2010) develop the quantile rearrangement technique. Bondell et al. (2010) consider minimizing a weighted sum of check loss functions subject to the non-crossing condition over a grid of quantiles. Building on this idea, Ando and Li (2025) derive a more convenient representation of the linear quantile model and propose a non-crossing estimator. Koenker and Ng (2005) and Parker (2019) consider linear quantile regression models with affine constraints, but their frameworks do not accommodate shape restrictions across the quantile index, such as non-crossing. Some studies, including Kitahara et al. (2021) and Dai et al. (2023), explore methods that simultaneously circumvent quantile crossing and impose other shape restrictions; however, they do not provide theoretical guarantees.

Our approach relies on the variational characterization of quantile regression coefficients developed in Carlier et al. (2016), Carlier et al. (2017), and Carlier et al. (2020). This framework formulates (vector) quantile regression from the perspective of optimal transport theory (see Villani (2003), Villani (2009)). While their focus is on the identification of quantile regression coefficients, we develop a practical estimation method and establish its theoretical guarantees.

The structural estimation of first-price sealed-bid auctions has also received extensive attention since Guerre et al. (2000). We adopt the linear quantile regression specification proposed by Gimenes and Guerre (2022), whereas Marmer and Shneyerov (2012) consider a nonparametric version of a similar model. Luo and Wan (2018), Pinkse and Schurter (2019), and Ma et al. (2021) explore constrained estimation of the valuation distribution using the quantile rearrangement. Our paper is closely related to Henderson et al. (2012), as both approaches enforce monotonicity of the bidding strategy. However, the methods differ fundamentally: Henderson et al. (2012) develop an estimator using constraint-weighted bootstrapping (Hall and Huang (2001)), whereas our approach imposes the restriction through a functional LP framework.

1.1 Motivating Example

We begin with a simple example to illustrate that unconstrained estimators can produce results inconsistent with economic theory in first-price auction models.

Suppose an econometrician observes data from first-price sealed-bid auctions $\ell = 1, \ldots, L$, each with $i = 1, \ldots, I$ risk-neutral bidders. Let Z_ℓ denote a one-dimensional auction-specific characteristic for simplicity, and $B_{\ell i}$ the bid of bidder i in auction ℓ. The econometrician observes $((Z_\ell, B_{\ell 1}, \ldots, B_{\ell I}))_{1 \le \ell \le L}$, assumed to be independent across auctions. Section 6 allows for multiple covariates.

We adopt the independent private value paradigm: each bidder draws a private value independently and submits a bid based solely on their own value. The valuation distribution $F_V(\cdot \mid z)$ conditional on auction characteristics z is assumed to be common knowledge.

Let $F_B(\cdot \mid z)$ and $f_B(\cdot \mid z)$ denote the conditional distribution and density of bids. It is well-known that a unique symmetric Bayesian-Nash equilibrium bidding strategy exists and is strictly increasing in private values (e.g., Athey and Haile (2007)). Specifically, the equilibrium strategy is given by the inverse of the map

$$\xi(b \mid z) = b + \frac{F_B(b \mid z)}{(I-1) f_B(b \mid z)},$$

which is strictly increasing in b at equilibrium. Note that $\xi(b \mid z)$ gives the private value of a bidder whose bid is b. It is not hard to check that the conditional quantile function of valuations, $Q_V(\cdot \mid z)$, satisfies

$$Q_V(u \mid z) = \xi(Q_B(u \mid z) \mid z) = Q_B(u \mid z) + \frac{u}{I-1}\left(D_u Q_B(u \mid z)\right), \tag{1}$$

where $Q_B(\cdot \mid z)$ is the conditional quantile function of bids and D_u denotes the derivative with respect to u.

Our goal is to estimate the distribution of private values. For this purpose, most papers take two steps: first, estimate the conditional distribution of bids B given Z; second, plug the resulting estimate $\hat{Q}_B$ into (1) to obtain

$$\hat{Q}_V(u \mid z) = \hat{Q}_B(u \mid z) + \frac{u}{I-1}\left(D_u \hat{Q}_B(u \mid z)\right).$$

Quantile regression models are often used in the first step. For example, Marmer and Shneyerov (2012) consider nonparametric quantile regression, while Gimenes (2017) and Gimenes and Guerre (2022) propose linear models. Following Gimenes and Guerre (2022), we adopt a linear quantile regression specification:

$$Q_B(u \mid z) = \beta_0(u) + z\beta_1(u). \tag{2}$$

This linear specification avoids the curse of dimensionality when multiple covariates are considered, as we do in Section 6, unlike nonparametric approaches.

By definition, the quantile functions $Q_B(\cdot \mid z)$ and $Q_V(\cdot \mid z)$ are increasing. Thus, their estimates, $\hat{Q}_B$ and $\hat{Q}_V$, should also satisfy the same monotonicity. Otherwise, not only would it be difficult to interpret them as valid probability distributions, but the estimator $\hat{\xi} = \hat{Q}_V \circ \hat{Q}_B^{-1}$ of the inverse bidding strategy could be non-monotone or even ill-defined. Such behavior would represent a deviation from the underlying economic model.

Nevertheless, unconstrained estimators frequently violate these conditions. To illustrate this, we apply the local polynomial estimator of Gimenes and Guerre (2022) to the model (2) under a simple data generating process (DGP):

$$\begin{cases} I = 2, L = 107 \\ Z_\ell \sim U[0,1] \quad \text{for } \ell = 1, \ldots, 107 \\ V_{\ell i} \mid Z_\ell \sim U\left[Z_\ell, Z_\ell + 2\right] \quad \text{for } i = 1, 2 \text{ and } \ell = 1, \ldots, 107 \end{cases}.$$

A reasonable interpretation of this DGP is that the valuation of bidder i at auction ℓ is the baseline value Z_ℓ plus an independent idiosyncratic shock drawn from $U[0,2]$. In this case, the equilibrium inverse bidding strategy is

$$\xi(b \mid z) = 2b - z \quad \text{if } z \le b \le z + 1,$$

and the bid distribution is

$$B_{\ell i} \mid Z_\ell \sim U[Z_\ell, Z_\ell + 1] \quad \text{for } i = 1, 2 \text{ and } \ell = 1, \ldots, 107.$$

Here, $L = 107$ mirrors the timber auction dataset analyzed in Section 6.

We generate datasets $t = 1, \dots 1000$ from the DGP and compute the estimates $\hat{Q}_B^{\mathrm{GG}}(\cdot \mid z)$ and $\hat{Q}_V^{\mathrm{GG}}(\cdot \mid z)$ of Gimenes and Guerre (2022)[1] in each iteration. For each dataset t, we record the number of auctions where the estimated conditional quantile functions are not increasing:

$$\begin{aligned} N_{B,t}^{\mathrm{GG}} &\coloneqq \#\{\ell = 1, \dots, L \mid \hat{Q}_B^{\mathrm{GG}}(\cdot \mid Z_\ell) \text{ is not increasing}\}, \\ \text{and} \\ N_{V,t}^{\mathrm{GG}} &\coloneqq \#\{\ell = 1, \dots, L \mid \hat{Q}_V^{\mathrm{GG}}(\cdot \mid Z_\ell) \text{ is not increasing}\}, \end{aligned} \tag{3}$$

for $t = 1, \dots, 1000$.

Figure 1 shows histograms of $N_{B,t}^{\mathrm{GG}}$ and $N_{V,t}^{\mathrm{GG}}$. Quantile crossings occur frequently, especially for the valuation estimates. Remarkably, in a substantial number of iterations, crossings appear in all auctions, highlighting a severe violation of monotonicity. For bids, crossings are less frequent but still occur in nearly ten percent of simulations, which remains a non-negligible concern.

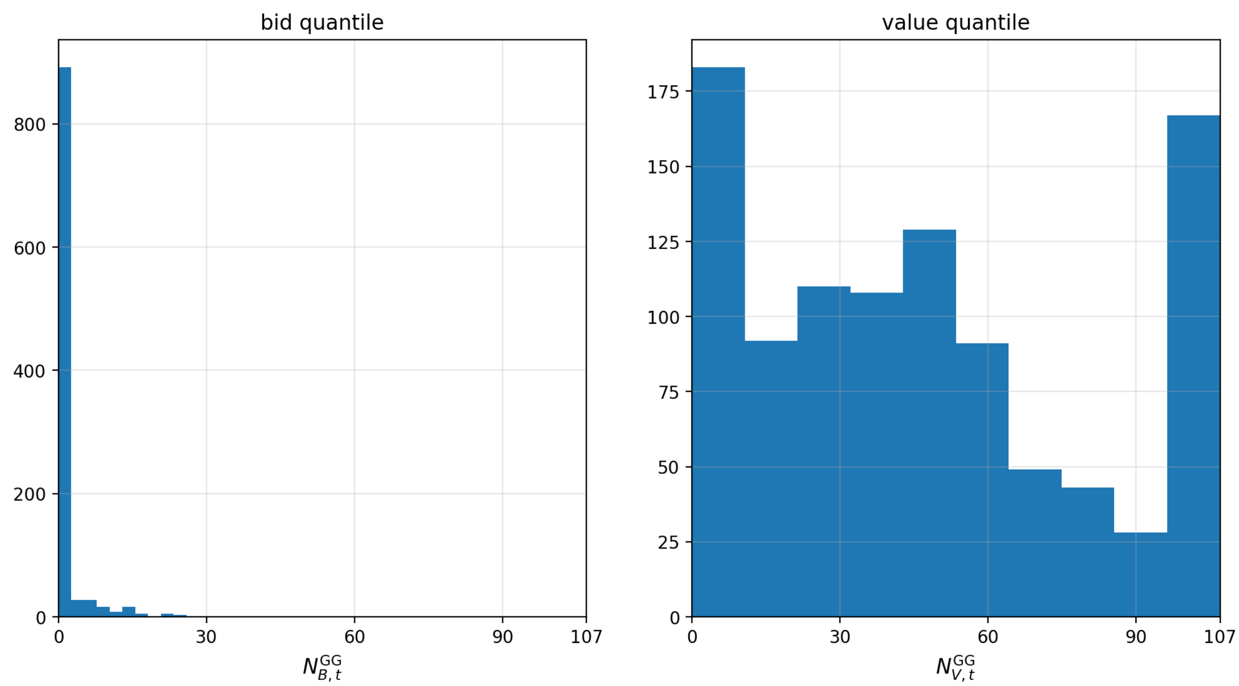

Figure 1: The histograms of the number of auctions with quantile crossings in the estimates of Gimenes and Guerre (2022). The left panel is the histogram of $N_{B,1}^{\mathrm{GG}}, \dots, N_{B,1000}^{\mathrm{GG}}$, the number of auctions with quantile crossings in the estimated bid distribution. The right panel is the histogram of $N_{V,1}^{\mathrm{GG}}, \dots, N_{V,1000}^{\mathrm{GG}}$, the number of auctions with quantile crossings in the estimated valuation distributions.

[1] We use the degree of the local polynomial $d = 2$ and the bandwidth $h = 0.3$ as Gimenes and Guerre (2022) suggest.

Our method can enforce the monotonicity directly within the estimation of the bid quantile regression. As a result, the estimated bid and valuation quantile functions, as well as the implied bidding strategy, are increasing for any auction-specific characteristics.

Before turning to formal results, Figure 2 provides a preview of how our estimator performs relative to Gimenes and Guerre's estimator in recovering the conditional quantile functions of bids and valuations at covariate values $z \in \{0.1, 0.5, 0.9\}$ for a representative simulated dataset. Both methods successfully recover the shape of the bid quantiles, demonstrating comparable performance in this dimension. The contrast is much clearer in the valuation quantile functions. While Gimenes and Guerre's estimator produces roughly increasing valuation quantile functions for lower covariate values, it exhibits severe non-monotonicity when $z = 0.9$. By construction, our estimator ensures that both the bid and valuation quantile functions are increasing for any z. Moreover, due to the imposed shape constraints, our estimator yields a closer approximation to the true conditional quantile functions.

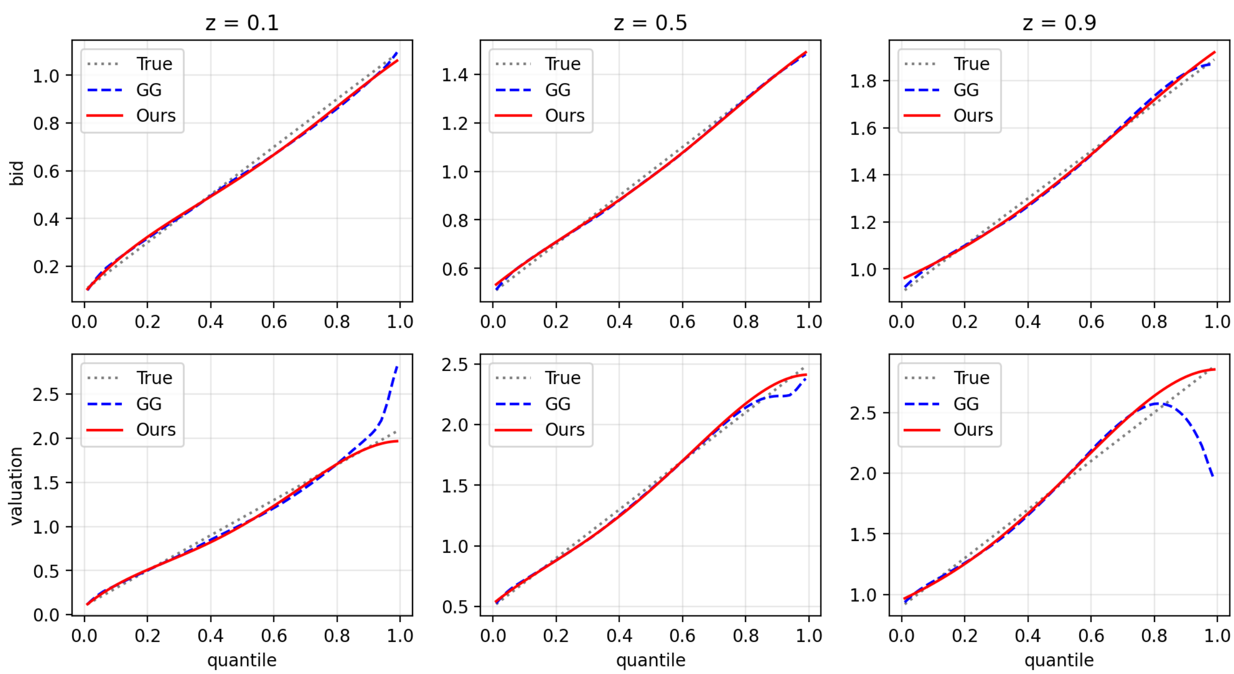

Figure 2: The conditional quantile functions of bids and valuations. The top panels show the truth (grey dotted), Gimenes and Guerre's estimate (blue dashed), and our estimate (red solid) of the conditional quantile function $Q_B(\cdot \mid z)$ of bids for $z \in \{0.1, 0.5, 0.9\}$. The bottom panels show the same for the conditional quantile function $Q_V(\cdot \mid z)$ of valuations.

Finally, it is worth mentioning that rearranging the unconstrained estimates of the conditional quantiles $Q_B(\cdot \mid z)$ and $Q_V(\cdot \mid z)$ is not a satisfactory solution. Such rearranged estimates lose the linear structure (2) and violate the model-based identity (1), leading to inconsistencies not only with the underlying economic model but also with the statistical model implied by the linear quantile regression specification.

2 Variational Characterization of Quantile Regression

2.1 The Model

We have i.i.d. observations (Y_i, Z_i) for $i = 1, \dots, n$, where Y_i is a scaler random variable and Z_i is a q-dimensional non-degenerated random vector. Let $X_i \coloneqq f(Z_i)$ be a p-dimensional vector of regressors where f is a prespecified transformation of Z_i. We assume that the first component of X_i is one. Let $\mathcal{Y} \subset \mathbb{R}$, $\mathcal{Z} \subset \mathbb{R}^q$, and $\mathcal{X} \subset \mathbb{R}^p$ be the support of Y, Z, and X, respectively. Throughout the paper, we assume $\mathbb{E}[Y_i^2] + \mathbb{E}[\|X_i\|^2] < \infty$.

For $u \in (0, 1)$, let $Q_{Y|X}(u \mid x)$ be the u-th conditional quantile of Y given $X = x$. It is well-known that the conditional quantile solves the following minimization problem

$$Q_{Y|X}(u \mid x) \in \operatorname*{argmin}_{q \in \mathbb{R}} \mathbb{E}\left[\rho_u(Y - q) \mid X = x\right]$$

where $\rho_u : \mathbb{R} \to \mathbb{R}$ is the check function defined as $\rho_u(\varepsilon) \coloneqq \varepsilon(u - \mathbf{1}\{\varepsilon < 0\})$.

The conditional quantile function is commonly assumed to be linear in x, i.e., $Q_{Y|X}(u \mid x) = x'\beta(u)$ for quantile regression coefficient $\beta : (0, 1) \to \mathbb{R}^p$. In this case, the function β is characterized as

$$\beta(u) \in \operatorname*{argmin}_{\beta \in \mathbb{R}^p} \mathbb{E}\left[\rho_u(Y - X'\beta)\right]. \tag{4}$$

However, the minimization problem makes sense for a generic distribution F_{YX} even when the linearity assumption does not hold. In what follows, we refer to the solution β to (4) as the quantile regression coefficient, regardless of whether linearity is satisfied.

In general, the map $u \mapsto x'\beta(u)$ is not equal to the conditional quantile function $Q_{Y|X}(\cdot \mid x)$. Nevertheless, Angrist et al. (2006) show that the map is the best linear approximation of the conditional quantile function under a weighted mean-square error. This result justifies regarding the map $u \mapsto x'\beta(u)$ as a surrogate object of the conditional quantile function without the linearity assumption.

The main object of this paper is the quantile regression coefficient β. Throughout the paper, we focus on cases where the set of minimizers is singleton for all $u \in (0, 1)$.

Assumption QR. For $u \in (0, 1)$, there exists a unique solution to (4).

Assumption QR is usually adopted in the literature on quantile regression under misspecification; see, e.g., Kim and White (2003) and Angrist et al. (2006). Sufficient conditions for Assumption QR are given by Kim and White (2003) and Carlier et al. (2016).

A fundamental feature of the conditional quantile is that $Q_{Y|X}(\cdot \mid x)$ is non-decreasing. This is a critical requirement for the conditional distribution of Y given $X = x$ to be well-defined. However, it is not automatically satisfied for the linear approximation, which makes the interpretation of the map $u \mapsto x'\beta(u)$ less trivial. To avoid this, we assume the following, which is a key assumption of this paper.

Assumption QL. The joint distribution F_{YX} satisfies the *quasi-linearity*, that is, $(0,1) \ni u \mapsto \beta(u)$ is continuous and $(0,1) \ni u \mapsto x'\beta(u)$ is strictly increasing for each $x \in \mathcal{X}$.

The term, quasi-linearity, is taken from Carlier et al. (2016). Ando and Li (2025) refer to this property as the approximate linearity. Assumption QL forces the linear approximation not to face the quantile crossing problem. Clearly, Assumption QL is satisfied if the conditional quantile function $Q_{Y|X}(u \mid x)$ is linear in x, i.e.,

$$Q_{Y|X}(u \mid x) = x'\beta(u), \tag{5}$$

and it is strictly increasing in u.

Assumption QL allows for misspecification in the sense that it also accommodates nonlinear conditional quantile functions as the following lemma shows.

Lemma 2.1. *Suppose that $\tilde{\beta} : (0,1) \to \mathbb{R}^p$ is a continuous function such that $v \mapsto x'\tilde{\beta}(v)$ is strictly increasing for all $x \in \mathcal{X}$. Let U be a continuous random variable drawn from a probability measure μ supported on an interval $\mathcal{U} \subset \mathbb{R}$. Let $g : \mathcal{U} \times [0,1] \to \mathbb{R}$ be a continuous function such that $u \mapsto g(u,w)$ is strictly increasing for all $w \in [0,1]$. Consider the data generating process*

$$Y = X'\tilde{\beta}(g(U,|Z|)), \quad (Z,U) \sim U[-1,1] \otimes \mu$$

where $X = (1,Z)'$. Then, the conditional quantile function $Q_{Y|X}(u \mid x) = x'\tilde{\beta}(g(u,|z|))$ of Y given $X = x$ is nonlinear in x in general. However, the quantile regression coefficient of the joint distribution of (Y,X) is given by $\beta = \tilde{\beta} \circ \tilde{F}^{-1}$, where $\tilde{F}^{-1}$ is the quantile function of $g(U,|Z|)$, and as a result, it satisfies Assumption QL.

More concretely, let us consider the DGP

$$Y = (2 - Z)(|Z| + 1)U, \quad (Z,U) \sim U[-1,1] \otimes U[0,1]. \tag{6}$$

Lemma 2.1 implies that Assumption QL is satisfied with $\beta(u) = Q_{U[1,2]\otimes U[0,1]}(u)(2,-1)'$, where $Q_{U[1,2]\otimes U[0,1]}(\cdot)$ is the quantile function of $U[1,2] \otimes U[0,1]$. Figure 3 shows a scatter plot of $n = 1000$ i.i.d. observations drawn from (6) as well as the conditional quantile curves

$z \mapsto Q_{Y|Z}(u|z)$ and the linear approximation $z \mapsto (1, z)\beta(u)$ for $u = 0.25, 0.5, 0.75$. We observe that the conditional quantile curves, drawn in dashed red, are highly nonlinear. The linear approximation, drawn in solid blue, does not encounter the quantile crossing problem, as Lemma 2.1 predicts.

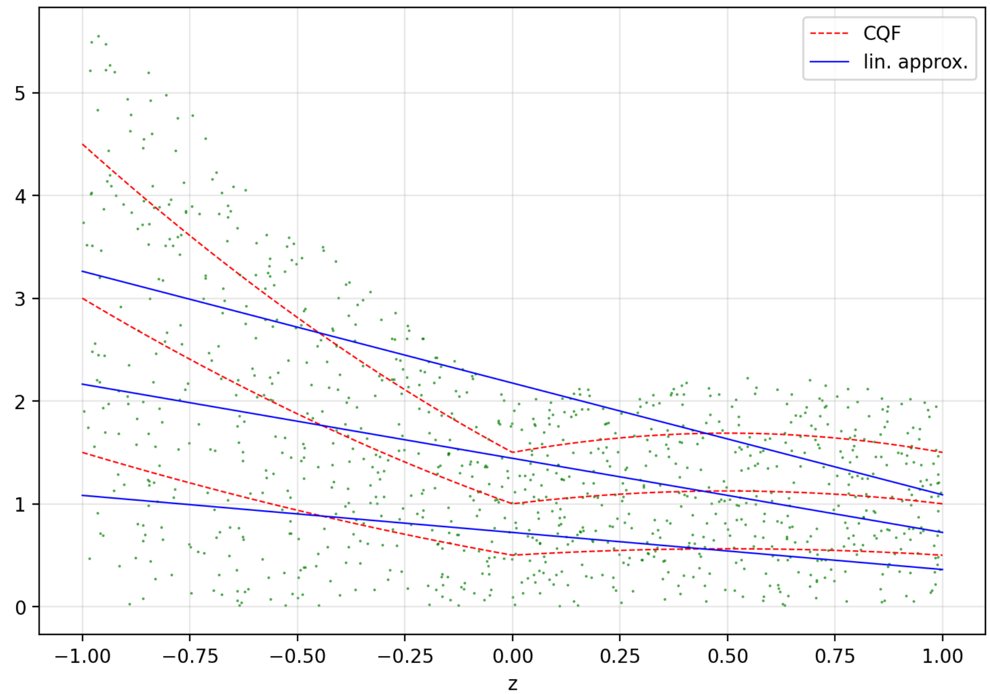

Figure 3: The red dashed lines represent the conditional quantile curves of the data generating process (6) for $u = 0.25, 0.5, 0.75$. The blue solid lines represent the linear approximation. The green dots represent the observed data points.

Heuristically, the quasi-linearity requires that the covariate transformation f is sufficiently rich. For example if it includes higher-order polynomials, then the true conditional quantile function $Q_{Y|X}(\cdot \mid x)$ can be well approximated by $x'\beta(\cdot)$ for some β. Since $Q_{Y|X}(\cdot \mid x)$ is increasing, it is expected that its linear approximation $x'\beta(\cdot)$ to be approximately increasing as well. A violation of Assumption QL may therefore indicate that the transformation f is not rich enough to capture the essential features of the true conditional quantile function.

We end this subsection with the probabilistic representation of the quasi-linearity. First, recall that if the conditional quantile function is linear in x, then the joint distribution of (Y, X) admits a probabilistic representation

$$Y = X'\beta(U), \quad U \mid X \sim U[0, 1]$$

for some uniform random variable U. Assumption QL has a similar representation. Before showing it, we need to introduce a regularity condition that is largely maintained through the rest of the paper.

Assumption LS. The conditional distribution $Y \mid X = x$ admits a continuous density $f_{Y|X}(\cdot \mid x)$ such that the set $\mathcal{X}_0 := \{x \in \mathcal{X} \mid f_{Y|X}(x'\beta(u) \mid x) > 0 \text{ for all } u \in (0,1)\}$ satisfies $F_X(\mathcal{X}_0) > 0$.

Assumption LS is a relaxation of the following condition. This is commonly imposed in the literature including Koenker and Bassett (1978) and Kim and White (2003).

Assumption LS'. The conditional distribution $Y \mid X = x$ admits a continuous density $f_{Y|X}(\cdot \mid x)$ such that $f_{Y|X}(x'\beta(u) \mid x) > 0$ for all $u \in (0,1)$ and F_X-almost all x.

Assumption LS' is more demanding than Assumption LS since it requires $F_X(\mathcal{X}_0) = 1$. The difference between these two assumptions can be crucial. For example, the DGP considered in (6) below satisfies Assumption LS but not Assumption LS'.

Now, we are ready to obtain the probabilistic representation of the quasi-linearity. The following is a slight generalization of Carlier et al. (2016).

Lemma 2.2. *Under Assumptions QR, LS, if Assumption QL holds, the random variable* $\bar{U} := (X'\beta(\cdot))^{-1}(Y)$ *satisfies*

$$\bar{U} \sim U[0,1] \quad \textit{and} \quad \mathbb{E}[X \mid \bar{U}] = \mathbb{E}[X].$$

Conversely, if there exist a continuous function $\bar{\beta} : (0,1) \to \mathbb{R}^p$ *and a random variable* $\bar{U} \sim U[0,1]$ *such that* $\bar{u} \mapsto x'\bar{\beta}(\bar{u})$ *is strictly increasing for all* $x \in \mathcal{X}$,

$$Y = X'\bar{\beta}(\bar{U}), \quad \textit{and} \quad \mathbb{E}[X \mid \bar{U}] = \mathbb{E}[X],$$

then $\beta = \bar{\beta}$ *holds, and as a result, Assumption QL is satisfied.*

Lemma 2.2 shows that Assumption QL requires X to be mean-independent of $\bar{U} = (X'\beta(\cdot))^{-1}(Y)$, that is, $\mathbb{E}[X \mid \bar{U}] = \mathbb{E}[X]$. In contrast, the linearity assumption (5) imposes the stronger condition $U \perp\!\!\!\perp X$. Since independence implies mean independence, this relationship confirms that linearity is a sufficient condition for quasi-linearity.

2.2 Identification under Quasi-Linearity

We first introduce an assumption on the covariate distribution

Assumption FR. $\mathbb{E}[XX' \mid X \in \mathcal{X}_0]$ exists and is full-rank.

Assumption FR seems not standard in the literature, but under Assumption LS', it reduces to the following condition, which is common in regression analysis.

Assumption FR'. $\mathbb{E}[XX']$ exists and is full-rank.

Although the full-rankness of $\mathbb{E}[XX']$ is weaker than that of $\mathbb{E}[XX' \mid X \in \mathcal{X}_0]$, Assumption FR is fairly weak. It is satisfied, for example, if $\mathcal{X}_0$ is not contained in any affine space of dimension $p-2$.

In many propositions throughout this paper, Assumptions LS and FR are imposed simultaneously. These propositions remain valid if Assumptions LS' and FR' are used instead. The two sets of assumptions involve a trade-off: Assumption LS is weaker than Assumption LS', while Assumption FR is stronger than Assumption FR'. We choose to work with the first pair because the practical benefit of Assumption FR' over Assumption FR appears limited. On the other hand, Assumption LS' does not accommodate our important example (6) discussed in the previous subsection, which is covered by Assumption LS.

Under the quasi-linearity, we have a variational characterization of the quantile regression coefficient within the framework developed by Carlier et al. (2016).

Lemma 2.3. *Suppose that Assumptions QR, QL, LS hold. The infinite-dimensional linear program (LP)*

$$\begin{aligned} &\inf_{\substack{\psi \in C(\mathcal{X}\times\mathcal{Y}) \\ \sigma \in (C([0,1]) \cap L^1_0[0,1])^p}} \int \psi \, dF_{YX} \\ &\text{s.t.} \quad uy \le \psi(x,y) + x'\sigma(u) \text{ for } (u,x,y) \in [0,1]\times\mathcal{X}\times\mathcal{Y} \end{aligned} \tag{7}$$

admits a solution $(\psi_\beta, \sigma_\beta)$ defined as

$$\sigma_\beta(u) \coloneqq \int_0^u \beta(v)dv - \int_0^1 \left(\int_0^{\tilde{u}} \beta(v)dv \right) d\tilde{u} \quad \text{and} \quad \psi_\beta(x,y) \coloneqq \sup_{u\in[0,1]} \left(uy - x'\sigma_\beta(u) \right).$$

Moreover, under Assumption FR, $(\psi,\sigma) = (\psi_\beta,\sigma_\beta)$ is the unique solution to (7) such that $u \mapsto x'\sigma(u)$ is convex for each $x \in \mathcal{X}$.

The first part of Lemma 2.3 guarantees the existence of a solution to (7). This follows immediately from the results obtained by Carlier et al. (2016). See the proof in Appendix C.3 for details.

To the best of our knowledge, the uniqueness part of Lemma 2.3 is new. This result relates the quantile regression coefficient β to the solution of the LP (7). To understand its usefulness, observe that the LP is determined solely by the joint distribution F_{YX}, which is explicitly identified. Given the joint distribution, suppose that we could find a solution $\bar{\sigma} \in (C([0,1]) \cap L_0^1[0,1])^p$ such that $u \mapsto x'\bar{\sigma}(u)$ is convex for each $x \in \mathcal{X}$. The uniqueness implies that the solution necessarily satisfies $\bar{\sigma} = \sigma_\beta$, and as a result, $\beta(\cdot) = D\bar{\sigma}(\cdot)$. This procedure provides a way to recover the quantile regression coefficient β from the joint distribution F_{YX} of the observable variables. It should be compared with the standard definition of the quantile regression coefficient, where the joint distribution F_{YX} defines a minimization problem (4), and the solution to the problem is β.

Given the above discussion, we can see that Lemma 2.3, particularly its uniqueness result, opens a new path from F_{YX} to β, distinct from the standard identification based on check function minimization. See Figure 4 for a diagram that illustrates this point.

$$\begin{tikzcd}
 & \min_\beta \mathbb{E}[\rho_u(Y - X'\beta)] \arrow[drr, "\exists!\ \text{sol.}"] & & \\
F_{YX} \arrow[ur, "\text{def.}"] \arrow[dr, "\text{def.}"'] & & & \beta(\cdot) \\
 & \inf_{\psi,\sigma} \int \psi dF_{YX} \arrow[r, "\exists!\ \text{sol.}"'] & \sigma(\cdot) \arrow[ur, "D"'] &
\end{tikzcd}$$

Figure 4: The above path: The joint distribution F_{YX} defines the minimization problem (4) using the check loss function. The quantile regression coefficient is defined as the solution to this problem. The below path: The joint distribution F_{YX} defines the infinite-dimensional LP (7). The quantile regression coefficient β is the derivative of the unique solution σ such that $u \mapsto x'\sigma(u)$ is convex.

A remarkable feature of the characterization given in Lemma 2.3 is that a single problem (7) pins down the quantile regression coefficients at all possible quantile levels $u \in (0,1)$ simultaneously. This is an aspect distinct from the check function based approach, where the problem (4) determines only the u-th coefficient $\beta(u)$. Due to this feature, additional constraints across different u's, including the non-crossing condition and the monotonicity of the equilibrium bidding strategy in auction models, can be easily imposed on the estimators we propose in the next section.

3 Shape-Constrained Estimation

3.1 Class of Shape Restrictions

Mathematical and economic theory often imposes restrictions on the admissible forms of conditional quantile functions. The non-crossing condition is one such example. In this subsection, we discuss a broad class of shape restrictions are commonly of interest in practice.

Recall that the covariate X is defined as the transformation of Z by f. Here, we are interested in shape restrictions on the linear approximation $(u,z) \mapsto f(z)'\beta(u)$. We focus on cases where the quantile regression coefficients $\beta(\cdot)$ takes values on $(C^k[0,1])^p =: \mathbb{B}^{(k)}$ for some $k \geq 1$, where $C^k[0,1]$ is the space of k-times continuously differentiable functions on $(0,1)$ whose derivatives are continuously extended to $[0,1]$. The order k is chosen depending on the shape constraints of interest; see the examples below. The space $\mathbb{B}^{(k)}$ equipped with the supremum norm $\|\beta\|_{\mathbb{B}^{(k)}} := \sum_{\ell=0}^{k} \sup_{u\in[0,1]} \left\|D_u^\ell \beta(u)\right\|$ is a Banach space.

Let $(\mathbb{D}, \|\cdot\|_{\mathbb{D}})$ be another Banach space. For a continuous linear operator $\mathcal{A}_f : \mathbb{B}^{(k)} \to \mathbb{D}$, a closed convex cone $K \subset \mathbb{D}$, and $\delta \in \mathbb{D}$, we consider restrictions in the form of

$$\mathcal{A}_f \beta + \delta \in K. \tag{8}$$

Call this form of shape restriction a conic restriction. Although this class of shape restrictions is abstract, it contains a variety of interesting examples as follows.

Example 1 (Non-crossing)**.** Let $k \geq 1$, and let $\mathbb{D} = C([0,1] \times \mathcal{X})$ be equipped with the supremum norm. Define $(\mathcal{A}_f^{\mathrm{NC}} \beta)(u,x) := x' D_u \beta(u)$, $\delta^{\mathrm{NC}} := 0$, and $K^{\mathrm{NC}} := \{k \in \mathbb{D} \mid k(u,x) \geq 0 \text{ for all } (u,x) \in [0,1] \times \mathcal{X}\}$. It is clear that $\mathcal{A}_f^{\mathrm{NC}}$ is linear and continuous and that K^{NC} is a closed and convex cone. The non-crossing condition is equivalent to β satisfying $\mathcal{A}_f^{\mathrm{NC}} \beta + \delta^{\mathrm{NC}} \in K^{\mathrm{NC}}$. Moreover, if the support of X_{-1}, where $X = (1, X_{-1}')'$, is a polytope whose vertices are denoted by $\bar{x}_1, \ldots, \bar{x}_v$, then the convex cone can be reduced to $K = \{k \in \mathbb{D} \mid k(u,x) \geq 0 \text{ for all } (u,x) \in [0,1] \times \{\bar{x}_1, \ldots, \bar{x}_V\}\}$, as observed by Bondell et al. (2010).

Example 2 (Range of outcome variable Y)**.** Suppose that we know that Y takes values greater than or equal to $\underline{y} \in \mathbb{R}$. Let $k \geq 1$, and let $\mathbb{D} = C([0,1] \times \mathcal{X})$ be equipped with the supremum norm. Define $(\mathcal{A}_f^{\mathrm{LB}} \beta)(u,x) := x'\beta(u)$, $\delta^{\mathrm{LB}} := -\underline{y}$, and $K^{\mathrm{LB}} := \{k \in \mathbb{D} \mid k(u,x) \geq 0 \text{ for all } (u,x) \in [0,1] \times \mathcal{X}\}$. Then, the restriction on the lower bound is represented as

$$x'\beta(u) \geq \underline{y} \text{ for } (u,x) \in [0,1] \times \mathcal{X} \iff \mathcal{A}_f^{\mathrm{LB}} \beta + \delta^{\mathrm{LB}} \in K^{\mathrm{LB}}.$$

Example 3 (Monotonicity in z)**.** For simplicity, we focus on univariate Z. The extension to other cases is straightforward. Assume that the map $z \mapsto f(z)'\beta(u)$ is non-decreasing and f is assumed to be smooth. Let $k \geq 1$, and let $\mathbb{D} = C([0,1] \times \mathcal{Z})$ be equipped with the supremum norm. Define $(\mathcal{A}_f^{\uparrow} \beta)(u,z) := (D_z f(z))'\beta(u)$, $\delta^{\uparrow} := 0$, and $K^{\uparrow} := \{k \in \mathbb{D} \mid k(u,z) \geq 0 \text{ for all } (u,z) \in [0,1] \times \mathcal{Z}\}$. This gives a representation of the non-decreasingness in z:

$$D_z(f(z)'\beta(u)) \geq \delta^{\uparrow} \text{ for } (u,z) \in [0,1] \times \mathcal{Z} \iff \mathcal{A}_f^{\uparrow} \beta + 0 \in K^{\uparrow}.$$

Example 4 (Convexity in z)**.** Consider the convexity of the map $z \mapsto f(z)'\beta(u)$, assuming the smoothness of f. Let $k \geq 1$, and let $\mathbb{D}^{\mathrm{CVX}} = C([0,1] \times \mathcal{Z} \times S^{q-1})$ be equipped with the supremum norm, where $S^{q-1} \subset \mathbb{R}^q$ is the unit $(q-1)$-surface of radius one. Define $\mathcal{A}_f^{\mathrm{CVX}} : \mathbb{B}^{(k)} \to \mathbb{D}^{\mathrm{CVX}}$ as

$$(\mathcal{A}_f^{\mathrm{CVX}}\beta)(u,z,v) \coloneqq \sum_{i=0}^{p-1} \left(v' D_z^2 f_i(z) v\right) \beta_i(u),$$

where $f = (f_0, \ldots, f_{p-1})'$ and $\beta = (\beta_0, \ldots, \beta_{p-1})'$, and let $\delta^{\mathrm{CVX}} \coloneqq 0$ and $K^{\mathrm{CVX}} \coloneqq \{k \in \mathbb{D}^{\mathrm{CVX}} \mid k(u,z,v) \geq 0 \text{ for all } (u,z,v) \in [0,1] \times \mathcal{Z} \times S^{q-1}\}$. Now, we have the representation of the convexity restriction as follows:

$$D_z^2(f(z)'\beta(u)) \text{ is positive semi-definite for } (u,z) \in [0,1] \times \mathcal{Z} \iff \mathcal{A}_f^{\mathrm{CVX}}\beta + \delta^{\mathrm{CVX}} \in K^{\mathrm{CVX}}.$$

Example 5 (Monotonicity of equilibrium bidding strategy)**.** In our example of auction estimation, we impose the monotonicity of the valuation quantile function (1). Note that since $Q_V = \xi \circ Q_B$, where ξ is the equilibrium inverse bidding strategy, the monotonicity of Q_V is equivalent to that of ξ. Our framework covers this constraint. For $k \geq 2$, define $\mathcal{A}_f^{\uparrow\xi} : \mathbb{B}^{(k)} \to \mathbb{D}^{\uparrow\xi} \coloneqq C([0,1] \times \mathcal{X})$ as

$$(\mathcal{A}_f^{\uparrow\xi}\beta)(u,x) \coloneqq x' D_u \left(\beta(u) + \frac{u}{I-1} D_u\beta(u)\right) = x' \left(\frac{I}{I-1} D_u\beta(u) + \frac{u}{I-1} D_u^2\beta(u)\right).$$

Then, the monotonicity of the bidding strategy has a conic representation $\mathcal{A}_f^{\uparrow\xi}\beta + \delta^{\uparrow\xi} \in K^{\uparrow\xi}$ where $\delta^{\uparrow\xi} \coloneqq 0$ and $K^{\uparrow\xi} \coloneqq \{k \in \mathbb{D}^{\uparrow\xi} \mid k(u,x) \geq 0 \text{ for all } (u,x) \in [0,1] \times \mathcal{X}\}$.

Example 6 (Multiple restrictions)**.** Since the non-crossing condition is a mathematical requirement, it is always of interest once Assumption QL is imposed. If an additional shape restriction is implied by economic theory, multiple restrictions must be considered simultaneously. Our framework is sufficiently general to incorporate such joint restrictions. Suppose that multiple restrictions $\mathcal{A}_f^{(\ell)}\beta + \delta^{(\ell)} \in K^{(\ell)}$ for $\ell = 1, \ldots, L$ are of interest, where $\mathcal{A}_f^{(\ell)} : \mathbb{B}^{(k)} \to \mathbb{D}^{(\ell)}$ is continuous and linear and $\mathbb{D}^{(\ell)}$ is Banach. This situation is reduced to the current framework by setting $\mathcal{A}_f \coloneqq (\mathcal{A}_f^{(1)}, \ldots, \mathcal{A}_f^{(L)})$, $\delta \coloneqq (\delta^{(1)}, \ldots, \delta^{(L)})$, $K \coloneqq K^{(1)} \times \cdots \times K^{(L)}$, and $\mathbb{D} \coloneqq \mathbb{D}^{(1)} \times \cdots \times \mathbb{D}^{(L)}$.

So far, we have discussed various shape restrictions on (the linear approximation of) conditional quantile functions. Before turning to the estimation of β, we first consider when and why economists might care directly about these restrictions.

By the inverse probability integral transform, there exists a random variable U independent of Z such that $Y = Q_{Y|Z}(U|Z)$ holds almost surely. This representation is purely a reduced form and does not necessarily reflect the underlying structural relationship between Y and Z. Under the exogeneity assumption, the structural form is typically written as $Y = g(Z, \varepsilon)$, where ε is a potentially infinite-dimensional unobserved heterogeneity and is independent of Z. In many economic applications, the object of real interest is the shape of the function $z \mapsto g(z, \varepsilon)$ rather than the shape of the reduced-form mapping $z \mapsto Q_{Y|Z}(u \mid z)$.

Does it then make sense to impose the same shape constraints on conditional quantiles as on the structural function? The answer depends on the nature of the constraint: some features of the structural function $z \mapsto g(z, \varepsilon)$ are inherited by the reduced-form function $z \mapsto Q_{Y|Z}(u \mid z)$, while others are not.

The most straightforward example of the former case is a bound on the support of g, as illustrated in Example 2. if $g(z, \varepsilon) \geq \underline{y}$ for all (z, ε), then it follows that $Q_{Y|Z}(u \mid z) \geq \underline{y}$ for all (z, u).

More interestingly, linear restrictions on the first-order derivatives of g in z are also transmitted to the reduced form.

Lemma 3.1. *Assume that $z \mapsto g(z, \varepsilon)$ is smooth. Let $a : \mathcal{Z} \to \mathbb{R}^q$ and $b : \mathcal{Z} \to \mathbb{R}$ be measurable functions. Then, if $(D_z g(z, \varepsilon))a(z) \geq b(z)$ holds for all (ε, z), it follows that $(D_z Q_{Y|Z}(u \mid z))a(z) \geq b(z)$ holds for all (u, z).*

Lemma 3.1 provides a strong justification for imposing the same shape restriction on the conditional quantile function as on the structural function. In particular, the monotonicity of $z \mapsto Q_{Y|Z}(u \mid z)$, as in Example 3, can be of great interest when the structural function is believed to be monotone.

However, shape restrictions on higher-order derivatives of g are not generally inherited by the reduced form without additional assumptions. For example, consider the convexity constraint discussed in Example 4. The reduced form $z \mapsto Q_{Y|Z}(u \mid z)$ may fail to be convex even if $z \mapsto g(z, \varepsilon)$ is convex. This reflects the fact that the lower envelope of convex functions need not be convex. A concrete example is given in the following lemma.

Lemma 3.2. *Consider the DGP*

$$Y = g(Z, \varepsilon) := Z^2 + (1 - 4\varepsilon)Z + 2\varepsilon^2, \quad (Z, \varepsilon) \sim U[0, 1] \otimes U[0, 1].$$

The function $z \mapsto g(z, \varepsilon)$ is convex for all $\varepsilon \in [0, 1]$, but the conditional quantile function of Y given $Z = z$ is not convex in z at some quantile level.

Imposing shape constraints on higher-order derivatives is justified when ε is univariate and g is monotone in ε, in which case $Q_{Y|Z}(u \mid z) = g(z, h(u))$ holds for some monotone function h. Some papers implicitly make this assumption and develop estimation methods for quantile regression with covariate convexity (see, e.g., Koenker and Ng (2005), Wang et al. (2014), Kitahara et al. (2021)). However, when quantile regression is used in structural models, whether such constraints are justified should be carefully examined.

3.2 The Estimator

We aim to construct an estimator of the quantile regression coefficient $\beta(\cdot)$ that respects a shape restriction $A_f\beta + \delta \in K$. Our construction is based on Lemma 2.3. The core idea is to solve the empirical counterpart of the LP (7)

$$\begin{aligned} &\inf_{\substack{\psi \in C(\mathcal{X}\times\mathcal{Y}) \\ \sigma \in (C([0,1]) \cap L_0^1[0,1])^p}} \frac{1}{n}\sum_{i=1}^{n} \psi(X_i, Y_i) \\ &\text{s.t.} \quad uy \leq \psi(x,y) + x'\sigma(u) \text{ for } (u,x,y) \in [0,1]\times\mathcal{X}\times\mathcal{Y} \end{aligned} \tag{9}$$

and recover β as the derivative of its solution. However, this construction faces two major challenges. First, the optimization problem is computationally difficult to solve due to the infinite-dimensional nature of the variables ψ and σ. While ψ can be effectively reduced to a finite-dimensional object—since it only enters the objective function through the values $\psi(X_1, Y_1), \ldots, \psi(X_n, Y_n)$—this reduction does not resolve the main issue, as σ remains an infinite-dimensional function

Moreover, even if the true quantile regression coefficient β satisfies some shape restriction, the plug-in estimator $\hat{\beta} = D\hat{\sigma}$ is not guaranteed to do so. Such violations of shape constraints can hinder interpretability and undermine the economic validity of the estimation results.

We address these issues. First, in order to overcome the computational difficulty, we introduce a finite-dimensional sieve scheme that approximates functions in the space $C[0,1] \cap L_0^1[0,1]$. For an integer $J \geq 1$, define $m^J : [0,1] \to \mathbb{R}^J$ as

$$m^J(\cdot) \coloneqq (m_1(\cdot), \ldots, m_J(\cdot))' \quad \text{where} \quad m_j(u) \coloneqq \frac{1}{j}\left(u^j - \frac{1}{j+1}\right).$$

It is easy to see that $m_j \in C[0,1] \cap L_0^1[0,1]$. By the Weierstrass approximation theorem, for any $\sigma_k \in C[0,1] \cap L_0^1[0,1]$, there exists a vector $\lambda_k \in \mathbb{R}^J$ such that $\lambda_k' m^J(\cdot)$ approximates $\sigma_k(\cdot)$ in the supremum norm for large J. Hence, it is expected that optimizing the objective function of (9) over $\sigma \in (C[0,1] \cap L_0^1[0,1])^p$ can be approximately replaced by the optimization over $\Lambda = (\lambda_1, \ldots, \lambda_p)' \in \mathbb{R}^{p \times J}$ in view of $\sigma(\cdot) \approx \Lambda m^J(\cdot)$. The same approximation scheme is introduced by Spady and Stouli (2018) in a similar context, but we develop a richer duality theory and convergence analysis; see Appendix B. The same argument below extends to any independent functional basis whose linear combinations form a dense subspace of $C[0,1] \cap L_0^1[0,1]$.

We next consider imposing the shape restriction $\mathcal{A}_f\beta + \delta \in K$ on the estimate. Since the quantile regression coefficient β is the derivative of the solution σ to the LP (7), as shown in Lemma 2.3, we add the constraint $\mathcal{A}_f(D_u\sigma(\cdot)) + \delta \in K$ to the LP. To impose the shape restriction on the polynomial approximation, we optimize over the set of Λ's that satisfy $\mathcal{A}_f(\Lambda D_u m^J(\cdot)) + \delta \in K$, since $D_u\sigma(\cdot) \approx D_u(\Lambda m^J(\cdot)) = \Lambda D_u m^J(\cdot)$.

Now, we are ready to define our estimator. Let $J = J_n$ be the degree of the polynomial approximation basis that can vary with the sample size n. Consider the following problem

$$\inf_{\substack{\psi \in \mathbb{R}^n \\ \Lambda \in \mathbb{R}^{p \times J_n}}} \frac{1}{n} \sum_{i=1}^{n} \psi_i$$
$$\text{s.t.} \quad \begin{cases} uY_i \le \psi_i + X_i' \Lambda m^{J_n}(u) \text{ for } (u,i) \in [0,1] \times \{1, \ldots, n\} \\ \mathcal{A}_f(\Lambda D_u m^{J_n}(\cdot)) + \delta \in K \end{cases} . \tag{10}$$

This minimization problem falls into the class of semi-infinite LP (Shapiro (2009))—it involves finite-dimensional variables and infinite-dimensional constraints.

We first establish the existence of a solution to the LP (10) in the following lemma.

Lemma 3.3. *Suppose that Assumption FR holds and that the LP (10) is feasible. Then, it admits a solution with probability one.*

Let $\hat{\Lambda}_n \in \mathbb{R}^{p \times J_n}$ be a solution to the LP (10). We propose an estimator

$$\hat{\beta}_n(\cdot) := \hat{\Lambda}_n D_u m^{J_n}(\cdot). \tag{11}$$

Note that the solution $\hat{\Lambda}_n$ to (10) is not unique in general, and consequently, $\hat{\beta}_n$ is not uniquely defined. However, regardless of which solution is selected, the resulting estimator satisfies the desired asymptotic properties. Details are provided in Section 4.

3.3 Implementation

The first constraint of the LP (10) consists of several functional inequalities. Typically, the shape restriction $\mathcal{A}_f(\Lambda D_u m^{J_n}(\cdot)) + \delta \in K$ also involve some inequalities. For example, the non-crossing condition $\mathcal{A}_f^{\mathrm{NC}}(\Lambda D_u m^{J_n}(\cdot)) \in K^{\mathrm{NC}}$ is equivalent to

$$x'\Lambda D_u^2 m^{J_n}(u) \geq 0 \text{ for all } (u,x) \in [0,1] \times \mathcal{X}.$$

Similarly, the monotonicity $\mathcal{A}_f^{\uparrow}(\Lambda D_u m^{J_n}(\cdot)) \in K^{\uparrow}$ can be written as

$$(D_z f(z))'\Lambda D_u m^{J_n}(u) \geq 0 \text{ for all } (u,z) \in [0,1] \times \mathcal{Z}$$

A variety of numerical methods have been developed to solve such problems with convergence guarantees; see, e.g., López and Still (2007). Below, we briefly explain a discretization-based algorithm and outline how it works, following the approach in López and Still (2007). In Section 5, we implement our estimator using this algorithm.

Write the inequality constraints in (10) as

$$g(\psi,\Lambda; i,u,x) \geq 0 \quad \text{for} \quad (i,u,x) \in \{1,\dots,n\} \times [0,1] \times \mathcal{X} =: T.$$

We consider a grid $T_k \coloneqq \{1,\dots,n\} \times \left\{0, \frac{1}{2^k}, \dots, \frac{2^k-1}{2^k}, 1\right\} \times \mathcal{X}_k$, where $\mathcal{X}_k$ is a growing fine subset of $\mathcal{X}$. Given this grid and a small tolerance level $\varepsilon > 0$, we run the following algorithm.

1. Let $k = 1$.

2. Solve the LP (10) subject to the constraints over T_k instead of T. Let $(\psi^{(k)}, \Lambda^{(k)})$ be a solution to the problem.

3. Stop the algorithm if the inequalities for $(\psi^{(k)}, \Lambda^{(k)})$ approximately hold on T, i.e.,

$$g(\psi^{(k)}, \Lambda^{(k)}; i,u,x) > -\varepsilon \quad \text{for} \quad (i,u,x) \in \{1,\dots,n\} \times [0,1] \times \mathcal{X}.$$

   Otherwise, repeat Step 2 with $k+1$.

By Theorem 10 of López and Still (2007), it is guaranteed that the output of this algorithm successfully approximates the real solution of (10) if the grid $\mathcal{X}_k$ is properly constructed.

3.4 Comparison with Other Non-Crossing Estimators

Solving the LP (10) with the restriction $(\mathcal{A}_f^{\mathrm{NC}}, \delta^{\mathrm{NC}}, K^{\mathrm{NC}})$ is not the only way to impose the non-crossing condition on the estimator. Bondell et al. (2010) propose what may be a more intuitive approach. They estimate several quantile coefficients jointly by solving

$$(\hat{\beta}^{\mathrm{BRW}}(u_\ell))_{\ell=1,\dots,L} \in \begin{array}{c} \underset{(\beta_\ell)_{\ell=1,\dots,L}\in\mathbb{R}^{p\times L}}{\operatorname{argmin}} \sum_{\ell=1}^{L}\sum_{i=1}^{n} \rho_{u_\ell}(Y_i - X_i'\beta_\ell) \\ \text{s.t.} \quad x'\beta_{\ell+1} \geq x'\beta_\ell \quad \text{for} \quad (x,\ell) \in \mathcal{X}_k \times \{1,\dots,L-1\} \end{array}, \tag{12}$$

where $0 < u_1 < \cdots < u_L < 1$ are quantile levels and $\mathcal{X}_k \subset \mathcal{X}$ is a finite set of covariate values at which the non-crossing condition is imposed. Without the inequality constraints, this reduces to L separate standard quantile regressions. The constraints ensure that crossings do not occur on the designated region, while minimizing the total check loss.

Although intuitive, this method has limitations. To establish convergence, Bondell et al. (2010) assume that L, the number of quantiles considered, is not too large relative to n, the sample size. Moreover, their results only guarantee pointwise convergence, which does not directly extend to (local) uniform convergence of the quantile process $\beta(\cdot)$.

In contrast, our estimator (11) imposes restrictions over the entire unit interval and achieves local uniform consistency, as shown in Theorem 4.1. By exploiting information across all quantile levels, it is expected to deliver superior finite-sample performance compared with Bondell et al.'s estimator. This expectation is confirmed by our numerical simulations in Section 5.

In many applications, restrictions on the derivative of β are of interest. For instance, the monotonicity of the bidding strategy in Sections 1.1 and 6 involves conditions on derivatives. Such restrictions cannot be enforced within $\hat{\beta}^{\mathrm{BRW}}$, since it only accounts for finitely many quantiles. While discrete approximations of derivative restrictions may be used in practice, analyzing their convergence becomes even more delicate.

We further compare the computational efficiency of the two estimators by examining the number of variables, the number of constraints, and the sparsity index of each formulation. Our estimator generally involves a smaller set of variables and constraints, whereas Bondell et al.'s method tends to yield a lower (i.e., more favorable) sparsity index. It is difficult to draw a general conclusion about which approach is computationally superior. Further details are provided in Appendix E.

Another stream of research on quantile regression estimation without crossings is the rearrangement approach proposed by Chernozhukov et al. (2010). Similar ideas are studied by Dette and Volgushev (2008). The procedure is roughly as follows. First, construct a preliminary estimate $\hat{Q}_{Y|X}(\cdot \mid x)$ of the conditional quantile function without imposing non-crossing. Then, "rearrange" the estimate by defining another estimator $\tilde{Q}_{Y|X}(\cdot \mid x)$ as

$$\tilde{Q}_{Y|X}(u_\ell \mid x) \coloneqq \ell\text{-th lowest number in } \left\{\hat{Q}_{Y|X}(u_1 \mid x), \ldots, \hat{Q}_{Y|X}(u_L \mid x)\right\},$$

where $0 < u_1 < \cdots < u_L < 1$ is a fine grid of the unit interval. Although $\hat{Q}_{Y|X}(\cdot \mid x)$ is not necessarily increasing, $\tilde{Q}_{Y|X}(\cdot \mid x)$ is increasing by construction.

The rearrangement approach is conceptually simple and computationally attractive. Moreover, Chernozhukov et al. (2010) show that the rearranged estimator is asymptotically equivalent to the preliminary estimator. However, it cannot be directly extended to shape restrictions beyond the non-crossing condition. For example, it is unclear how to enforce monotonicity of the bidding strategy when estimating the bid quantile function in the auction model.

Rearranged quantile estimates are typically non-smooth in covariates z, even if the original estimates are smooth. This potentially undermines the interpretability of the estimates in practice. See the second panel of Figure 7.

Another difficulty arises from the fact that the rearrangement method was developed for nonparametric estimation of conditional quantile functions. When the conditional quantile function is assumed to be (quasi-)linear, as in this paper, the quantile regression coefficient β is central because it captures the conditional quantile elasticity with respect to covariates. If rearrangement is applied after estimating β without the non-crossing restriction, this interpretability of β is lost.

4 Asymptotic Properties

4.1 Uniform Consistency

Due to the second constraint of the LP (10), the estimator (11) respects the desired shape restriction. If the population linear approximation $u \mapsto x'\beta(u)$ of the conditional quantile satisfies the same shape restriction, the estimator is uniformly consistent as shown in Theorem 4.1. To state the result, we first provide several assumptions for it.

Assumption BS. The support of F_X is bounded.

Assumption NC. The restriction $(\mathcal{A}_f, \delta, K)$ ensures the non-crossing condition, that is, it is written as

$$\mathcal{A}_f = (\mathcal{A}_f^{\mathrm{NC}}, \mathcal{A}'_f), \quad \delta = (\delta^{\mathrm{NC}}, \delta'), \quad K = K^{\mathrm{NC}} \times K',$$

where the restriction $(\mathcal{A}_f^{\mathrm{NC}}, \delta^{\mathrm{NC}}, K^{\mathrm{NC}})$ is the one given in Example 1 and $(\mathcal{A}'_f, \delta', K')$ is a conic restriction on a Banach space $\mathbb{D}'$.

Assumption SR. The quantile regression coefficient $\beta \in \mathbb{B}^{(k)}$ satisfies

$$\mathcal{A}'_f \beta + \delta' \in K'.$$

Assumption IF. There exists a sequence $\Lambda_n^{\mathrm{IF}} \in \mathbb{R}^{p \times J_n}$ such that

$$\mathcal{A}_f(\Lambda_n^{\mathrm{IF}} D_u m^{J_n}(\cdot)) + \delta \in \mathrm{Int}_{\mathbb{D}}(K)$$

holds for each n.

Assumption CP. There exist a constant $M > 0$ and a sequence $\Lambda_n^{\mathrm{CP}} \in \mathbb{R}^{p \times J_n}$ such that

$$\mathcal{A}_f(\Lambda_n^{\mathrm{CP}} D_u m^{J_n}(\cdot)) + \delta \in K \quad \text{and} \quad \inf_{(u,x) \in [0,1] \times \mathcal{X}} x' \Lambda_n^{\mathrm{CP}} m^{J_n}(u) > -M$$

hold for each n.

Assumption BB. There exists a sequence $\Lambda_n^{\mathrm{BB}} \in \mathbb{R}^{p \times J_n}$ such that

$$\sup_{n \in \mathbb{N}} \left\| \Lambda_n^{\mathrm{BB}} \right\| < \infty \quad \text{and} \quad \inf_{\substack{n \in \mathbb{N} \\ \rho \in K^* \\ \|\rho\| = 1}} \left\langle \rho, \mathcal{A}_f(\Lambda_n^{\mathrm{BB}} D_u m^{J_n}(\cdot)) \right\rangle > 0,$$

where $K^* := \{k^* \in \mathbb{D}^* \mid \forall k \in K, \langle k^*, k \rangle \geq 0\}$, D^* is the topological dual of $\mathbb{D}$, and $\langle \cdot, \cdot \rangle$ is the dual pairing. Conventionally, if the set over which an infimum is taken is empty, the value is defined as $+\infty$.

Assumption BS together with the condition $\beta \in \mathbb{B}^{(k)}$ in Assumption SR requires the support of F_{XY} to be bounded. Combining Assumption SR with Assumptions QL, NC implies that the quantile regression coefficient β satisfies

$$\mathcal{A}_f \beta + \delta \in K,$$

which implies that the shape restriction is assumed to be satisfied in population. When only the non-crossing condition is of interest, one can set

$$\mathbb{D}' = \mathbb{R}, \quad \mathcal{A}'_f = 0, \quad \delta' = 0, \quad \text{and} \quad K' = \mathbb{R}.$$

Assumptions IF, CP, BB are mild technical conditions on shape restrictions. Assumption IF ensures that the empirical problem (10) has a feasible solution Λ_n^{IF} such that any small perturbation from Λ_n^{IF} is also feasible. This assumption is not too restrictive, given that all shape restrictions considered in Section 3.1 satisfy it when the transformation f forms a polynomial basis, i.e., $f(z) = (1, z, \cdots, z^{p-1})'$. See in Appendix F for details. A similar argument applies to other transformation functions straightforwardly. Assumption IF is critical to exploit the duality structure of (10) since it plays a role of Slater's condition. See Appendix B.

Assumption CP is a mild condition that ensures that the family of convex functions $u \mapsto x'\hat{\Lambda}_n m^{J_n}(u)$ is sequentially relatively compact. A simple but strong sufficient condition for Assumption CP is $\delta = 0$, in which case, it is satisfied with $\Lambda_n^{\text{CP}} = 0$ and any $M > 0$.

Assumption BB is another technical condition that is used to prove the convergence of the dual problem of (10). It guarantees that if the value of $\mathcal{A}_f$ never approaches zero along any sequence in the feasible set that does not converge to zero. See Lemma B.7 for details. Appendix F shows that for the shape restrictions that appear in Section 3.1, the choice $\Lambda_n^{\text{BB}} = \Lambda_n^{\text{IF}}$ satisfies Assumption BB.

Now, we are ready to state the uniform consistency of our estimator.

Theorem 4.1. *Suppose that Assumptions QR, QL, LS, FR, BS, NC, SR, IF, CP, BB hold and $J_n \to \infty$. Let $\hat{\beta}_n(\cdot)$ be defined by (11). Then, for any compact set $K \subset (0,1)$,*

$$\lim_{n\to\infty} \sup_{u\in K} \left\| \hat{\beta}_n(u) - \beta(u) \right\| = 0$$

holds almost surely.

Notably, Theorem 4.1 establishes uniform consistency of our estimator without imposing any growth rate condition on the degree of the polynomial basis J_n. This is a feature contrast to the usual nonparametric regression. In general, when a nonparametric function is estimated using a polynomial basis of degree J_n, an excessively fast growth of J_n relative to the sample size n can lead to high variance and a failure of uniform consistency. To prevent this, the growth rate of J_n is typically controlled; see, e.g., Newey (1997).

By contrast, Theorem 4.1 does no require a rate condition on J_n, despite the linear program (10) serves as a polynomial sieve approximation to its population counterpart. This is made possible by the non-crossing condition (Assumption NC), specifically the monotonicity of $u \mapsto x'\hat{\beta}_n(u)$. Thanks to the monotonicity, the function $u \mapsto x'\hat{\beta}_n(u)$ cannot oscillate even when the degree J_n is large, which effectively regularizes the estimator. Such automatic regularization through shape constraints is documented in the literature on nonparametric least squares estimation too; see, for example, Feng et al. (2022).

Appendix G presents numerical simulations comparing the performance of our estimator across different values of J_n. The results show that the estimates are largely similar unless J_n is chosen too small.

The standard framework of sieve estimation can also be used to establish uniform consistency. However, it typically requires control of the covering number of the sieve space, which in turn implies that J_n cannot be too large. See Remark 3.3 and Condition 3.5M in Chen (2007), for example.

4.2 Asymptotic Normality

Next, we investigate the limit distribution of $\hat{\beta}_n$. Throughout this subsection, we assume that the polynomial approximation of the population coefficient is exact.

Assumption PL. There exist an integer $J_* > 0$ and a matrix $\Lambda_* \in \mathbb{R}^{p \times J_*}$ such that

$$\beta(u) = \Lambda_* D_u m^{J_*}(u)$$

for $u \in [0, 1]$.

Theorem 4.2. *Suppose Assumptions PL, QR, QL, NC, CP and $J_n = \bar{J}$ for some $\bar{J} \geq J_*$. Moreover, suppose that $\mathcal{A}_f'\beta + \delta' \in \mathrm{Int}_{\mathbb{D}'}(K')$ and that there exists a constant $\varepsilon > 0$ such that*

$$\inf_{(u,x) \in [0,1] \times \mathcal{X}} x' D_u \beta(u) > \varepsilon, \tag{13}$$

and that the matrix $\mathbb{E}[XX' \mid \bar{U}]$ is invertible almost surely, where $\bar{U} := (X'\beta(\cdot))^{-1}(Y)$. Then, for the estimator $\hat{\beta}_n$ defined by (11), $\sqrt{n}\left(\hat{\beta}_n(\cdot) - \beta(\cdot)\right)$ weakly converges to a tight centered Gaussian process $\mathbb{G}(\cdot)$ in $\ell^\infty([0,1], \mathbb{R}^p)$, of which covariance function is given by

$$\mathbb{E}[\mathbb{G}(u_1)\mathbb{G}(u_2)'] = ((D_u m^{\bar{J}}(u_1))' \otimes I_p)(V^{-1}WV^{-1})(D_u m^{\bar{J}}(u_2) \otimes I_p),$$

where

$$W := \mathbb{E}\left[\left(X \otimes m^{\bar{J}}(\bar{U})\right)\left(X \otimes m^{\bar{J}}(\bar{U})\right)'\right]$$

and

$$V := \mathbb{E}\left[\frac{1}{X' D_u \beta(\bar{U})}\left(X \otimes D_u m^{\bar{J}}(\bar{U})\right)\left(X \otimes D_u m^{\bar{J}}(\bar{U})\right)'\right].$$

Clearly, Assumption PL is restrictive, but we believe that it is general enough in practical applications. Developing a distributional theory of the estimator (11) without Assumption PL is an important open question.

The condition (13) assumes that the slope of the (linear approximation of the) conditional quantile function $u \mapsto x'\beta(u)$ is uniformly away from zero. When the linear specification (5) holds, this condition is equivalent to the uniform boundedness of the conditional density of Y given X, because

$$\sup_{(u,x)\in[0,1]\times\mathcal{X}} f_{Y|X}(x'\beta(u) \mid x) = \sup_{(u,x)\in[0,1]\times\mathcal{X}} \frac{1}{x'D_u\beta(u)} < 1/\varepsilon.$$

The invertibility of $\mathbb{E}[XX' \mid \bar{U}]$ is used to ensure that V is full-rank. It is easy to show that under Assumption FR, the this condition is satisfied if the conditional quantile function is linear in x.

The asymptotic variance in Theorem 4.2 also appears in existing studies on parametric quantile regression. For example, Frumento and Bottai (2016) consider exactly the same specification of quantile regression coefficients as Assumption PL. Their estimator is obtained by minimizing an integrated empirical check loss function over the parameter space. Although this approach is completely different from ours, it converges to the same asymptotic distribution. Parametric models that allow for nonlinear parameter dependence are studied by Firpo et al. (2022). One of their estimators is defined as the function in the parametric model that is closest to the KB estimator with respect to a weighted squared loss. This estimator is also shown to have the same asymptotic distribution as ours.

For statistical inference based on the asymptotic normality, we need to estimate W and V. Let $\hat{U}_i := (X_i'\hat{\beta}_n(\cdot))^{-1}(Y_i)$ for $i = 1, \ldots, n$, and define

$$\hat{W} := \frac{1}{n}\sum_{i=1}^{n} \left(X_i \otimes m^{\bar{J}}(\hat{U}_i)\right)\left(X_i \otimes m^{\bar{J}}(\hat{U}_i)\right)' \tag{14}$$

and

$$\hat{V} := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{X_i'D_u\hat{\beta}_n(\hat{U}_i)} \left(X_i \otimes D_u m^{\bar{J}}(\hat{U}_i)\right)\left(X_i \otimes D_u m^{\bar{J}}(\hat{U}_i)\right)'. \tag{15}$$

Proposition 4.3. *Suppose that Assumptions QR, QL, LS, BS, NC, SR, IF, CP, BB hold. Furthermore, assume that the considered conic restriction contains* $(\mathcal{A}_f^{(\varepsilon)}, \delta^{(\varepsilon)}, K^{(\varepsilon)})$ *where*

$$\mathcal{A}_f^{(\varepsilon)} : \mathbb{B}^{(k)} \to C([0,1]\times\mathcal{X}); \quad (\mathcal{A}_f^{(\varepsilon)}\beta)(u,x) := x'D_u\beta(u),$$

$\delta^{(\varepsilon)} := -\varepsilon$, *and*

$$K^{(\varepsilon)} := \{k \in C([0,1]\times\mathcal{X}) \mid k(u,x) \geq 0 \text{ for all } (u,x) \in [0,1]\times\mathcal{X}\}.$$

Then, $\hat{W} \to W$ and $\hat{V} \to V$ hold almost surely.

The restriction $\mathcal{A}_f^{(\varepsilon)}\beta + \delta^{(\varepsilon)} \in K^{(\varepsilon)}$ implies that the average rate of change of $u \mapsto x'\beta(u)$ is no less than ε. Hence, it is a stricter restriction than the non-crossing restriction $(\mathcal{A}_f^{\mathrm{NC}}, \delta^{\mathrm{NC}}, K^{\mathrm{NC}})$, which is equivalent to $(\mathcal{A}_f^{(\varepsilon)}, \delta^{(\varepsilon)}, K^{(\varepsilon)})$ for $\varepsilon = 0$. This tighter restriction incorporates the lower bound (13) of the population slope function. We use it to establish $\hat{V} \to V$, whereas $\hat{W} \to W$ does not require it.

5 Monte Carlo Simulations

We compare the performance of our estimator $\hat{\beta}_n$ defined by (11) to that of the estimators proposed by Koenker and Bassett (1978) ($\hat{\beta}_n^{\mathrm{KB}}$) and Bondell et al. (2010) ($\hat{\beta}_n^{\mathrm{BRW}}$).

DGP1. First, we consider the following DGP.

(DGP1) The data is independently drawn from

$$Y_i = X_i'(\alpha + \gamma\Phi^{-1}(U_i)), \quad X_i = (1, Z_i'), \quad (Z_i, U_i) \sim U[0,1]^4 \otimes U[0,1]$$

for $i = 1, \ldots, n$, where Φ^{-1} is the quantile function of the standard Gaussian distribution, $\alpha = (1,1,1,1,1)'$ and $\gamma = (1, 0.1, 0.1, 0.1, 0.1)'$. This DGP is taken from Bondell et al. (2010). In this case, the quantile regression coefficient is $\beta(u) = \alpha + \gamma\Phi^{-1}(u)$.

Let $(u_1, u_2, u_3, u_4, u_5, u_6, u_7) = (0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99)$. The estimator $\hat{\beta}_n^{\mathrm{KB}}$ of Koenker and Bassett (1978) is computed for each quantile level u_k for $k = 1, \ldots, 7$ separately. Bondell et al.'s estimator $\hat{\beta}_n^{\mathrm{BRW}}$ is implemented so that the quantile crossing does not happen only at these quantile levels; see equations (2)-(3) of Bondell et al. (2010). In (DGP1), our estimator $\hat{\beta}_n$ is implemented with the non-crossing condition. We set the degree of the polynomial approximation to $J_n = 8$ and use this degree in all subsequent simulations. The choice of J_n has little impact on the results unless it is chosen too small. Appendix G provides a comparison of the estimator's performance across different values of J_n. For an estimator $\tilde{\beta} = (\tilde{\beta}_0, \ldots, \tilde{\beta}_{p-1})'$, define

$$L_\beta(\tilde{\beta}, u_k) := \sqrt{\frac{1}{p}\sum_{i=0}^{p-1}\Big(\tilde{\beta}_i(u_k) - \beta_i(u_k)\Big)^2} \tag{16}$$

for $k = 1, \ldots, 7$. We compare the three estimators based on this loss function.

Setting the sample size $n = 100$, we generate a dataset $((X_i, Y_i))_{i=1}^n$, estimate β by the three estimators, and compute the loss function (16) for each of them. We repeat this procedure 1000 times.

Figure 5 presents the boxplot of the loss function values (as defined in equation (16)) for various quantile levels $u = u_1, \ldots, u_7$ under DGP1. The performance of the three estimators is compared. Across all quantile levels, the proposed method ("Ours") consistently achieves lower median loss values than the estimators of Koenker and Bassett (1978) ("KB") and Bondell et al. (2010) ("BRW"), with especially notable improvements in the tails (i.e., at $u = 0.01, 0.99$). Moreover, the worst-case losses tend to be smaller for the proposed estimator across most quantile levels. This suggests that the method not only improves median performance but also mitigates extreme errors.

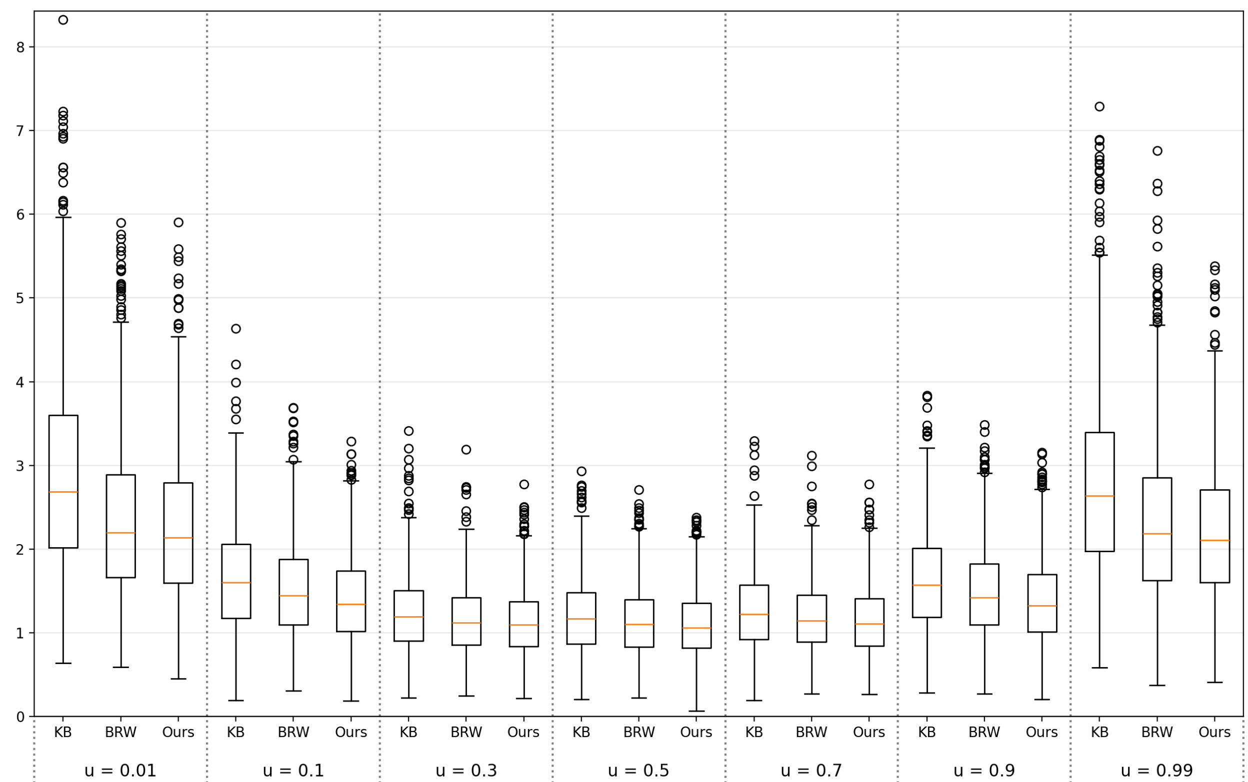

Figure 5: Boxplot of the loss values for the estimators $\hat{\beta}_n^{\mathrm{KB}}$, $\hat{\beta}_n^{\mathrm{BRW}}$, and $\hat{\beta}_n$ at different quantile levels $u = u_1, \ldots, u_7$, obtained from samples generated by DGP1. The horizontal axis represents different pairs of an estimator and a quantile level.

DGP2. Next, we consider the DGP defind in (6), where the true conditional quantile function is nonlinear in x, but Assumption QL is satisfied. We generate a dataset for $n = 30$, estimate β with the three estimators, and compute the loss function (16) for each of them.

We repeat this process for 1000 times. Our estimator is implemented with the non-crossing restriction only.

Figure 6 reports the results. Like the previous experiment, we observe that the proposed estimator tends to outperform the existing estimators, though the differences are less significant compared with (DGP1). It is also worth mentioning that, in the worst-case scenario for each quantile level, the loss value of our estimator is always smaller than those of the other two.

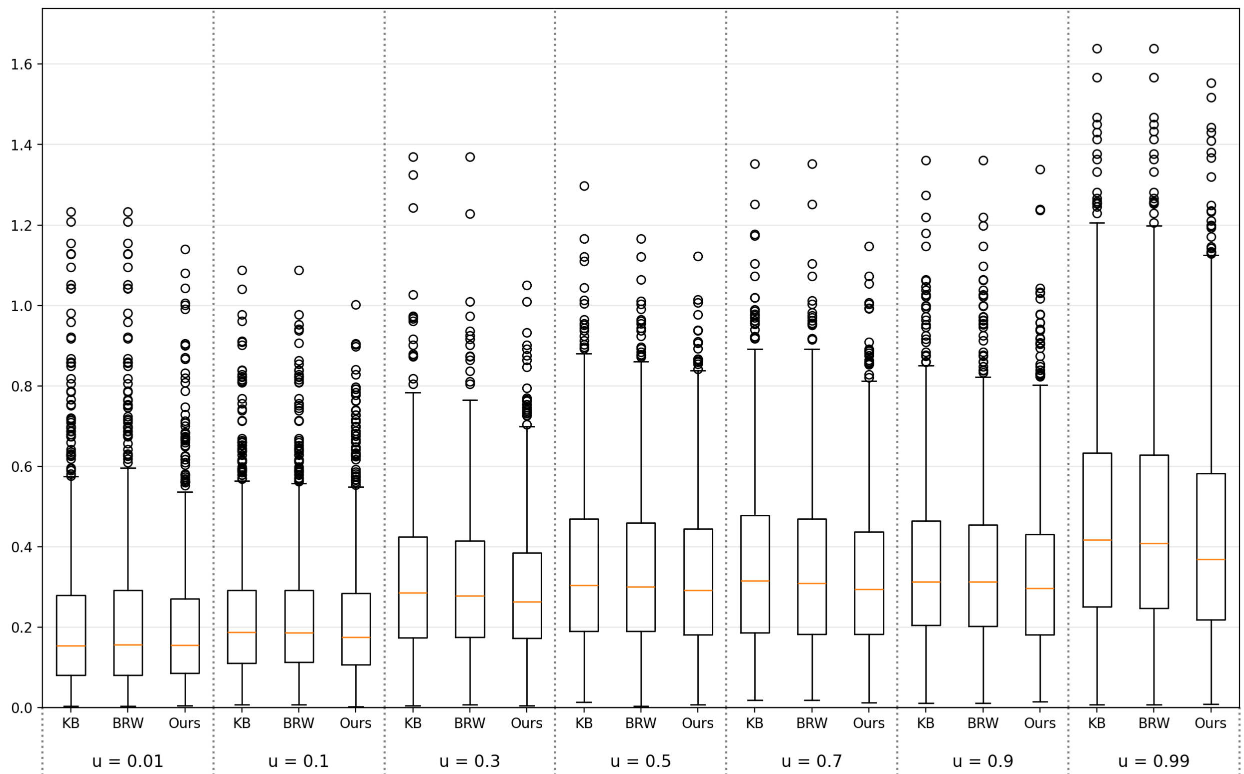

Figure 6: Boxplot of the loss values for the estimators $\hat{\beta}_n^{\text{KB}}$, $\hat{\beta}_n^{\text{BRW}}$, and $\hat{\beta}_n$ at different quantile levels $u = u_1, \ldots, u_7$, obtained from samples generated by (6). The horizontal axis represents different pairs of an estimator and a quantile level.

DGP3. Finally, we investigate how shape restrictions affect estimated conditional quantile functions. We consider the following DGP.

(DGP3) The data is independently drawn from

$$Y_i = 1 + U_i(e^{2(Z_i+1)} - 1), \quad (Z_i, U_i) \sim \text{Beta}(3, 3) \otimes U[0, 1]$$

for $i = 1, \ldots, n$.

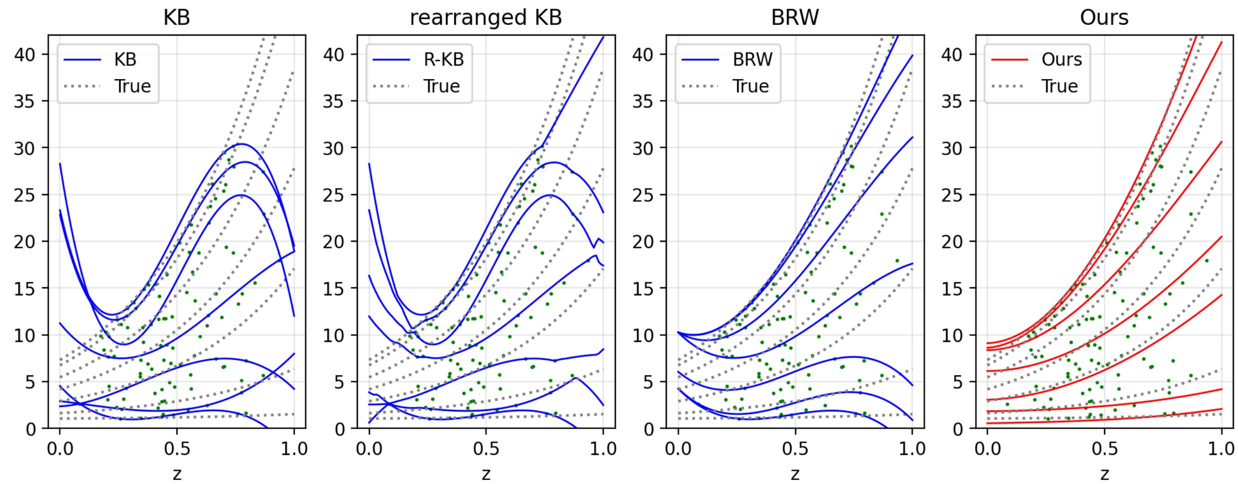

Figure 7: The fitted conditional quantile functions at quantile levels $u = u_1, \ldots, u_7$ are represented by solid lines. The dotted lines represent the true quantile curves. The green dots represent the generated data points.

We apply each estimation method to the polynomial regression model of degree three. That is, we estimate

$$\beta(u) = \underset{\beta \in \mathbb{R}^4}{\operatorname{argmin}}\, \mathbb{E}\left[\rho_u\left(Y - X'\beta\right)\right],$$

where $X = (1, Z, Z^2, Z^3)'$. Observe also that the map $z \mapsto 1 + u(e^{2(z+1)} - 1)$ is convex and increasing on $z \in [0, 1]$ for all $u \in [0, 1]$. Assuming that this fact is implied by some economic theory, we impose the convexity and non-decreasingness in z as well as the non-crossing condition on our estimator $\hat{\beta}_n$.

Figure 7 shows the fitted conditional quantile curves of the three estimators. Clearly, the classical estimator ("KB") suffers from the quantile crossing problem.

The second panel ("rearranged KB") reports the estimates of KB's estimator with the rearrangement proposed by Chernozhukov et al. (2010). It is observed that the shape restrictions of interest are violated, even though the quantile crossing problem is resolved.

The estimator of Bondell et al. (2010) ("BRW") resolves the crossing problem, but the shape restrictions are not satisfied especially in the region close to the boundary. Moreover, the quantile curve estimates meet at the boundary, though they do not cross. This phenomenon, called quantile convergence by Ando and Li (2025), contradicts to the assumption that the conditional distribution of Y given Z is continuous.

The proposed estimator ("Ours") does not encounter these issues and effectively captures the shape of the true quantile curves. Also, the quantile convergence does not occur. Note that the fitting is not intended to be perfect, given that the true conditional quantile is not linear in x.

Finally, we evaluate the estimation precision with respect to the conditional quantile function for each estimator. We generate datasets $((Y_i^{(b)}, Z_i^{(b)}))_{i=1}^{100}$ for $b = 1, \ldots, 1000$ from DGP3. For each estimator $\tilde{Q}_{Y|Z}$ of the conditional quantile function $Q_{Y|Z}(u \mid z) \coloneqq 1 + u(e^{2(z+1)} - 1)$, we calculate a loss function

$$L_Q^{(b)}(\tilde{Q}_{Y|Z}, u_k) \coloneqq \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_{Y|Z}(u_k \mid Z_i^{(b)}) - \tilde{Q}_{Y|Z}(u_k \mid Z_i^{(b)})\right)^2},$$

where $b = 1, \ldots, 1000$ and $k = 1, \ldots, 7$.

Figure 8 presents the boxplot of the values of $L_Q^{(b)}$. The KB estimator and its rearranged version ("RKB") yield similar results, with the latter performing slightly better, in line with the findings of Chernozhukov et al. (2010). The BRW estimator, which enforces non-crossing over a grid of quantiles, tends to achieve lower loss values than KB and RKB, particularly around the middle quantiles. Among the four methods, our estimator attains the smallest loss values, highlighting the advantage of incorporating information about the ground truth into the estimation procedure whenever it is theoretically justified.

6 Application to Estimation of First-price Auctions

6.1 Model and Estimator

The objective of this subsection is to illustrate how our estimator can be applied to the structural estimation of auction models. We begin by recalling the setup from Section 1.1. Consider datasets from first-price sealed-bid auctions indexed by $\ell = 1, \ldots, L$. Each auction involves a fixed number I of risk-neutral bidders.[2] For auction ℓ, let $Z_\ell \in \mathbb{R}^q$ denote the auction-specific characteristics, and let $B_{\ell i}$ be the bid submitted by bidder i at auction ℓ. The econometrician observes data $((Z_\ell, B_{\ell 1}, \ldots, B_{\ell I}))_{1 \le \ell \le L}$, generated independently across auctions.

At auction ℓ with covariate $Z_\ell = z$, each bidder's private value is drawn independently from the valuation distribution $F_V(\cdot \mid z)$. As discussed in Section 1.1, the unique symmetric Bayesian-Nash equilibrium bidding strategy is strictly increasing and given by the inverse of

[2]The analysis can be extended to risk-averse bidders by assuming a constant relative risk aversion utility function. However, identifying the degree of risk aversion typically requires additional structure or data; see Gimenes and Guerre (2022) for details. We expect that our method remains valid under this generalization, but we do not pursue this extension in the present paper.

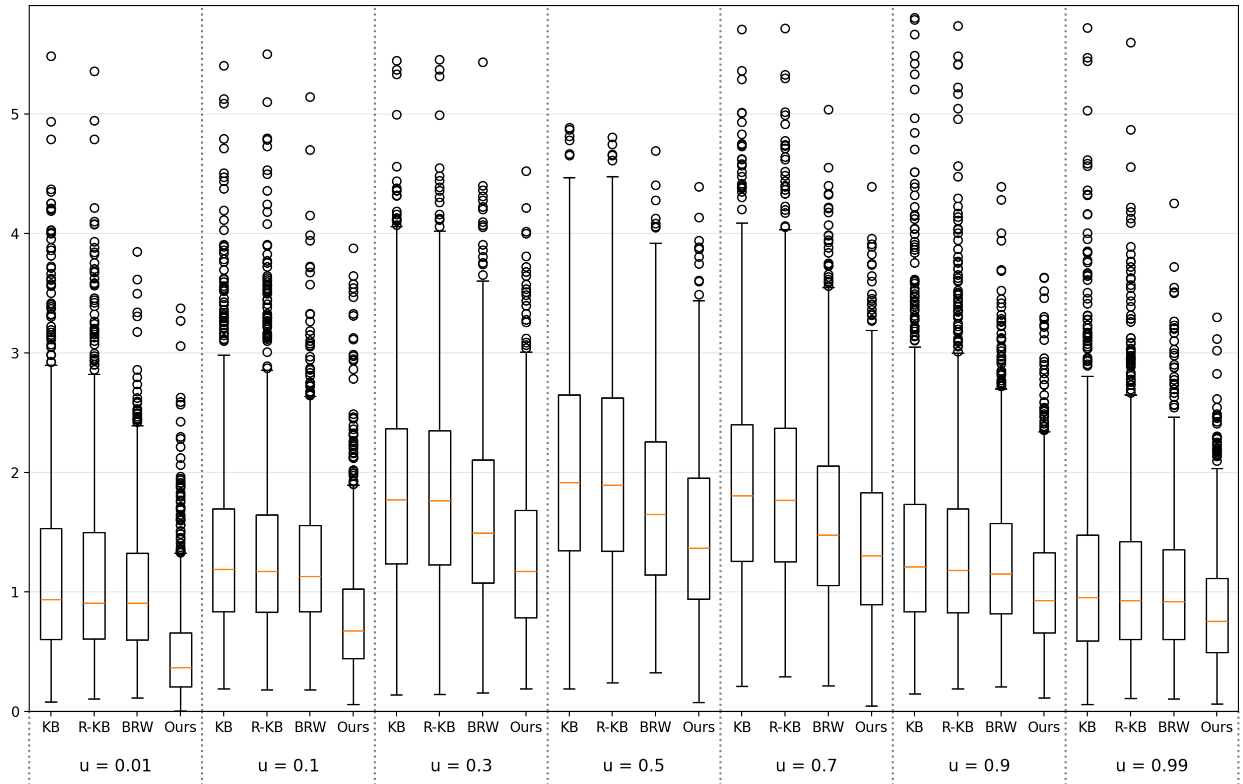

Figure 8: Boxplot of the loss values for the estimates of conditional quantile functions. Loss values are calculated for KB, rearranged KB (RKB), and BRW estimators as well as our estimator at quantile levels $u = u_1, \ldots, u_7$. The horizontal axis represents different pairs of an estimator and a quantile level. For example, the box at BRW-0.5 is the loss distribution of $\hat{\beta}_n^{\mathrm{BRW}}$ at $u = 0.5$.

the map

$$\xi(b \mid z) = b + \frac{F_B(b \mid z)}{(I-1) f_B(b \mid z)}, \tag{17}$$

where $F_B(\cdot \mid z)$ and $f_B(\cdot \mid z)$ are the conditional distribution and density function of bids. In what follows, we maintain the assumption that bidders adopt this equilibrium bidding strategy.

Following Gimenes and Guerre (2022), we consider the linear quantile regression model

$$Q_B(u \mid x) = x'\beta(u),$$

where $x = (1, z')' \in \mathbb{R}^p$ with $p = q + 1$. Using the identities $F_B(Q_B(u \mid x) \mid x) = u$ and $f_B(Q_B(u \mid x)) = (D_u Q_B(u \mid x))^{-1}$, the inverse bidding strategy (17) satisfies

$$\xi(Q_B(u \mid x) \mid x) = Q_B(u \mid x) + \frac{u}{I-1} \left(D_u Q_B(u \mid x) \right) = x'\gamma(u),$$

where

$$\gamma(u) := \beta(u) + \frac{u}{I-1} D_u \beta(u).$$

Thus, by estimating the bid quantile regression model to obtain β, one can recover the private value of bidders who submit bid $Q_B(u \mid x)$. Moreover, this leads to the valuation quantile function $Q_V(u \mid x) = x'\gamma(u)$. Gimenes and Guerre (2022) implement this approach using a local polynomial estimator for the bid quantile regression.

A limitation of the estimator in Gimenes and Guerre (2022) is that it does not incorporate shape constraints that necessarily hold in population. First, the definition of a quantile function requires that

$$\text{the map } u \mapsto Q_B(u \mid x) = x'\beta(u) \text{ is increasing.} \tag{18}$$

Second, the economic model implies that the equilibrium strategy ξ is monotone, which in turn requires

$$\text{the map } u \mapsto \xi(Q_B(u \mid x) \mid x) = x' \left(\beta(u) + \frac{u}{I-1} D_u \beta(u) \right) \text{ is increasing.} \tag{19}$$

Now, we estimate the bid quantile regression coefficient β with these two shape restrictions imposed. Let $\varepsilon \geq 0$ be a non-negative number. Consider the following linear program:

$$\inf_{\substack{\psi \in \mathbb{R}^{LI} \\ \Lambda \in \mathbb{R}^{p \times J_n}}} \frac{1}{LI} \sum_{\ell=1}^{L} \sum_{i=1}^{I} \psi_{\ell i}$$

$$\text{s.t.} \quad \begin{cases} u B_{\ell i} \leq \psi_{\ell i} + X_\ell' \Lambda m^{J_n}(u) \text{ for } (u, \ell, i) \in [0,1] \times \{1, \ldots, L\} \times \{1, \ldots, I\} \\ x' \Lambda D_u^2 m^{J_n}(u) \geq \varepsilon \text{ for } (u, x) \in [0,1] \times \mathcal{X} \\ x' \Lambda \left(\frac{I}{I-1} D_u^2 m^{J_n}(u) + \frac{u}{I-1} D_u^3 m^{J_n}(u) \right) \geq \varepsilon \text{ for } (u, x) \in [0,1] \times \mathcal{X} \end{cases}, \tag{20}$$

where the second and third constraints correspond to restrictions (18) and (19), respectively, as we see below in more details.

Let $\hat{\Lambda}_n^{(\varepsilon)}$ be a solution to LP (20). We estimate $\beta(\cdot)$ with

$$\hat{\beta}_n^{(\varepsilon)}(\cdot) := \hat{\Lambda}_n^{(\varepsilon)} D_u m^{J_n}(\cdot).$$

The bid quantile functions is estimated by

$$\hat{Q}_{B,n}^{(\varepsilon)}(u \mid x) := x' \hat{\beta}_n^{(\varepsilon)}(u),$$

which implies an estimator of the equilibrium inverse bidding strategy ξ

$$\hat{\xi}_n^{(\varepsilon)}(b \mid x) \coloneqq b + \frac{\hat{F}_{B,n}^{(\varepsilon)}(b \mid x)}{(I-1)\hat{f}_{B,n}^{(\varepsilon)}(b \mid x)},$$

where

$$\hat{F}_{B,n}^{(\varepsilon)}(b \mid x) \coloneqq \left(x'\hat{\beta}_n^{(\varepsilon)}\right)^{-1}(b) \quad \text{and} \quad \hat{f}_{B,n}^{(\varepsilon)}(b \mid x) \coloneqq \frac{1}{x' D_u \hat{\beta}_n^{(\varepsilon)}(\hat{F}_{B,n}^{(\varepsilon)}(b \mid x))}.$$

These estimators satisfy the imposed shape restrictions (18) and (19). In fact, the second constraint of (20) implies

$$D_u \hat{Q}_{B,n}^{(\varepsilon)}(u \mid x) = x'\hat{\Lambda}_n^{(\varepsilon)} D_u^2 m^{J_n}(u) \geq \varepsilon \geq 0,$$

and by the third constraint, we have

$$D_u \hat{\xi}_n^{(\varepsilon)}(\hat{Q}_{B,n}^{(\varepsilon)}(u \mid x) \mid x) = x'\hat{\Lambda}_n^{(\varepsilon)} \left(\frac{I}{I-1} D_u^2 m^{J_n}(u) + \frac{u}{I-1} D_u^3 m^{J_n}(u)\right) \geq \varepsilon \geq 0.$$

This estimation method also provides an estimator of the quantile function Q_V of private values. To see this, notice first that the value quantile function is given by $Q_V(u \mid x) = \xi(Q_B(u \mid x) \mid x)$ due to the monotonicity of Q_B and ξ. Thus, it can be estimated as

$$\hat{Q}_{V,n}^{(\varepsilon)}(u \mid x) \coloneqq \hat{\xi}_n^{(\varepsilon)}(\hat{Q}_{B,n}^{(\varepsilon)}(u \mid x) \mid x) = x' \left(\hat{\beta}_n^{(\varepsilon)}(u) + \frac{u}{I-1} D_u \hat{\beta}_n^{(\varepsilon)}(u)\right) \eqqcolon x'\hat{\gamma}_n^{(\varepsilon)}(u).$$

Because $\hat{\xi}_n^{(\varepsilon)}$ is monotone, the estimated quantile function of private values is automatically increasing. In particular, it never exhibits quantile crossings.

By Theorem 4.1, we can show the uniform consistency of these estimators.

Theorem 6.1. *Suppose that Assumptions QR, LS, BS, FR hold. Fix $x \in \mathcal{X}$. For $\varepsilon \geq 0$, suppose that $D_u Q_V(u \mid x), D_u Q_B(u \mid x) \geq \varepsilon$ hold for $(u, x) \in [0,1] \times \mathcal{X}$. Then, $\hat{\beta}_n^{(\varepsilon)}(\cdot) \to \beta(\cdot)$, $\hat{\gamma}_n^{(\varepsilon)}(\cdot) \to \gamma(\cdot)$, $\hat{Q}_{B,n}^{(\varepsilon)}(\cdot \mid x) \to Q_B(\cdot \mid x)$, and $\hat{Q}_{V,n}^{(\varepsilon)}(\cdot \mid x) \to Q_V(\cdot \mid x)$ hold locally uniformly in $(0,1)$ almost surely. Moreover, under the same assumption, if $\varepsilon > 0$, then the estimated bidding strategy $(\hat{\xi}_n^{(\varepsilon)})^{-1}(\cdot \mid x)$ is locally uniformly consistent on the interior $(\underline{v}, \overline{v})$ of $F_V(\cdot \mid x)$, and so is its inverse $\hat{\xi}_n^{(\varepsilon)}(\cdot \mid x)$ on $(\xi^{-1}(\underline{v} \mid x), \xi^{-1}(\overline{v} \mid x))$.*

It is remarkable that our estimators $(\hat{Q}_{B,n}^{(\varepsilon)}, \hat{Q}_{V,n}^{(\varepsilon)}, \hat{\xi}_n^{(\varepsilon)})$ are all increasing and satisfy the identity $\hat{\xi}_n^{(\varepsilon)}(\hat{Q}_{B,n}^{(\varepsilon)}(u \mid x) \mid x) = \hat{Q}_{V,n}^{(\varepsilon)}(u \mid x)$. Both properties are consistent with the underlying economic theory. Many existing estimators stemming from Guerre et al. (2000) (nearly) satisfy the identity, but their estimates of ξ are not monotone in general. Rearrangement-based estimators are increasing by construction, but they violate the identity, representing a deviation from the economic model under consideration.

6.2 Timber Auction Data

We apply our estimator to data from U.S. timber auctions studied by Lu and Perrigne (2008). The dataset contains records from 215 first-price sealed-bid auctions, including the number of bidders, the estimated timber volume (in thousands of board feet), and the appraisal value (in dollars per unit of volume). Following Gimenes and Guerre (2022), we focus on the subsample of auctions with exactly two bidders, which yields $L = 107$ observations. As auction-specific covariates, we use the log of volume and the appraisal value.

Before presenting our estimator, we first examine the extent of quantile crossings in the local polynomial quantile regression estimator of Gimenes and Guerre (2022). Table 1 reports the proportion of auctions in which the estimated bid or valuation quantile function is not increasing. Formally, we compute

$$\frac{\#\{\ell = 1, \ldots, L \mid \hat{Q}_B^{\mathrm{GG}}(\cdot \mid Z_\ell) \text{ is not increasing}\}}{L}$$

and

$$\frac{\#\{\ell = 1, \ldots, L \mid \hat{Q}_V^{\mathrm{GG}}(\cdot \mid Z_\ell) \text{ is not increasing}\}}{L}$$

for different degree-bandwidth pairs. We vary the bandwidth $h \in \{0.2, 0.3, 0.4, 0.5\}$ and the degree of the local polynomial $d \in \{1, 2, 3\}$. The proportion of auctions with crossings is highly sensitive to these tuning parameters. For higher-order polynomials ($d = 3$), violations of monotonicity are extremely frequent. Crossings are less common for lower-order polynomials, but this comes at the cost of limited approximation power. For the specification $(d, h) = (2, 0.3)$, which is used in the numerical studies of Gimenes and Guerre (2022), the proportion is relatively modest. However, given the instability observed for other tuning parameters and in our simulations in Section 1.1, this case appears more accidental than structural. Overall, the results show that monotonicity violations are not only a simulation artifact but also a serious issue in real data.

Even in the $(d, h) = (2, 0.3)$ case, where the crossing frequency is modest, crossings are far from negligible. Figure 9 shows the region in the covariate space where the estimated conditional quantile function is not increasing. Although only a handful of observed data points fall into this region (red x's), the estimated quantile curves cross once evaluated even slightly outside the support of the sample. This finding underscores the fragility of the

	$d=1$				$d=2$				$d=3$			
h	0.2	0.3	0.4	0.5	0.2	0.3	0.4	0.5	0.2	0.3	0.4	0.5
bid	0.037	0.009	0.009	0.299	0.234	0.037	0.187	0.402	0.047	0.411	0.626	0.364
value	0.093	0.037	0.009	0.000	0.486	0.037	0.131	0.374	0.935	0.757	0.757	0.720

Table 1: Proportions of auctions with quantile crossings in estimates of Gimenes and Guerre (2022) for different degrees of the local polynomial $d \in \{1, 2, 3\}$ and bandwidths $h \in \{0.2, 0.3, 0.4, 0.5\}$.

estimator: small extrapolations beyond the data domain can render the results invalid.

We now turn to the results from our proposed estimator, which imposes the relevant shape restrictions directly within the estimation procedure.

Figure 10 displays the estimated quantile regression coefficients β for the bid function, comparing our results ("Ours") with the unconstrained Koenker and Bassett (1978) ("KB") estimator and the Gimenes and Guerre (2022) ("GG") estimator. Visually, our estimates are closely aligned with those from the GG method, particularly in the central quantile range, where the consistency of our estimator is guaranteed. Both our estimator and the GG estimator produce smoother coefficient functions compared to the more erratic KB estimates.

It is worth noting that this smoothness is often desirable, as it suggests more stable coefficient estimates across neighboring quantiles. The crucial advantage of our method is that it achieves this smoothness while simultaneously guaranteeing the non-crossing property—the very property that the unconstrained GG estimator often violates, as demonstrated in Table 1. Our approach, therefore, retains the flexibility of modern methods while imposing the necessary theoretical discipline.

This pattern continues in Figure 11, which presents the corresponding coefficient estimates $\gamma(\cdot)$ for the valuation quantile function. Here, we only compare our results to the GG estimator, as the KB framework does not readily provide the derivative estimates necessary to compute the valuation coefficients. Once again, the estimated coefficients are similar. This suggests that imposing the shape constraints does not fundamentally distort the estimated economic relationships but rather corrects them to be theoretically coherent.

Finally, we plot the estimated bid and valuation quantile functions and the equilibrium inverse bidding strategy for an observed auction with covariates $z = (4.32, 5.67)'$ as an illustrative example. As shown in Figure 9, the GG estimates of the bid and valuation quantile functions are both non-monotonic. The left panel of Figure 12 compares the bid quantile functions estimated by GG and our estimators. Both estimators appear similar,

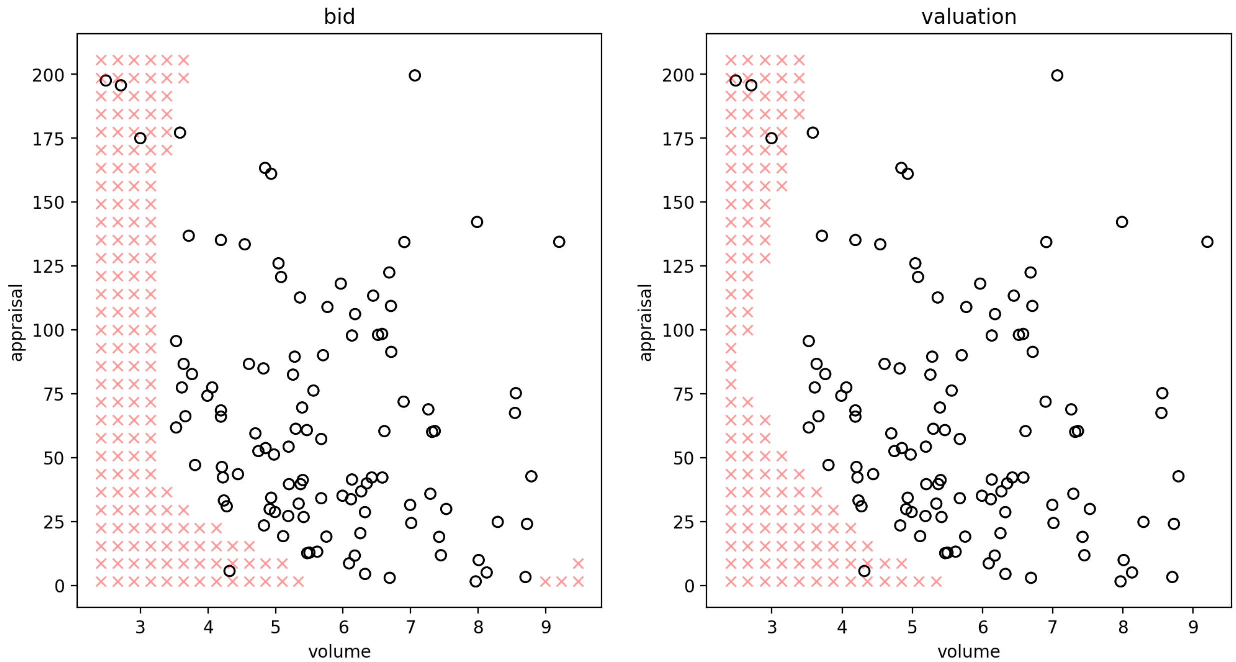

Figure 9: Regions of quantile crossing in the covariate space for Gimenes and Guerre's estimator with $(d, h) = (2, 0.3)$. The left panel is for the bid quantile function, and the right panel is for the valuation quantile function. Crossings occur in areas marked by red x's. Black circles represent observed data points.

although the GG estimate exhibits a small quantile crossing at low quantile levels. The middle panel shows the valuation quantile functions, where the difference is more pronounced. Our estimator produces a monotone quantile estimate, while the GG estimate suffers from a severe crossing problem. The right panel displays the implied equilibrium inverse bidding strategy. Here, the GG estimate[3] peaks around $b = 42$, whereas our estimate increases smoothly over the entire domain.

7 Concluding Remarks

This paper developed a new estimator for quantile regression coefficients that enforces the non-crossing condition and accommodates a broad class of economically motivated shape restrictions, such as monotonicity and convexity. Our approach builds on the variational characterization of linear quantile regression introduced by Carlier et al. (2016), which enables constraints to be imposed jointly across covariates and across the quantile index.

[3]The GG estimate of the inverse bidding strategy is constructed using formula (17). To obtain the bid distribution function, we invert the estimated quantile function. Since the GG quantile estimate is not monotone, we apply a rearrangement to define its inverse.

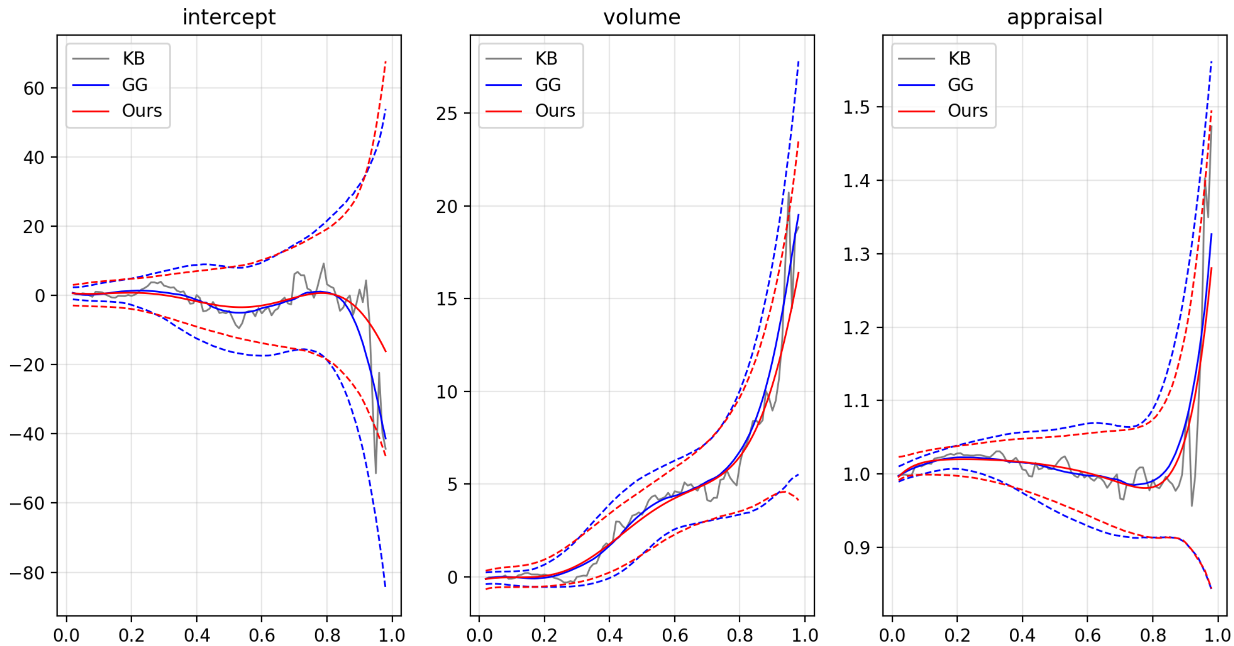

Figure 10: Bid coefficient estimates $\beta(\cdot)$. KB, GG, and our estimates are shown in gray, blue and red, respectively. Pointwise bootstrap 90% confidence intervals are shown in dashed lines for GG and our estimates.

We constructed a polynomial approximation scheme to the associated infinite-dimensional linear program, allowing for computational feasibility while preserving the structural integrity of the model. Leveraging linear programming duality, we established uniform consistency of the proposed estimator under general conditions. In addition, under a polynomial specification of the quantile process, we derived the asymptotic normality of our estimator.

Monte Carlo simulations confirmed the practical advantages of our estimator. Compared to existing methods, it yields more accurate estimates across quantile levels. These findings highlight the effectiveness of incorporating structural information directly into the estimation procedure.

Our empirical application to U.S. timber auction data further illustrates the usefulness of the framework. While existing estimators frequently produce non-monotone bid and value quantiles, our method enforces the required shape restrictions while remaining closely aligned with unconstrained estimates in regions where they are reliable. In doing so, it provides a coherent and stable basis for downstream economic analysis, as illustrated by the monotone inverse bidding strategies recovered in our application.

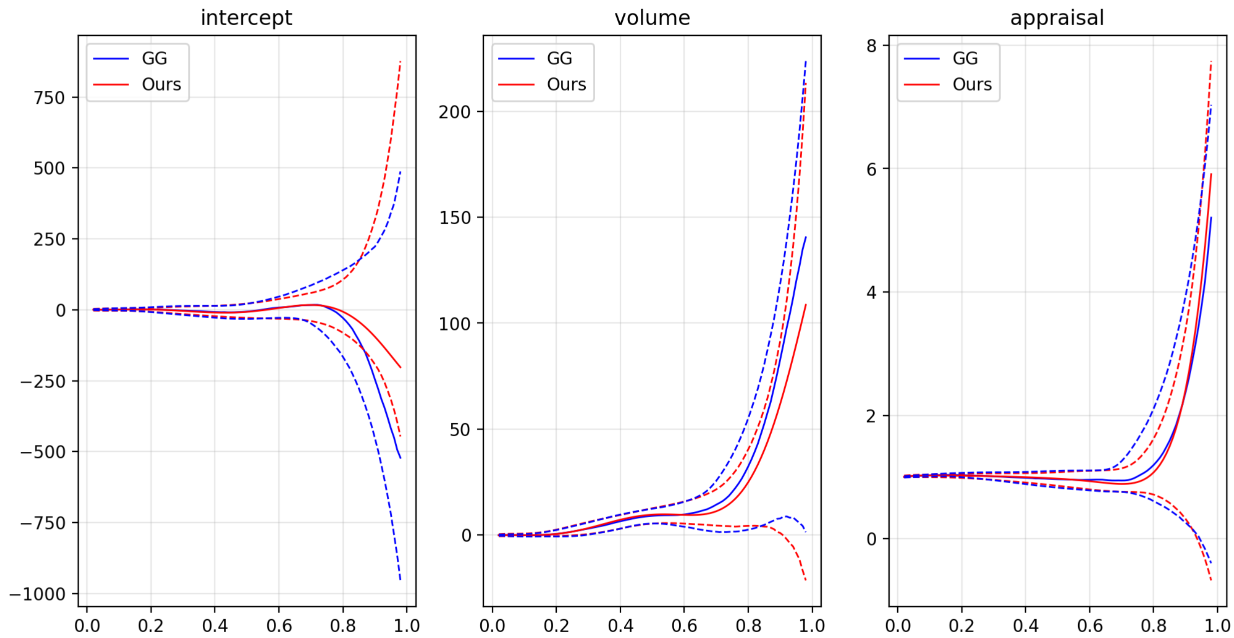

Figure 11: Valuation coefficient estimates $\gamma(\cdot)$. GG and our estimates are shown in blue and red, respectively. Pointwise bootstrap 90% confidence intervals are shown in dashed lines.

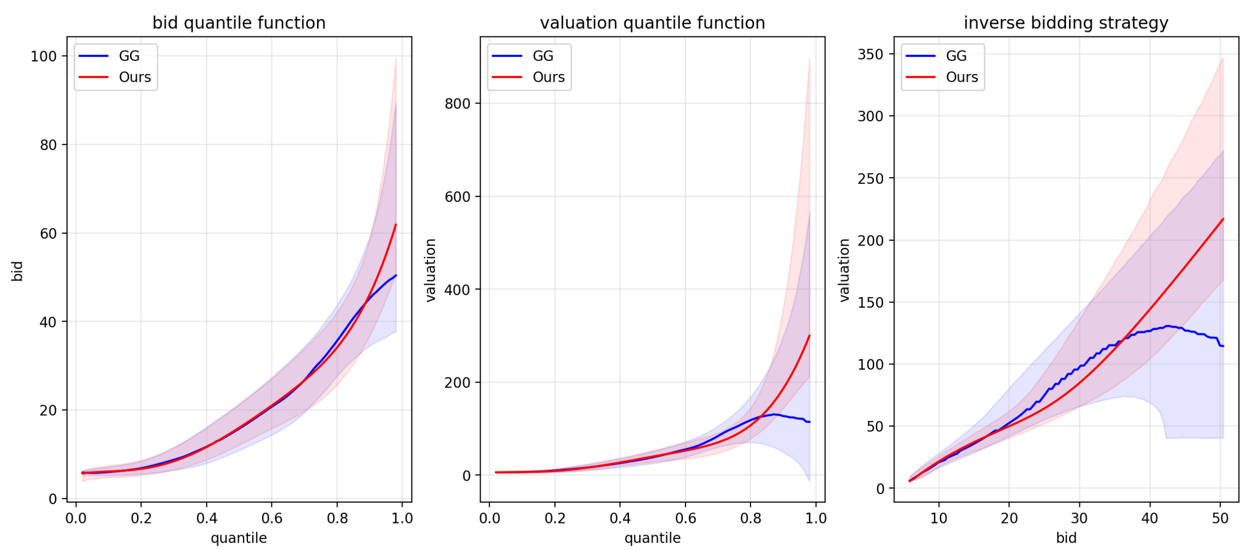

Figure 12: Estimated bid and valuation conditional quantile functions (left and middle, respectively), and the induced inverse bidding strategy (right). The GG estimates are shown in blue. Our estimates are shown in red. The shaded regions represent pointwise bootstrap 90% confidence intervals.

A Notation

Let $C[0,1]$ be the set of continuous functions on $[0,1]$. For an integer $k \geq 1$, let $C^k[0,1]$ be the set of k-times continuously differentiable functions on $(0,1)$ whose derivatives are continuously extended to $[0,1]$. Let $L_0^1[0,1]$ be the space of Lebesgue integrable functions with mean zero. For a set X and a metric space Y, let $\ell^\infty(X,Y)$ be the space of bounded functions from X to Y. For a continuously differentiable function $f : U \to \mathbb{R}^m$, where $U \subset \mathbb{R}^n$ is an open subset, let $D_x f(x) \coloneqq \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{i=1,\dots m, j=1,\dots,n}$ be the Jacobian matrix of f at $x \in U$. For a twice continuously differentiable function $g : U \to \mathbb{R}$, where $U \subset \mathbb{R}^n$ is an open subset, let $D_x^2 g(x) \coloneqq \left(\frac{\partial^2 g}{\partial x_i \partial x_j}(x)\right)_{i,j=1,\dots n}$ be the Hessian matrix of g at $x \in U$. For a vector $x \in \mathbb{R}^n$, denote the Euclidean norm of x by $\|x\| \coloneqq \left(\sum_{i=1}^n x_i^2\right)^{1/2}$. For a set X and a subset $A \subset X$, let $\mathbf{1}_A : X \to \{0,1\}$ be the indicator function of A, i.e., $\mathbf{1}_A(x) = 1$ if $x \in A$, and $\mathbf{1}_A(x) = 0$ otherwise. Conventionally, we write $\mathbf{1}\{x \in A\} \coloneqq \mathbf{1}_A(x)$ if the context is clear. For a square matrix A, denote the smallest eigenvalue of A by $\lambda_{\min}(A)$. For two probability distributions P, Q, the product measure is denoted by $P \otimes Q$. For a topological space X and its subset $A \subset X$, let $\mathrm{Int}_X(A)$ be the interior of A in X. For a vector of random elements $(R_1, \dots, R_m)$, its joint distribution is denoted by $F_{R_1 \dots R_m}$. Since the products of random variables are not considered in this paper, this notation does not cause a confusion. For two random variables R_1, R_2, let $F_{R_1|R_2}$ be the conditional distribution of R_1 given R_2, whenever it is well-defined. For a measurable space (X, Σ), let $\mathcal{P}(X, \Sigma)$ be the set of all probability measures on (X, Σ). When the σ-field Σ is evident in the context, it is also written as $\mathcal{P}(X)$. For a Banach space $(\mathbb{D}, \|\cdot\|)$, let $\mathbb{D}^*$ denote its topological dual equipped with the dual norm, and let $\langle \cdot, \cdot \rangle$ denote the dual pairing between them. Although this notation does not explicitly indicate the underlying norm, no ambiguity arises because we will not introduce more than one norm for any given Banach space.

B Duality Theory and Approximation Scheme

Theorem 4.1 shows that the solution to the sample linear program (10) converges to that of the population problem (7). Proving this convergence directly is nontrivial. To address this, we analyze the dual formulations of the two problems. Establishing convergence for the dual problems turns out to be more tractable, and once this is obtained, the convergence of the original problem follows from the primal-dual relationship through complementary slackness.

In this section, we develop the duality framework that forms the basis for the proof of Theorem 4.1, which is provided in detail in Appendix C.7.

B.1 Approximate Problems

Let $(\mathcal{A}_f, \delta, K)$ be a conic restriction of interest that is defined in Section 3. Define a linear continuous operator $\bar{\mathcal{A}}_f : \mathbb{R}^{p\times J_n} \to \mathbb{D}$ as $\bar{\mathcal{A}}_f(\Lambda) := \mathcal{A}_f(\Lambda D_u m^{J_n}(\cdot))$. Then, the LP (10) is equivalent to

$$\begin{aligned} &\inf_{\substack{\psi\in\mathbb{R}^n \\ \Lambda\in\mathbb{R}^{p\times J_n}}} \frac{1}{n}\sum_{i=1}^{n}\psi_i \\ &\text{s.t.} \quad \begin{cases} uY_i \leq \psi_i + X_i'\Lambda m^{J_n}(u) \text{ for } (u,i)\in[0,1]\times\{1,\dots,n\} \\ \bar{\mathcal{A}}_f(\Lambda)+\delta\in K \end{cases}. \end{aligned} \tag{21}$$

Let $\mathbb{D}^*$ be the topological dual of $\mathbb{D}$, and let $\bar{\mathcal{A}}_f^* : \mathbb{D}^* \to \mathbb{R}^{J_n\times p}$ be the adjoint operator of $\bar{\mathcal{A}}_f$. Let K^* be the dual cone of K, that is,

$$K^* := \{k^*\in\mathbb{D}^* \mid \forall k\in K, \langle k^*, k\rangle \geq 0\},$$

where $\langle\cdot,\cdot\rangle$ is the dual pairing. The LP (21) is the dual problem of

$$\begin{aligned} &\sup_{\substack{\gamma\in\mathcal{P}([0,1]\times\mathcal{X}\times\mathcal{Y}) \\ \rho\in K^*}} \int uy d\gamma - \langle\rho,\delta\rangle \\ &\text{s.t.} \quad \begin{cases} \gamma_{23} = \frac{1}{n}\sum_{i=1}^{n}\delta_{(X_i,Y_i)} \\ \int m^{J_n}(u)x' d\gamma_{12}(u,x) + \bar{\mathcal{A}}_f^*\rho = 0 \end{cases}. \end{aligned} \tag{22}$$

[4]

Although we started with the LP (21) (or equivalently (10)), we call (22) and (21) the primal and dual, respectively, to make the duality relationship consistent with the one used in the literature (Carlier et al. (2016)).

Under the interior feasibility condition, (22) admits a solution, and the strong duality holds.

Lemma B.1. *Suppose that Assumptions NC, IF hold. Then, the primal problem (22) has a solution, and the strong duality holds, i.e, the value of (21) is equal to that of (22).*

A solution to the dual problem (21) is connected to a solution to the primal problem (22) through the relationship described in the following lemma. This relationship is commonly referred to as complementary slackness.

[4] When no shape restriction is involved, the LP (22) is reduced to equation (22) of Spady and Stouli (2018).

Lemma B.2. *Suppose that (ψ, Λ) is a solution to the dual (21) and (γ, ρ) is a solution to the primal (22), and that the strong duality holds. Let $\bar{\psi} \in C(\mathcal{X} \times \mathcal{Y})$ be such that $\bar{\psi}(X_i, Y_i) = \psi_i$ for $i = 1, \ldots, n$. Then,*

$$uy = \bar{\psi}(x, y) + x'\Lambda m^{J_n}(u) \quad \gamma\text{-a.s.}$$

and

$$\left\langle \rho, \bar{A}_f(\Lambda) + \delta \right\rangle = 0$$

hold.

A trivial but important observation is that when multiple restrictions are considered as in Example 6, the second part of the complementary slackness holds for each restriction, i.e.,

$$\left\langle \rho^{(i)}, \bar{A}_f^{(i)}(\Lambda) + \delta^{(i)} \right\rangle = 0$$

for $i = 1, \ldots, I$.

B.2 Exact Problems

As the limit of the approximate LP (21), we consider the following LP.

$$\begin{aligned} &\inf_{\substack{\psi \in C(\mathcal{X} \times \mathcal{Y}) \\ \sigma \in (C^{k+1}[0,1] \cap L_0^1[0,1])^p}} \int \psi dF_{XY} \\ &\text{s.t.} \quad \begin{cases} uy \leq \psi(x, y) + x'\sigma(u) \text{ for } (u, x, y) \in [0,1] \times \mathcal{X} \times \mathcal{Y} \\ \mathcal{A}_f(D\sigma(\cdot)) + \delta \in K \end{cases} \end{aligned} . \tag{23}$$

Observe that this LP is equivalent to the LP (7) with an additional shape restriction. Hence, if the restriction is correct, that is, if the solution to (7) satisfies the second constraint of (7), then it remains to solve (23).

Lemma B.3. *Suppose that Assumptions QR, QL, NC, SR hold. Then, the population dual problem (23) admits a unique solution, and it is given by the solution to (7).*

Next, we consider the limit of the approximate primal (22). Let $\mathcal{A}_f^* : \mathbb{D}^* \to (\mathbb{B}^{(k)})^*$ be the adjoint operator of $\mathcal{A}_f$. We consider the following LP.

$$\begin{aligned} &\sup_{\substack{\gamma \in \mathcal{P}([0,1] \times \mathcal{X} \times \mathcal{Y}) \\ \rho \in K^*}} \int uy d\gamma - \langle \rho, \delta \rangle \\ &\text{s.t.} \quad \begin{cases} \gamma_{23} = F_{XY} \\ \left(\int \mathbf{1}\{u \leq v\} x d\gamma_{12}(v, x) - (1 - u) \int x d\gamma_2(x) \right) du + \mathcal{A}_f^* \rho = 0 \end{cases} \end{aligned} . \tag{24}$$

Note that the second constraint should be interpreted as an equation in $(\mathbb{B}^{(k)})^*$. That is, it means that for any $b \in \mathbb{B}^{(k)}$,

$$\int_0^1 b(u)\left(\int \mathbf{1}\{u \leq v\} x d\gamma_{12}(v,x) - (1-u)\int x d\gamma_2(x)\right) du + (\mathcal{A}_f^* \rho)(b) = 0$$

holds.

If the shape restriction of interest is correct in population, the duality and complementary slackness holds between (23) and (24).

Lemma B.4. *Suppose that Assumptions QR, QL, LS, NC, SR hold. Then, the strong duality between (23) and (24) holds.*

Lemma B.5. *Suppose that (ψ, σ) is a solution to the dual (23) and (γ, ρ) is a solution to the primal (24), and that the strong duality holds. Then,*

$$uy = \psi(x,y) + x'\sigma(u) \quad \gamma\text{-}a.s.$$

and

$$\langle \rho, \mathcal{A}_f(D\sigma(\cdot)) + \delta \rangle = 0$$

hold.

Furthermore, the solution to (24) is pined down by the quasi-linear representation of F_{XY} as the following lemma shows.

Lemma B.6. *Suppose that Assumptions QR, QL, LS, NC, SR hold. Then, $(\gamma_*, 0)$ is a solution to the population primal problem (24), where $\gamma_* \in \mathcal{P}([0,1] \times \mathcal{X} \times \mathcal{Y})$ is the joint distribution of $(\bar{U}, X, Y)$ where $\bar{U} = (X'\beta(\cdot))^{-1}(Y)$. Moreover, if (γ, ρ) is a solution to (24), then $\gamma = \gamma_*$ holds.*

B.3 Convergence of Approximation Scheme

Now, we claim that the sample primal problem (22) converges to the population problem (24) in the following sense.

Lemma B.7. *Suppose that Assumptions QR, QL, LS, BS, NC, SR, IF, BB hold. Let (γ_n, ρ_n) be a solution the primal (22). If $J_n \to \infty$, then γ_n converges weakly to γ_* almost surely.*

As an easy corollary of Lemma B.7, we have the convergence of the dual value.

Corollary B.8. *Suppose that Assumptions QR, QL, LS, BS, NC, SR, IF, BB hold. The value of the sample dual problem (21) converges to that of the population problem (24) almost surely.*

Finally, we review the results shown in this section. They are summarized in Figure 13. The empirical dual (21) and the empirical primal (22) have solutions as shown in Lemma 3.3 and Lemma B.1, respectively, and there is duality between them (Lemmas B.1, B.2). The existence and uniqueness of solution to the population dual (23) and the population primal (24) are shown in Lemma B.3 and Lemma B.6, respectively. The duality between the population problems is given in Lemmas B.4, B.5. The convergence of empirical primal problems is proven by Lemma B.7. In Theorem 4.1, we prove the convergence of empirical dual solutions using the results stated in this section—see Appendix C.7 for details.

$$(21)\quad \inf_{\substack{\psi\in\mathbb{R}^n\\ \Lambda\in\mathbb{R}^{p\times J_n}}} \frac{1}{n}\sum_{i=1}^{n}\psi_i \qquad \text{s.t.} \begin{cases} uY_i \le \psi_i + X_i'\Lambda m^{J_n}(u) \\ \bar{\mathcal{A}}_f(\Lambda)+\delta\in K \end{cases}$$

($\exists$ sol. : Lemma 3.3)

$$\xleftrightarrow[\text{Lemma B.2}]{\text{Lemma B.1}}$$

$$(22)\quad \sup_{\substack{\gamma\in\mathcal{P}([0,1]\times\mathcal{X}\times\mathcal{Y})\\ \rho\in K^*}} \int uyd\gamma - \langle\rho,\delta\rangle \qquad \text{s.t.} \begin{cases} \gamma_{23} = \frac{1}{n}\sum_{i=1}^{n}\delta_{(X_i,Y_i)} \\ \int m^{J_n}(u)x'd\gamma + \bar{\mathcal{A}}_f^*\rho = 0 \end{cases}$$

($\exists$ sol. : Lemma B.1)

$$(21)\ \Big\downarrow\ \text{Theorem 4.1}\ (23) \qquad\qquad (22)\ \Big\downarrow\ \text{Lemma B.7}\ (24)$$

$$(23)\quad \inf_{\substack{\psi\in C(\mathcal{X}\times\mathcal{Y})\\ \sigma\in(C^{k+1}[0,1]\times L_0^1[0,1])^p}} \int \psi dF_{XY} \qquad \text{s.t.} \begin{cases} uy \le \psi(x,y) + x'\sigma(u) \\ \mathcal{A}_f(D\sigma)+\delta\in K \end{cases}$$

($\exists!$ sol. : Lemma B.3)

$$\xleftrightarrow[\text{Lemma B.5}]{\text{Lemma B.4}}$$

$$(24)\quad \sup_{\substack{\gamma\in\mathcal{P}([0,1]\times\mathcal{X}\times\mathcal{Y})\\ \rho\in K^*}} \int uyd\gamma - \langle\rho,\delta\rangle \qquad \text{s.t.} \begin{cases} \gamma_{23} = F_{XY} \\ \Big(\int \mathbf{1}\{u\le v\}xd\gamma_{12}(v,x) \\ \quad -(1-u)\int xd\gamma_2(x)\Big)du + \mathcal{A}_f^*\rho(du) = 0 \end{cases}$$

($\exists!$ sol. : Lemma B.6)

Figure 13: Duality and approximation stability.

C Proofs

C.1 Proof of Lemma 2.1

Let $\bar{U} \coloneqq \tilde{F}(g(U, |Z|))$. It is clear that $\bar{U} \sim U[0,1]$. We also have

$$Y = X'\tilde{\beta}(\tilde{F}^{-1}(U)) = X'\beta(\bar{U}).$$

Since $Z \stackrel{d}{=} -Z$ and $U \perp\!\!\!\perp Z$, we have $(Z, U) \stackrel{d}{=} (-Z, U)$, and therefore,

$$(Z, \bar{U}) = (Z, \tilde{F}(g(U, |Z|))) \stackrel{d}{=} (-Z, \tilde{F}(g(U, |-Z|))) = (-Z, \tilde{F}(g(U, |Z|))) = (-Z, \bar{U}).$$

Hence, we obtain $\mathbb{E}[Z \mid \bar{U}] = \mathbb{E}[-Z \mid \bar{U}]$, which implies $\mathbb{E}[Z \mid \bar{U}] = 0 = \mathbb{E}[Z]$. Since $x'\beta(\bar{u}) = (x'\tilde{\beta})(\tilde{F}^{-1}(\bar{u}))$ is strictly increasing in $\bar{u}$, Lemma 2.2 concludes that the joint distribution of (Y, X) satisfies Assumption QL.

C.2 Proof of Lemma 2.2

By Theorem 3.3 (i) of Carlier et al. (2016), Assumptions LS, QL imply that there exists a random variable $\bar{U} \sim U[0,1]$ on the same probability space as (Y, X) such that $Y = X'\beta(\bar{U})$ almost surely and $\mathbb{E}[X \mid \bar{U}] = \mathbb{E}[X]$. It is clear that $\bar{U} = (X'\beta(\cdot))^{-1}(Y)$ holds.

Conversely, suppose that there exist a continuous function $\bar{\beta} : (0,1) \to \mathbb{R}^p$ and a random variable $\bar{U} \sim U[0,1]$ defined on the same probability space as (Y, X) such that $\bar{u} \mapsto x'\bar{\beta}(\bar{u})$ is strictly increasing for all $x \in \mathcal{X}$,

$$Y = X'\bar{\beta}(\bar{U}), \quad \text{and} \quad \mathbb{E}[X \mid \bar{U}] = \mathbb{E}[X].$$

For each $u \in (0,1)$, observe

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}\{Y > X'\bar{\beta}(u)\}X\right] &= \mathbb{E}\left[\mathbf{1}\{X'\bar{\beta}(\bar{U}) > X'\bar{\beta}(u)\}X\right] \\
&= \mathbb{E}\left[\mathbf{1}\{\bar{U} > u\}X\right] \\
&= (1-u)\mathbb{E}[X].
\end{aligned}$$

By considering the first-order condition, it follows that $\beta = \bar{\beta}(u)$ is a solution to (4). Assumption QL holds by the properties of $\bar{\beta}$.

C.3 Proof of Lemma 2.3

By Theorem 3.3 (i) of Carlier et al. (2016), there exists a random variable $\bar{U} \sim U[0,1]$ on the same probability space as (X, Y) such that $Y = X'\beta(\bar{U})$ almost surely and $\mathbb{E}[X \mid \bar{U}] = \mathbb{E}[X]$. By Assumption QL, Theorem 3.2 of Carlier et al. (2016) implies that the pair $(\sigma_\beta, \psi_\beta)$ solves (7).

To show the latter statement, suppose that $\bar{\sigma} \in (C([0,1]) \cap L_0^1[0,1])^p$ solves the problem (7) and $u \mapsto x'\bar{\sigma}(u)$ is convex for each $x \in \mathcal{X}$. Let $\bar{\psi}(x,y) \coloneqq \sup_{u\in[0,1]}(uy - x'\bar{\sigma}(u))$.

Theorem 3.3 (i) of Carlier et al. (2016) implies that the LP

$$\begin{aligned} &\sup_{\gamma \in \mathcal{P}([0,1]\times\mathcal{X}\times\mathcal{Y})} \int uy d\gamma \\ &\text{s.t.} \quad \begin{cases} \gamma_{23} = F_{XY} \\ \left(\int \mathbf{1}\{u \le v\} x d\gamma_{12}(v,x) - (1-u)\int x d\gamma_2(x)\right) du = 0 \end{cases} . \end{aligned} \tag{25}$$

is the dual problem of (7), and a solution to (25) is given by the joint distribution $\bar{\gamma}$ of $(\bar{U}, X, Y)$, where $\bar{U} = (X'\beta(\cdot))^{-1}(Y)$. Furthermore, the strong duality holds, i.e., $\int \psi_\beta dF_{XY} = \int uy d\bar{\gamma}$. Since $(\bar{\psi}, \bar{\sigma})$ is also optimal, we have $\int \bar{\psi} dF_{XY} = \int uy d\bar{\gamma}$. By the complementary slackness, $uy = \bar{\psi}(x,y) + x'\bar{\sigma}(u)$ holds $\bar{\gamma}$-almost surely. By Fenchel's inequality and the convexity of $x'\bar{\sigma}(\cdot)$, $y \in \partial_u(x'\bar{\sigma})(u)$ holds $\bar{\gamma}$-almost surely.[5] Since $y = x'\beta(u)$ holds for $\bar{\gamma}$-almost all (u,x,y), we have

$$1 = \int \mathbf{1}\{x'\beta(u) \in \partial_u(x'\bar{\sigma})(u)\} d\bar{\gamma} = \int \left(\int \mathbf{1}\{x'\beta(u) \in \partial_u(x'\bar{\sigma})(u)\} d\bar{\gamma}_{1|2}(u \mid x) \right) dF_X(x),$$

which implies that there exist sets $A \subset \mathcal{X}$ and $B_x \subset [0,1]$, which is defined for each $x \in A$, such that $F_X(A) = 1$, $\bar{\gamma}_{1|2}(B_x \mid x) = 1$, and $x'\beta(u) \in \partial_u(x'\bar{\sigma})(u)$ holds for $u \in B_x$ and $x \in A$. Let $x \in A \cap \mathcal{X}_0$, where $A \cap \mathcal{X}_0$ is non-empty since $F_X(\mathcal{X}_0) > 0$ by Assumption LS. By Assumption QL and $\bar{\gamma}_{23} = F_{XY}$,

$$\bar{\gamma}_{1|2}([0,u] \mid x) = \int \mathbf{1}\{v \le u\} d\bar{\gamma}_{1|2}(v \mid x) = \int \mathbf{1}\{(x'\beta)^{-1}(y) \le u\} d\bar{\gamma}_{1|2}(y \mid x) = F_{Y|X}(x'\beta(u) \mid x)$$

holds for $u \in (0,1)$. Then, $u \mapsto \bar{\gamma}([0,u] \mid x)$ is strictly increasing on $(0,1)$ since for $0 < u < \tilde{u} < 1$,

$$\bar{\gamma}_{1|2}([0,\tilde{u}] \mid x) - \bar{\gamma}_{1|2}([0,u] \mid x) = \int_0^1 f_{Y|X}((1-t)x'\beta(u) + tx'\beta(\tilde{u}) \mid x) dt \cdot (x'\beta(\tilde{u}) - x'\beta(u)) > 0,$$

[5] The same observation also follows from Theorem 3.3 of Carlier et al. (2017).

which implies that $\text{supp}(\bar{\gamma}(du \mid x)) = [0,1]$. Since $\bar{B}_x \subset [0,1] = \text{supp}(\bar{\gamma}(du \mid x)) \subset \bar{B}_x$ by the definition of support, we have $\bar{B}_x = [0,1]$. Hence, $x'\beta(u) \in \partial_u(x'\bar{\sigma})(u)$ holds for u in a dense set of $(0,1)$. By the continuity of β, Theorem 24.4 of Rockafellar (1970) implies that $x'\beta(u) = D_u(x'\bar{\sigma})(u)$ holds for $u \in (0,1)$. Combining this with the fact that $x'\bar{\sigma}(\cdot)$ is differentiable almost everywhere on $(0,1)$ yields that $D_u(x'\sigma_\beta(\cdot)) = x'\beta(\cdot) = D_u(x'\bar{\sigma}(\cdot))$ holds almost everywhere on $(0,1)$. By Lemma 2.1 of del Barrio and Loubes (2019), we have $x'\sigma_\beta(u) = x'\bar{\sigma}(u)$ for $u \in (0,1)$, as $x'\sigma_\beta(\cdot), x'\bar{\sigma}(\cdot) \in L_0^1[0,1]$. Remember that this holds for all $x \in A \cap \mathcal{X}_0$. By Assumption FR, we obtain $\bar{\sigma} = \sigma_\beta$, which implies that $\bar{\sigma}(\cdot)$ is continuously differentiable and $D\bar{\sigma}(\cdot) = \beta(\cdot)$ on $(0,1)$.

C.4 Proof of Lemma 3.1

Let $(u,z) \in [0,1] \times \mathcal{Z}$. Then, we have

$$\begin{aligned}
(D_z Q_{Y|Z}(u \mid z))a(z) &= \mathbb{E}[D_z g(Z,\varepsilon) \mid Z = z, Y = Q_{Y|Z}(u \mid z)]a(z) \\
&= \mathbb{E}[(D_z g(Z,\varepsilon))a(Z) \mid Z = z, Y = Q_{Y|Z}(u \mid z)] \\
&\geq \mathbb{E}[b(Z) \mid Z = z, Y = Q_{Y|Z}(u \mid z)] \\
&= b(z),
\end{aligned}$$

where the first equality holds by Theorem 2.1 of Hoderlein and Mammen (2007).

C.5 Proof of Lemma 3.2

Let

$$E \coloneqq \{(z,y) \in [0,1] \times \mathbb{R} \mid y \geq \inf \text{supp}(F_{Y|Z=z})\}.$$

It is easy to see $(0,\eta), (1,\eta) \in E$ and $(1/2,\eta) \notin E$ for small $\eta > 0$. See Figure 14. For a sufficiently small quantile level $u > 0$, we have

$$Q_{Y|Z}(u \mid 0) < \eta, \quad Q_{Y|Z}(u \mid 1/2) > \eta, \quad Q_{Y|Z}(u \mid 1) < \eta,$$

which implies that the map $z \mapsto Q_{Y|Z}(u \mid z)$ is not convex.

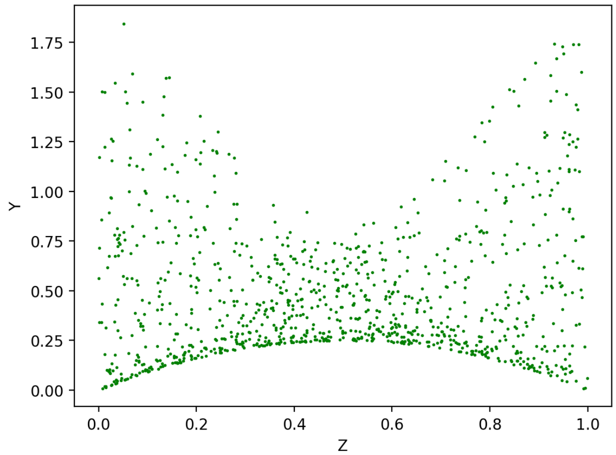

Figure 14: Scatter plot of an i.i.d. observation of (Z, Y).

C.6 Proof of Lemma 3.3

By Assumption FR, $\frac{1}{n}\sum_{i=1}^{n} X_i X_i'$ is full-rank almost surely. Lemma 3.3 follows by applying Lemma D.1 for $c(u,x,y) = uy$ and

$$\mathcal{L} = \left\{\Lambda \in \mathbb{R}^{p\times J_n} \mid \mathcal{A}_f(\Lambda D_u m^{J_n}(\cdot)) + \delta \in K\right\},$$

which is closed and non-empty.

By the proof of Lemma D.1, any solution lies in the set

$$\left\{\Lambda \in \mathbb{R}^{p\times J_n} \mid \inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\Lambda m^{J_n}(u) \geq -2C + \inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\bar{\Lambda} m^{J_n}(u)\right\}, \tag{26}$$

where $\bar{\Lambda}$ is a feasible solution.

C.7 Proof of Theorem 4.1

By Varadarajan's theorem (e.g., Theorem 11.4.1 of Dudley (2002)), the empirical measure $n^{-1}\sum_{i=1}^{n}\delta_{(X_i,Y_i)}$ converges weakly to F_{XY} almost surely. It suffices to show the stated uniform convergence for each realization of $\{(X_i,Y_i)\}_{i\in\mathbb{N}}$ along which the weak convergence occurs.

Fix $\varepsilon \in (0, 1/2)$. By (26) and Assumption CP, we observe

$$\hat{\Lambda}_n \in \left\{\Lambda \in \mathbb{R}^{p\times J_n} \mid -\inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\Lambda m^{J_n}(u) \leq 2C + M\right\} \tag{27}$$

for each n. By Assumption NC, $u \mapsto \inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\hat{\Lambda}_n m^{J_n}(u)$ is a smooth convex function on $[0,1]$ for each n and x. By Proposition D.4, we have

$$\sup_{n\in\mathbb{N},x\in\mathcal{X},u\in[\varepsilon,1-\varepsilon]}\left|x'\hat{\Lambda}_n m^{J_n}(u)\right|<\infty \quad \text{and} \quad \sup_{n\in\mathbb{N},x\in\mathcal{X},u\in[\varepsilon,1-\varepsilon]}\left|x'\hat{\Lambda}_n D_u m^{J_n}(u)\right|<\infty.$$

Since for $n\in\mathbb{N}$ and $u\in[\varepsilon,1-\varepsilon]$,

$$\sup_{n\in\mathbb{N},x\in\mathcal{X},u\in[\varepsilon,1-\varepsilon]}\left|x'\hat{\Lambda}_n m^{J_n}(u)\right|^2\geq\int_{\mathcal{X}}\left(x'\hat{\Lambda}_n m^{J_n}(u)\right)^2 dF_X(x)\geq\lambda_{\min}(\mathbb{E}[XX'])\left\|\hat{\Lambda}_n m^{J_n}(u)\right\|^2$$

holds, we have

$$\sup_{u\in[\varepsilon,1-\varepsilon]}\left\|\hat{\Lambda}_n m^{J_n}(u)\right\|\leq(\lambda_{\min}(\mathbb{E}[XX']))^{-1/2}\sup_{n\in\mathbb{N},x\in\mathcal{X},u\in[\varepsilon,1-\varepsilon]}\left|x'\hat{\Lambda}_n m^{J_n}(u)\right|<\infty$$

by Assumption FR. Similarly,

$$\sup_{u\in[\varepsilon,1-\varepsilon]}\left\|\hat{\Lambda}_n D m^{J_n}(u)\right\|<\infty$$

holds. By the local verision of the Arzelà–Ascoli theorem (e.g., Theorem 47.1 of Munkres (2000)), there exist a subsequence n' and a continuous function $\bar{\sigma}:(0,1)\to\mathbb{R}$ such that $\hat{\Lambda}_{n'}m^{J_{n'}}(\cdot)$ converges to $\bar{\sigma}$ locally uniformly on $(0,1)$. It is easy to see that $x'\bar{\sigma}(\cdot)$ is convex for each x. [6]

Let $\gamma_n,\gamma\in\mathcal{P}([0,1]\times\mathcal{X}\times\mathcal{Y})$ be solutions to (22) and (24), respectively. Notice that the existence is guaranteed by Lemmas B.1 and B.6. Fix $x\in\mathcal{X}_0$. Since the pair of functions

$$\sigma(u)=\int_0^u\beta(v)dv-\int_0^1\int_0^{\tilde{u}}\beta(v)dvd\tilde{u}\quad\text{and}\quad\psi(x,y)=\sup_{u\in[0,1]}(uy-x'\sigma(u))$$

is a solution to (7) by Lemma 2.3, the strong duality and complementary slackness (Lemmas B.5 and B.4) imply

$$uy=\psi(x,y)+x'\sigma(u)\text{ for }\gamma\text{-almost all }(u,x,y).$$

Since both $u\mapsto x'\sigma(u)$ and $y\mapsto\psi(x,y)$ are convex by Assumption NC and $D\sigma(\cdot)=\beta(\cdot)$,

[6] Probably, it is not possible to ensure $\int_0^1\bar{\sigma}=0$ at this point. Consider a convex function $f_n:[0,1]\to\mathbb{R}$ such that it takes values about -1 on a large portion of $(0,1)$ but grows near the boundary to satisfy $\int_0^1 f_n=0$. Let $g_n:[0,1]\to\mathbb{R}$ be a convex function that behaves similarly to f_n but takes values about -2 on the interior. The family of convex functions $\{f_n,g_n\}_{n\in\mathbb{N}}$ is uniformly bounded and equicontinuous on any compact subset of $(0,1)$. If f_n is chosen as a subsequence in the Arzelà-Ascoli theorem, the limit is $f\equiv-1$ while choosing g_n yields the limit $g\equiv-2$. Obviously, the zero mean condition is not satisfied in the limit, and moreover, the integral of the limit function depends on the choice of subsequence. It seems that there is no a priori regularity of $\hat{\Lambda}_n$ to avoid this phenomenon, and this is why we need to add constants to guarantee the convergence of the entire sequence of convex functions.

Fenchel's inequality implies

$$y = x'\beta(u) \text{ for } \gamma\text{-almost all } (u, x, y).$$

Now, let $u \in (0,1)$ and $y \in \mathcal{Y}$ be such that $(u, x, y) \in \operatorname{supp}(\gamma)$ and $y = x'\beta(u)$. Since $\gamma_{n'}$ weakly converges to γ by Lemma B.7, there exist $\delta > 0$ and $(u_{n'}, x_{n'}, y_{n'}) \in \operatorname{supp}(\gamma_{n'})$ such that $u_{n'} \in [\delta, 1-\delta]$, $y_{n'} = x'_{n'}\hat{\Lambda}_{n'} Dm^{J_{n'}}(u_{n'})$, for large n' and $(u_{n'}, x_{n'}, y_{n'}) \to (u, x, y)$, which follows from the strong duality and complementary slackness for approximate problems (Lemmas B.1 and B.2). It is clear that $x'_{n'}\hat{\Lambda}_{n'} m^{J_{n'}}(\cdot)$ converges to $x'\bar{\sigma}(\cdot)$ locally uniformly on $(0,1)$ by the construction of $\bar{\sigma}$. Define $f_{n'} : \mathbb{R} \to \bar{\mathbb{R}}$ as

$$f_{n'}(u) := \begin{cases} x'_{n'}\hat{\Lambda}_{n'} m^{J_{n'}}(u) & (u \in [0,1]) \\ +\infty & (u \notin [0,1]) \end{cases}.$$

Also, define $\bar{f}_x, f_x : \mathbb{R} \to \bar{\mathbb{R}}$ as

$$\bar{f}_x(u) := \begin{cases} x'\bar{\sigma}(u) & (u \in (0,1)) \\ +\infty & (u \notin (0,1)) \end{cases} \quad \text{and} \quad f_x(u) := \begin{cases} \bar{f}_x^{**}(u) & (u \in [0,1]) \\ +\infty & (u \notin [0,1]) \end{cases},$$

where $\bar{f}_x^{**}$ is the convex envelope of $\bar{f}_x$. Then, both $f_{n'}$ and f_x are proper lsc convex functions. Since $u \mapsto x'\bar{\sigma}(u)$ is convex on $(0,1)$, $f_x(u) = x'\bar{\sigma}(u)$ for $u \in (0,1)$. Hence, $f_{n'}$ converges to f_x locally uniformly on $(0,1)$. Also, since the effective domain of f_x contains $(0,1)$, $\partial f_{n'}$ converges to ∂f_x graphically by Lemma D.5. By construction, $y_{n'} = Df_{n'}(u_{n'})$ holds for large n'. Combining this with $(u_{n'}, y_{n'}) \to (u, y)$ yields $y \in \partial f_x(u)$. Recalling $y = x'\beta(u)$, we obtain $x'\beta(u) \in \partial f_x(u)$. Under Assumption LS, this holds for all $u \in (0,1)$. Since $u \mapsto \partial f_x(u)$ admits a continuous selection on $(0,1)$, the function $f_x(\cdot) = x'\bar{\sigma}(\cdot)$ is differentiable there. Hence, we obtain $x'\beta(\cdot) = D_u(x'\bar{\sigma}(\cdot))$ on $(0,1)$. Observing

$$x'\beta(u) = D_u\left(\int_0^u x'\beta(v)dv - \int_0^1 \int_0^{\tilde{u}} x'\beta(v)dvd\tilde{u}\right) \text{ for } u \in (0,1),$$

where the function in the parenthesis is convex in u by Assumption QL, Lemma 2.1 of del Barrio and Loubes (2019) implies that there exists a constant $C_{\bar{\sigma},x} \in \mathbb{R}$ such that

$$x'\bar{\sigma}(u) = \int_0^u x'\beta(v)dv + C_{\bar{\sigma},x}$$

holds for all $u \in (0,1)$, as $x'\bar{\sigma}(\cdot)$ is convex. By Assumption FR, we have

$$\bar{\sigma}(u) = \int_0^u \beta(v)dv + C_{\bar{\sigma}}$$

for some $C_{\bar{\sigma}} \in \mathbb{R}^p$. Notice that $\bar{\sigma}$ is the limit of a subsequence of $\hat{\Lambda}_n m^{J_n}(\cdot)$, but this equation reveals that the limit depends on the subsequence only through an additive constant, and hence, the entire sequence $\hat{\Lambda}_n m^{J_n}(\cdot) + C_n$, where

$$C_n \coloneqq \int_0^{1/2} \beta(v)dv - \hat{\Lambda}_n m^{J_n}(1/2),$$

converges to $\int_0^{\cdot} \beta(v)dv$ locally uniformly on $(0,1)$. In particular, $x'(\hat{\Lambda}_n m^{J_n}(\cdot) + C_n)$ converges to $x'\int_0^{\cdot} \beta(v)dv$ locally uniformly for each x. Theorem 25.7 of Rockafellar (1970) implies that $x'\hat{\Lambda}_n D_u m^{J_n}(\cdot) = x'\hat{\beta}_n(\cdot)$ converges to $x'\beta(\cdot)$ locally uniformly for each x. Let $K \subset (0,1)$ be a compact set. We observe

$$\begin{aligned}
\int_{\mathcal{X}} \sup_{u\in K} \left| x'\hat{\beta}_n(u) - x'\beta(u) \right|^2 dF_X(x) &= \int_{\mathcal{X}} \sup_{u\in K} \left((\hat{\beta}_n(u) - \beta(u))'xx'(\hat{\beta}_n(u) - \beta(u)) \right) dF_X(x) \\
&\geq \sup_{u\in K} \int_{\mathcal{X}} \left((\hat{\beta}_n(u) - \beta(u))'xx'(\hat{\beta}_n(u) - \beta(u)) \right) dF_X(x) \\
&\geq \lambda_{\min}(\mathbb{E}[XX']) \cdot \sup_{u\in K} \left\| \hat{\beta}_n(u) - \beta(u) \right\|^2.
\end{aligned}$$

Since for $\varepsilon > 0$ such that $K \subset [\varepsilon, 1-\varepsilon]$, we have

$$\sup_{u\in K} \left| x'\hat{\beta}_n(u) - x'\beta(u) \right|^2 \leq \sup_{x\in\mathcal{X}} \|x\| \cdot 2 \sup_{u\in[\varepsilon,1-\varepsilon]} \left(\left\| \hat{\Lambda}_n D_u m^{J_n}(u) \right\|^2 + \|\beta(u)\|^2 \right) < \infty,$$

Lebesgue's dominated convergence theorem implies

$$\lim_{n\to\infty} \int_{\mathcal{X}} \sup_{u\in K} \left| x'\hat{\beta}_n(u) - x'\beta(u) \right|^2 dF_X(x) = 0.$$

Assumption FR, which implies Assumption FR', concludes

$$\lim_{n\to\infty} \sup_{u\in K} \left\| \hat{\beta}_n(u) - \beta(u) \right\| = 0.$$

C.8 Proof of Theorem 4.2

Proof. Let $\lambda_* \coloneqq \mathrm{vec}(\Lambda_*')$ and $\hat{\lambda}_n \coloneqq \mathrm{vec}(\hat{\Lambda}_n')$.[7] We first show the asymptotic normality of $\hat{\lambda}_n$ by applying Example 3.2.12 of van der Vaart and Wellner (1996), which is summarized in Lemma D.6.

Without loss of generality, we may assume $\bar{J} = J_*$. Let

$$\Theta \coloneqq \mathrm{Int}\{\lambda \in \mathbb{R}^{pJ_*} \mid \bar{\mathcal{A}}_f(\lambda) + \delta \in K\}$$

where $\bar{\mathcal{A}}_f(\lambda) \coloneqq \mathcal{A}_f(\Lambda D_u m^{J_*}(\cdot))$ for $\lambda = \mathrm{vec}(\Lambda')$. Observe that $\lambda_* \in \Theta$ under Assumptions PL, QL, NC, SR. Let

$$g_\lambda(x,y) \coloneqq \sup_{u\in[0,1]} \left(uy - \lambda'(x \otimes m^{J_*}(u))\right).$$

For $\lambda_1, \lambda_2 \in \Theta$, we have

$$|g_{\lambda_1}(x,y) - g_{\lambda_2}(x,y)| \leq \left(\sup_{x\in\mathcal{X}} \|x\| \cdot \sup_{u\in[0,1]} \left\|m^{J_*}(u)\right\|\right) \cdot \|\lambda_1 - \lambda_2\|.$$

The problem (10) is equivalent to

$$\min_{\lambda\in\Theta} \frac{1}{n}\sum_{i=1}^n g_\lambda(X_i, Y_i),$$

and λ_* is a minimizer of $\lambda \mapsto \mathbb{E}[g_\lambda(X,Y)]$ by Assumption PL.

The consistency $\hat{\lambda}_n \to \lambda_*$ can be shown by a routine work using Corollary 3.2.3 (ii) of van der Vaart and Wellner (1996). The uniform tightness of $\hat{\lambda}_n$ follows by Assumption CP.

For $\lambda \in \Theta$ and $(x,y) \in \mathcal{X} \times \mathbb{R}$, let

$$u(\lambda; x, y) \coloneqq \operatorname*{argmax}_{u\in[0,1]} \left(uy - \lambda'(x \otimes m^{J_*}(u))\right).$$

Notice that this map is well-defined by Assumption QL. If $y \in \{\lambda'(x \otimes D_u m^{J_*}(u)) \mid u \in [0,1]\}$ is satisfied, then $u(\lambda; x, y)$ is characterized as the unique solution to $y = \lambda'(x \otimes D_u m^{J_*}(u))$. It also follows that

$$\begin{cases} y < \lambda'(x \otimes D_u m^{J_*}(0)) \Rightarrow u(\lambda; x, y) = 0 \\ y > \lambda'(x \otimes D_u m^{J_*}(1)) \Rightarrow u(\lambda; x, y) = 1 \end{cases}.$$

[7] For an $m \times n$ matrix $A = (a_1, \ldots, a_n)$ where $a_1, \ldots, a_n \in \mathbb{R}^m$, define $\mathrm{vec}(A) \coloneqq (a_1' \ldots a_n')'$.

By the implicit function theorem, the map $\lambda \mapsto u(\lambda; x, y)$ is continuously differentiable if

$$y \notin \{\lambda'(x \otimes D_u m^{J_*}(0)), \lambda'(x \otimes D_u m^{J_*}(1))\}. \tag{28}$$

It is easy to see

$$(D_\lambda u(\lambda; x, y))' = \begin{cases} -\frac{x \otimes D_u m^{J_*}(u(\lambda; x, y))}{\lambda'(x \otimes D_u^2 m^{J_*}(u(\lambda; x, y)))} & \text{if } y \in \left(\lambda'(x \otimes D_u m^{J_*}(0)), \lambda'(x \otimes D_u m^{J_*}(1))\right) \\ 0 & \text{if } y < \lambda'(x \otimes D_u m^{J_*}(0)) \text{ or } y > \lambda'(x \otimes D_u m^{J_*}(1)) \end{cases}$$

By the envelope theorem, we have

$$D_\lambda g_\lambda(x, y) = -x \otimes m^{J_*}(u(\lambda; x, y)).$$

holds. For each $\lambda \in \Theta$, if (x, y) satisfies (28), $\lambda \mapsto g_\lambda(x, y)$ is twice differentiable with

$$D_\lambda^2 g_\lambda(x, y) = -\left(x \otimes D_u m^{J_*}(u(\lambda; x, y))\right) D_\lambda u(\lambda; x, y).$$

Since

$$\left\| D_u m^{J_*}(u) - D_u m^{J_*}(v) \right\| \leq \left(\sum_{j=0}^{J_*} j^2 \right)^{1/2} |u - v|$$

we have

$$\begin{aligned} &\left\| \int_0^1 D_u m^{J_*}(tu(\lambda_* + h; x, y) + (1-t)u(\lambda_*; x, y))dt - D_u m^{J_*}(u(\lambda_*; x, y)) \right\| \\ &\qquad\qquad \leq \left(\sum_{j=0}^{J_*} j^2 \right)^{1/2} |u(\lambda_* + h; x, y) - u(\lambda_*; x, y)| \end{aligned}$$

By the equalities

$$\begin{aligned} 0 &= (\lambda_* + h)' \left(x \otimes D_u m^{J_*}(u(\lambda_* + h; x, y))\right) - \lambda_*' \left(x \otimes D_u m^{J_*}(u(\lambda_*; x, y))\right) \\ &= \lambda_*' \left(x \otimes \left(D_u m^{J_*}(u(\lambda_* + h; x, y)) - D_u m^{J_*}(u(\lambda_*; x, y))\right)\right) + h' \left(x \otimes D_u m^{J_*}(u(\lambda_*; x, y))\right) \\ &= \lambda_*' \left(x \otimes \left(\int_0^1 D_u^2 m^{J_*}(tu(\lambda_* + h; x, y) + (1-t)u(\lambda_*; x, y))dt\right)\right) (u(\lambda_* + h; x, y) - u(\lambda_*; x, y)) \\ &\qquad + h' \left(x \otimes D_u m^{J_*}(u(\lambda_*; x, y))\right), \end{aligned}$$

we have

$$u(\lambda_* + h; x, y) - u(\lambda_*; x, y)$$

$$= -h' \left(x \otimes D_u m^{J_*}(u(\lambda_*; x, y)) \right) \left(\lambda_*' \left(x \otimes \int_0^1 D_u^2 m^{J_*}(tu(\lambda_* + h; x, y) + (1-t)u(\lambda_*; x, y)) dt \right) \right)^{-1}.$$

Consequently, we obtain

$$\begin{aligned} & u(\lambda_* + h; x, y) - u(\lambda_*; x, y) - D_\lambda u(\lambda_*; x, y) h \\ = & -h' \left(x \otimes D_u m^{J_*}(u(\lambda_*; x, y)) \right) \\ \cdot & \left(\frac{1}{\lambda_*' \left(x \otimes \int_0^1 D_u^2 m^{J_*}(tu(\lambda_* + h; x, y) + (1-t)u(\lambda_*; x, y)) dt \right)} - \frac{1}{\lambda_*' \left(x \otimes D_u^2 m^{J_*}(u(\lambda_*; x, y)) \right)} \right). \end{aligned}$$

Hence, we obtain

$$|u(\lambda_* + h; x, y) - u(\lambda_*; x, y) - D_\lambda u(\lambda_*; x, y) h| \le \frac{2}{\varepsilon} \|h\| \left\| x \otimes D_u m^{J_*}(u(\lambda_*; x, y)) \right\|$$

by (13).

Combining these with

$$\begin{aligned} & D_\lambda g_{\lambda_* + h}(x, y) - D_\lambda g_{\lambda_*}(x, y) \\ = & -x \otimes \left(m^{J_*}(u(\lambda_* + h; x, y)) - m^{J_*}(u(\lambda_*; x, y)) \right) \\ = & - \left(x \otimes \int_0^1 D_u m^{J_*}(tu(\lambda_* + h; x, y) + (1-t)u(\lambda_*; x, y)) dt \right) (u(\lambda_* + h; x, y) - u(\lambda_*; x, y)) \end{aligned}$$

we find that, for small $\eta > 0$,

$$\sup_{\|h\| < \eta} \frac{1}{\|h\|} \left\| D_\lambda g_{\lambda_* + h}(x, y) - D_\lambda g_{\lambda_*}(x, y) - D_\lambda^2 g_{\lambda_*}(x, y) h \right\|$$

is bounded. Lemma D.7 implies

$$D_\lambda^2 \mathbb{E}[g_{\lambda_*}(X, Y)] = -\mathbb{E}\left[\left(X \otimes D_u m^{J_*}(\bar{U}) \right) D_\lambda u(\lambda_*; X, Y) \right]$$

where $Y = \lambda_*'(X \otimes D_u m^{J_*}(\bar{U}))$ holds almost surely by Lemma 2.2. Since

$$(D_\lambda u(\lambda_*; X, Y))' = -\frac{X \otimes D_u m^{J_*}(\bar{U})}{\lambda_*'(X \otimes D_u^2 m^{J_*}(\bar{U}))},$$

we obtain

$$D_\lambda^2 \mathbb{E}[g_{\lambda_*}(X, Y)] = \mathbb{E}\left[\frac{1}{\lambda_*'(X \otimes D_u^2 m^{J_*}(\bar{U}))} \left(X \otimes D_u m^{J_*}(\bar{U}) \right) \left(X \otimes D_u m^{J_*}(\bar{U}) \right)' \right]$$

$$= \mathbb{E}\left[\frac{1}{X' D_u \beta(\bar{U})}\left(X \otimes D_u m^{J_*}(\bar{U})\right)\left(X \otimes D_u m^{J_*}(\bar{U})\right)'\right]$$
$$= V.$$

To show that V is positive definite, suppose that $\lambda \in \mathbb{R}^{pJ_*}$ satisfies $\lambda' V \lambda = 0$. Since $X' D_u \beta(\bar{U}) > 0$ almost surely by Assumption QL, the fact that $\lambda' V \lambda = 0$ implies that $X' \Lambda D_u m^{J_*}(\bar{U}) = \lambda'(X \otimes D_u m^{J_*}(\bar{U})) = 0$ holds almost surely. Multiplying X from the left, taking the conditional expectation given $\bar{U}$, and multiplying the inverse of $\mathbb{E}[XX' \mid \bar{U}]$ from the left, we have $\Lambda D_u m^{J_*}(\bar{U}) = 0$ almost surely. Since the left-hand side is a vector of polynomials in $\bar{U}$, $\Lambda = 0$ is necessary. Hence, V is positive definite.

Fix (y, x) and let $u(\lambda) \coloneqq u(\lambda; x, y)$. We observe

$$\left|\frac{g_{\lambda_* + \delta h}(x, y) - g_{\lambda_*}(x, y)}{\delta} - D_\lambda g_{\lambda_*}(x, y) h\right| \leq \left|\frac{T_\delta(u(\lambda_* + \delta h)) - T_\delta(u(\lambda_*))}{\delta}\right|$$

where $T_\delta : [0, 1] \to \mathbb{R}$ is defined as

$$T_\delta(u) \coloneqq uy - (\lambda_* + \delta h)'(x \otimes m^{J_*}(u)).$$

Since

$$D_u T_\delta(u) = y - (\lambda_* + \delta h)'(x \otimes D_u m^{J_*}(u))$$

We have

$$\frac{T_\delta(u(\lambda_* + \delta h)) - T_\delta(u(\lambda_*))}{\delta}$$
$$= \int_0^1 \left(y - (\lambda_* + \delta h)'(x \otimes D_u m^{J_*}(tu(\lambda_* + \delta h) + (1 - t)u(\lambda_*)))\right) dt \cdot \frac{u(\lambda_* + \delta h) - u(\lambda_*)}{\delta}.$$

Observing

$$\begin{aligned}\frac{u(\lambda_* + \delta h) - u(\lambda_*)}{\delta} &= \left(\int_0^1 D_\lambda u(t(\lambda_* + \delta h) + (1 - t)\lambda_*) dt\right) h \\ &= -\left(\int_0^1 \frac{x \otimes D_u m^{J_*}(u(\lambda_* + t\delta h))}{(\lambda_* + t\delta h)'(x \otimes D_u^2 m^{J_*}(u(\lambda_* + t\delta h)))} dt\right) h\end{aligned}$$

and

$$\int_0^1 \left(y - (\lambda_* + \delta h)'(x \otimes D_u m^{J_*}(tu(\lambda_* + \delta h) + (1 - t)u(\lambda_*)))\right) dt$$
$$= \int_0^1 \left(\lambda_*'(x \otimes D_u m^{J_*}(u(\lambda_*))) - (\lambda_* + \delta h)'(x \otimes D_u m^{J_*}(tu(\lambda_* + \delta h) + (1 - t)u(\lambda_*)))\right) dt$$

$$= \lambda_*' \left(x \otimes \int_0^1 D_u m^{J_*}(u(\lambda_*)) - D_u m^{J_*}(tu(\lambda_* + \delta h) + (1-t)u(\lambda_*))dt \right) + O(|\delta|),$$

showing the L^2 convergence of

$$\left| \frac{g_{\lambda_* + \delta h}(X,Y) - g_{\lambda_*}(X,Y)}{\delta} - D_\lambda g_{\lambda_*}(X,Y)h \right|$$

is a routine work.

Now, Lemma D.6 implies $\sqrt{n}(\hat{\lambda}_n - \lambda_*) \Rightarrow N(0, V^{-1}WV^{-1})$ with the linear representation

$$\sqrt{n}(\hat{\lambda}_n - \lambda_*) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V^{-1} \left(X_i \otimes m^{\bar{J}}((X_i'\beta(\cdot))^{-1}(Y_i)) \right) + o_p(1).$$

Since

$$\hat{\beta}_n(\cdot) = \hat{\Lambda}_n D_u m^{\bar{J}}(\cdot) = \left(I_p \otimes \left(D_u m^{\bar{J}}(\cdot) \right)' \right) \hat{\lambda}_n,$$

the asymptotic normality stated in the theorem follows by the delta method. □

C.9 Proof of Proposition 4.3

By Lemma B.2, the empirical distribution γ_n of $(\hat{U}_i, X_i, Y_i)$ solves (22). By Lemmas B.6, B.7, γ_n converges weakly to γ_*, the joint distribution of $(\bar{U}_i, X_i, Y_i)$, almost surely. Hence, we have

$$\begin{aligned} \hat{W} &= \int \left(x \otimes m^{\bar{J}}(u) \right) \left(x \otimes m^{\bar{J}}(u) \right)' d\gamma_n(u,x,y) \\ &\to \int \left(x \otimes m^{\bar{J}}(u) \right) \left(x \otimes m^{\bar{J}}(u) \right)' d\gamma_*(u,x,y) \\ &= W \end{aligned}$$

as $n \to \infty$ almost surely.

$$\begin{aligned} \hat{V} - V = \int \frac{1}{x' D_u \hat{\beta}_n(u)} &\left(x \otimes D_u m^{\bar{J}}(u) \right) \left(x \otimes D_u m^{\bar{J}}(u) \right)' d\gamma_n(u,x,y) \\ &- \int \frac{1}{x' D_u \beta(u)} \left(x \otimes D_u m^{\bar{J}}(u) \right) \left(x \otimes D_u m^{\bar{J}}(u) \right)' d\gamma_*(u,x,y) \\ = \int \left(\frac{1}{x' D_u \hat{\beta}_n(u)} - \frac{1}{x' D_u \beta(u)} \right) &\left(x \otimes D_u m^{\bar{J}}(u) \right) \left(x \otimes D_u m^{\bar{J}}(u) \right)' d\gamma_n(u,x,y) \\ &+ \int \frac{1}{x' D_u \beta(u)} \left(x \otimes D_u m^{\bar{J}}(u) \right) \left(x \otimes D_u m^{\bar{J}}(u) \right)' d(\gamma_n - \gamma_*)(u,x,y) \end{aligned}$$

$$=: I_1 + I_2.$$

Since $(\mathcal{A}_f^{(\varepsilon)}, \delta^{(\varepsilon)}, K^{(\varepsilon)})$ is imposed, we have a lower bound

$$\inf_{(u,x)\in[0,1]\times\mathcal{X}} x' D_u \hat{\beta}_n(u) \geq \varepsilon.$$

The integrand of I_1 is uniformly bounded and converges to zero, as we have

$$\begin{aligned}
&\left|\left(\frac{1}{x' D_u \hat{\beta}_n(u)} - \frac{1}{x' D_u \beta(u)}\right)\left(x \otimes D_u m^{\bar{J}}(u)\right)\left(x \otimes D_u m^{\bar{J}}(u)\right)'\right| \\
&\leq \frac{1}{\varepsilon^2} \cdot \left| x' \left(\Lambda_* - \hat{\Lambda}_n\right) D_u^2 m^{\bar{J}}(u)\right| \cdot \left|\left(x \otimes D_u m^{\bar{J}}(u)\right)\left(x \otimes D_u m^{\bar{J}}(u)\right)'\right| \\
&\leq C \left\|\Lambda_* - \hat{\Lambda}_n\right\| \\
&\to 0
\end{aligned}$$

where $C > 0$ is some positive constant independent of n, and therefore, $I_1 \to 0$ holds almost surely. Since γ_n weakly converges to γ_* and $[0,1]\times\mathcal{X}\times\mathcal{Y}$ is compact, we have $I_2 \to 0$ almost surely. Thus, we obtain $\hat{V} \to V$ almost surely.

C.10 Proof of Theorem 6.1

The uniform consistency of $\hat{\beta}_n^{(\varepsilon)}$ is an immediate consequence of Theorem 4.1. It is easy to check that for any compact set $K \subset (0,1)$,

$$\lim_{n\to\infty} \sup_{u\in K} \left|\hat{Q}_{B,n}^{(\varepsilon)}(u \mid x) - Q_B(u \mid x)\right| = 0 \quad \text{a.s.} \tag{29}$$

holds. Let $\underline{u} > 0$ be a small positive number. Since

$$Q_V(u \mid x) = Q_B(u \mid x) + \frac{u}{I-1} D_u Q_B(u \mid x)$$

and

$$\hat{Q}_{V,n}^{(\varepsilon)}(u \mid x) = \hat{Q}_{B,n}^{(\varepsilon)}(u \mid x) + \frac{u}{I-1} D_u \hat{Q}_{B,n}^{(\varepsilon)}(u \mid x)$$

hold, for each $u \in (0,1)$, we have

$$\hat{\varphi}_{V,n}^{(\varepsilon)}(u \mid x) := \int_{\underline{u}}^{u} \hat{Q}_{V,n}^{(\varepsilon)}(v \mid x) dv$$

$$
\begin{aligned}
&= \frac{I-2}{I-1}\int_{\underline{u}}^{u} \hat{Q}_{B,n}^{(\varepsilon)}(v \mid x)dv + \frac{1}{I-1}\left(u\hat{Q}_{B,n}^{(\varepsilon)}(u \mid x) - \underline{u}\hat{Q}_{B,n}^{(\varepsilon)}(\underline{u} \mid x)\right) \\
&\to \frac{I-2}{I-1}\int_{\underline{u}}^{u} Q_B(v \mid x)dv + \frac{1}{I-1}\left(uQ_B(u \mid x) - \underline{u}Q_B(\underline{u} \mid x)\right) \\
&= \int_{\underline{u}}^{u} Q_V(v \mid x)dv =: \varphi_V(u \mid x)
\end{aligned}
$$

almost surely by (29). Since $\hat{Q}_{V,n}(\cdot \mid x)$ is increasing by construction, $\hat{\varphi}_{V,n}^{(\varepsilon)}$ is convex on $(0,1)$. By Theorem 25.7 of Rockafellar (1970), $\hat{Q}_{V,n}^{(\varepsilon)}(\cdot \mid x) = D\hat{\varphi}_{V,n}^{(\varepsilon)}(\cdot \mid x)$ converges to $Q_V(\cdot \mid x) = D\varphi_V(\cdot \mid x)$ locally uniformly. The consistency of $\hat{\gamma}_n^{(\varepsilon)}$ holds from that of $\hat{Q}_{V,n}^{(\varepsilon)}(\cdot \mid x)$ in the same way as the last part of the proof of Theorem 4.1.

Since $\hat{\xi}_{B,n}^{(\varepsilon)} = \hat{Q}_{V,n}^{(\varepsilon)} \circ \hat{F}_{B,n}^{(\varepsilon)}$, it suffices to show the locally uniform convergence of $\hat{F}_{B,n}^{(\varepsilon)}$. Fix $b \in (\xi_B^{-1}(\underline{v} \mid x), \xi_B^{-1}(\overline{v} \mid x))$. Since $Q_B(F_B(b \mid x) \mid x) = b = \hat{Q}_{B,n}^{(\varepsilon)}(\hat{F}_{B,n}^{(\varepsilon)}(b \mid x) \mid x)$, we have

$$
\begin{aligned}
&\hat{Q}_{B,n}^{(\varepsilon)}(F_B(b \mid x) \mid x) - Q_B(F_B(b \mid x) \mid x) \\
&= \left(\hat{Q}_{B,n}^{(\varepsilon)}(F_B(b \mid x) \mid x) - \hat{Q}_{B,n}^{(\varepsilon)}(\hat{F}_{B,n}^{(\varepsilon)}(b \mid x) \mid x)\right) + \left(\hat{Q}_{B,n}^{(\varepsilon)}(\hat{F}_{B,n}^{(\varepsilon)}(b \mid x) \mid x) - Q_B(F_B(b \mid x) \mid x)\right) \\
&= \int_0^1 D\hat{Q}_{B,n}^{(\varepsilon)}(tF_B(b \mid x) + (1-t)\hat{F}_{B,n}^{(\varepsilon)}(b \mid x) \mid x)dt \cdot \left(F_B(b \mid x) - \hat{F}_{B,n}^{(\varepsilon)}(b \mid x)\right).
\end{aligned}
$$

Let $[\underline{b}', \overline{b}'] \subset (\xi_B^{-1}(\underline{v} \mid x), \xi_B^{-1}(\overline{v} \mid x))$ be a compact interval. Then, we get

$$
\begin{aligned}
\sup_{b \in [\underline{b}', \overline{b}']} \left|\hat{F}_{B,n}^{(\varepsilon)}(b \mid x) - F_B(b \mid x)\right| &= \sup_{b \in [\underline{b}', \overline{b}']} \left| \frac{\hat{Q}_{B,n}^{(\varepsilon)}(F_B(b \mid x) \mid x) - Q_B(F_B(b \mid x) \mid x)}{\int_0^1 D\hat{Q}_{B,n}^{(\varepsilon)}(tF_B(b \mid x) + (1-t)\hat{F}_{B,n}^{(\varepsilon)}(b \mid x) \mid x)dt} \right| \\
&\le \frac{1}{\varepsilon} \sup_{u \in [F_B(\underline{b}' \mid x), F_B(\overline{b}' \mid x)]} \left|\hat{Q}_{B,n}^{(\varepsilon)}(u \mid x) - Q_B(u \mid x)\right| \\
&\to 0.
\end{aligned}
$$

The convergence of $(\hat{\xi}_{B,n}^{(\varepsilon)})^{-1}$ can be shown in the same way.

C.11 Proof of Lemma B.1

For $i = 1, \ldots, n$, define $\mathcal{A}_i : \mathbb{R}^n \times \mathbb{R}^{p \times J_n} \to C[0,1]$, $b_i \in C[0,1]$, and $K_i \subset C[0,1]$ as

$$
\mathcal{A}_i(\psi, \Lambda)(u) := X_i' \Lambda m^{J_n}(u) + \psi_i, \quad b_i(u) := -uY_i,
$$

and

$$
K_i := \{k \in C[0,1] \mid k(u) \ge 0 \text{ for all } u \in [0,1]\}.
$$

Also, define $\tilde{\mathcal{A}}_f : \mathbb{R}^n \times \mathbb{R}^{p \times J_n} \to \mathbb{D}$ as $\tilde{\mathcal{A}}_f(\psi, \Lambda) \coloneqq \bar{\mathcal{A}}_f(\Lambda)$. Then, if we set $\mathcal{A} \coloneqq (\mathcal{A}_1, \ldots, \mathcal{A}_n, \tilde{\mathcal{A}}_f)$, $b \coloneqq (b_1, \ldots, b_n, \delta)$, and $\mathcal{K} \coloneqq K_1 \times \cdots \times K_n \times K$, the constraint of LP (21) is equivalent to $\mathcal{A}(\psi, \Lambda) + b \in \mathcal{K}$. By Assumption IF, there exists some (ψ, Λ) such that $\mathcal{A}(\psi, \Lambda) + b \in \mathrm{Int}_{(C[0,1])^p \times \mathbb{D}}(\mathcal{K})$. By Theorem 2.9 of Shapiro (2001), the existence of a solution to (22) and the strong duality hold.

C.12 Proof of Lemma B.2

By the first constraint of (21) and the first constraint of (22), $uy - \bar{\psi}(x, y) - x'\Lambda m^{J_n}(u) \leq 0$ holds for γ-almost all (u, x, y). Integrating both sides of the inequality with respect to γ yields

$$0 \geq \int uy - \bar{\psi}(x, y) - x'\Lambda m^{J_n}(u) d\gamma = \int uy d\gamma - \frac{1}{n}\sum_{i=1}^{n} \psi_i - \int x'\Lambda m^{J_n}(u) d\gamma.$$

The second constraint of (22) implies

$$-\int x'\Lambda m^{J_n}(u) d\gamma = \left\langle \bar{\mathcal{A}}_f^* \rho, \Lambda \right\rangle = \left\langle \rho, \bar{\mathcal{A}}_f(\Lambda) \right\rangle = \left\langle \rho, \bar{\mathcal{A}}_f(\Lambda) + \delta \right\rangle - \left\langle \rho, \delta \right\rangle \geq -\left\langle \rho, \delta \right\rangle,$$

where the last inequality holds by the second constraint of (21) and $\rho \in K^*$. Moreover, the strong duality implies

$$\frac{1}{n}\sum_{i=1}^{n} \psi_i = \int uy d\gamma - \left\langle \rho, \delta \right\rangle.$$

From these (in)equalities, we have

$$0 \geq \int uy - \bar{\psi}(x, y) - x'\Lambda m^{J_n}(u) d\gamma \geq \int uy d\gamma - \frac{1}{n}\sum_{i=1}^{n} \psi_i - \left\langle \rho, \delta \right\rangle = 0.$$

Since the integrand of the second term is negative, the fact that the first inequality is an equality implies

$$uy = \bar{\psi}(x, y) + x'\Lambda m^{J_n}(u) \quad \gamma\text{-a.s.}.$$

The fact that the second inequality is an equality implies

$$\left\langle \rho, \bar{\mathcal{A}}_f(\Lambda) + \delta \right\rangle = 0.$$

C.13 Proof of Lemma B.3

Recall from Lemma 2.3 that $(\psi_\beta, \sigma_\beta)$ is a solution to (7). By Assumptions QL, NC, SR, σ_β satisfies $\mathcal{A}_f(D\sigma_\beta(\cdot)) + \delta \in K$, which implies that the pair $(\psi_\beta, \sigma_\beta)$ is feasible to (23) and consequently, is a solution. The uniqueness holds by Lemma 2.3.

C.14 Proof of Lemma B.4

Let $\gamma_* \in \mathcal{P}([0,1] \times \mathcal{X} \times \mathcal{Y})$ be the joint distribution of $(\bar{U}, X, Y)$, where $\bar{U} = (X'\beta(\cdot))^{-1}(Y)$. Assumptions QL, LS and Theorem 3.3 (i) of Carlier et al. (2016) imply that γ_* is a solution to the LP (25). Since $(\gamma_*, 0)$ is feasible to (24), the value of (25) is less than or equal to that of (24). It is easy to see that the LP (23) is equivalent to

$$
\begin{aligned}
&\inf_{\substack{\psi \in C(\mathcal{X} \times \mathcal{Y}) \\ \bar{\beta} \in \mathbb{B}^{(k)}}} \int \psi dF_{XY} \\
&\text{s.t.} \quad \begin{cases} uy \leq \psi(x,y) + x'\sigma_{\bar{\beta}}(u) \text{ for } (u,x,y) \in [0,1] \times \mathcal{X} \times \mathcal{Y} \\ \mathcal{A}_f \bar{\beta} + \delta \in K \end{cases},
\end{aligned} \tag{30}
$$

where

$$
\sigma_{\bar{\beta}}(u) \coloneqq \int_0^u \bar{\beta}(v)dv - \int_0^1 \left(\int_0^{\tilde{u}} \bar{\beta}(v)dv \right) d\tilde{u}.
$$

By (2.4) of Shapiro (2001), (24) is the dual of (30). By the weak duality, the value of (24) is less than or equal to that of (30). Recalling the equivalence between (23) and (30), Lemma B.3 implies that the value of (30) is equal to that of (7). Since the strong duality holds between (7) and (25) by Theorem C.2 of Carlier et al. (2016), the value of (7) is equal to that of (25). These are summarized as

$$
(25) \leq (24) \leq (30) = (23) = (7) = (25).
$$

Therefore, all of the values are equal, and in particular, the value of (23) is equal to that of (24).

C.15 Proof of Lemma B.5

By the first constraint of (23), $uy - \psi(x,y) - x'\sigma(u) \leq 0$ holds for all $(u,x,y) \in [0,1] \times \mathcal{X} \times \mathcal{Y}$. Integrating both sides of the inequality with respect to γ yields

$$
0 \geq \int uy - \psi(x,y) - x'\sigma(u) d\gamma = \int uy d\gamma - \int \psi dF_{XY} - \int x'\sigma(u) d\gamma.
$$

We also observe

$$\int_0^1 \left(\int \mathbf{1}\{u \le v\} x d\gamma_{12}(v,x) \right)' D_u \sigma(u) du = \int x' \left(\sigma(v) - \sigma(0)\right) d\gamma_{12}(v,x)$$
$$= \int x'\sigma(v) d\gamma_{12}(v,x) - \left(\int x' d\gamma_2(x) \right) \sigma(0),$$

and

$$\int_0^1 \left((1-u) \int x d\gamma_2(x) \right)' D_u \sigma(u) du = \left(\int x d\gamma_2(x) \right)' \left(\int_0^1 (1-u) D_u \sigma(u) du \right)$$
$$= - \left(\int x d\gamma_2(x) \right)' \sigma(0),$$

where the last equality holds by $\sigma \in (L_0^1[0,1])^p$. Combining these with the second constraint of (24) yields

$$-\int x'\sigma(u) d\gamma = \left\langle \mathcal{A}_f^* \rho, D\sigma \right\rangle = \left\langle \rho, \mathcal{A}_f(D\sigma(\cdot)) \right\rangle = \left\langle \rho, \mathcal{A}_f(D\sigma(\cdot)) + \delta \right\rangle - \left\langle \rho, \delta \right\rangle \ge - \left\langle \rho, \delta \right\rangle,$$

where the last inequality holds by the second constraint of (24) and $\rho \in K^*$. Moreover, the strong duality implies

$$\int \psi dF_{XY} = \int uy d\gamma - \left\langle \rho, \delta \right\rangle.$$

From these (in)equalities, we have

$$0 \ge \int uy - \psi - x'\sigma(u) d\gamma \ge \int uy d\gamma - \int \psi dF_{XY} - \left\langle \rho, \delta \right\rangle = 0.$$

Since the integrand of the second term is negative, the fact that the first inequality is an equality implies

$$uy = \psi(x,y) + x'\sigma(u) \quad \gamma\text{-a.s.}.$$

The fact that the second inequality is an equality implies

$$\left\langle \rho, \mathcal{A}_f(D\sigma(\cdot)) + \delta \right\rangle = 0.$$

C.16 Proof of Lemma B.6

By the first part of the proof of Lemma B.4, $(\gamma_*, 0)$ is a solution to (24). To show the uniqueness, suppose that (γ, ρ) is a solution to (24). Recall that $(\psi_\beta, \sigma_\beta)$ is a solution to

(23) by Lemma B.3. By Lemma B.5, it holds that

$$u \in \underset{\tilde{u}\in[0,1]}{\operatorname{argmax}} \left(\tilde{u}y - x'\sigma_\beta(\tilde{u})\right) \text{ for } \gamma\text{- and } \gamma_*\text{-almost all } (u,x,y).$$

Since $\tilde{u} \mapsto x'\sigma_\beta(\tilde{u})$ is strictly convex by Assumption QL, the argmax set is singleton for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$. Hence, $\gamma_{1|23}(\cdot \mid x,y) = \gamma_{*1|23}(\cdot \mid x,y)$ for all (x,y). Since $\gamma_{12} = F_{XY} = \gamma_{*12}$, we have $\gamma = \gamma_*$.

C.17 Proof of Lemma B.7

By the standard tightness argument and Prokhorov's theorem, γ_n converges weakly to $\gamma_\infty \in \mathcal{P}([0,1] \times \mathcal{X} \times \mathcal{Y})$ up to a subsequence. By Assumption BB, there exists $\varepsilon > 0$ such that

$$\operatorname{tr}\left(\Lambda_n^{\mathrm{BB}} \bar{\mathcal{A}}_f^* \rho_n\right) = \left\langle \rho_n, \bar{\mathcal{A}}_f \Lambda_n^{\mathrm{BB}} \right\rangle \geq \varepsilon \left\| \rho_n \right\|.$$

This inequality and the second constraint of (24) imply

$$-\operatorname{tr}\left(\Lambda_n^{\mathrm{BB}}\left(\int m^{J_n}(u)x' d\gamma_{n,12}(u,x)\right)\right) \geq \varepsilon \left\|\rho_n\right\| \geq 0.$$

By Assumptions BS, BB, the left-hand side is bounded uniformly over n. The Banach–Alaoglu theorem implies that ρ_n converges to $\rho_\infty \in \mathbb{D}^*$ with respect to the weak-$*$ topology up to a subsequence.

The second constraint of (24) is equivalent to the following condition.

$$\begin{gathered}\text{For any vector of polynomials } \varphi = (\varphi_1, \ldots, \varphi_p)' \text{ such that } \varphi_1, \ldots, \varphi_p \in L_0^1[0,1], \\ \int x'\varphi(u) d\gamma_{12}(u,x) + \langle \rho, \mathcal{A}_f(D\varphi) \rangle = 0.\end{gathered} \tag{31}$$

We shall show that the limit pair $(\gamma_\infty, \rho_\infty)$ is feasible to the LP (24) almost surely. To see this, observe that $\gamma_{n,12} = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i,Y_i)}$, which implies $\gamma_{\infty,12} = F_{XY}$ almost surely. Next, to consider condition (31), let $\varphi = (\varphi_1, \ldots, \varphi_p)'$ be a p-dimensional vector of polynomials such that $\varphi_1, \ldots, \varphi_p \in L_0^1[0,1]$. Since $J_n \to \infty$, for sufficiently large n, there exists $\Phi \in \mathbb{R}^{p \times J_n}$ such that $\varphi(\cdot) = \Phi m^{J_n}(\cdot)$. By the second constraint of (22), we have

$$\begin{aligned}\langle \rho_\infty, \mathcal{A}_f(D\varphi) \rangle &= \left\langle \rho_\infty, \bar{\mathcal{A}}_f(\Phi) \right\rangle \\ &= \lim_{n\to\infty} \left\langle \rho_n, \bar{\mathcal{A}}_f(\Phi) \right\rangle \\ &= \lim_{n\to\infty} \left\langle \bar{\mathcal{A}}_f^* \rho_n, \Phi \right\rangle\end{aligned}$$

$$
\begin{aligned}
&= -\lim_{n\to\infty} \operatorname{tr}\left(\Phi \int m^{J_n}(u)x' d\gamma_{n,12}(u,x)\right) \\
&= -\lim_{n\to\infty} \int x'\varphi(u) d\gamma_{n,12}(u,x) \\
&= -\int x'\varphi(u) d\gamma_{\infty,12}(u,x),
\end{aligned}
$$

which implies that condition (31) holds, and hence, $(\gamma_\infty, \rho_\infty)$ is feasible to the LP (24) almost surely.

Let D_n be the value of (21) and D_* be that of (7). By Lemma C.1 below, we have $\liminf_{n\to\infty} D_n \geq D_*$ almost surely. Now, we have

$$
\begin{aligned}
\int uy d\gamma_\infty - \langle \rho_\infty, \delta \rangle &= \liminf_{n\to\infty} \int uy d\gamma_n - \langle \rho_n, \delta \rangle \\
&= \liminf_{n\to\infty} D_n \\
&\geq D_* \\
&= \int \psi_\beta dF_{XY} \\
&= \int uy d\gamma_*,
\end{aligned}
$$

where the first equality holds since (γ_n, ρ_n) converges to $(\gamma_\infty, \rho_\infty)$ up to a subsequence, the second holds by Lemma B.1, and the fourth holds by Lemma B.4. Since $(\gamma_\infty, \rho_\infty)$ is feasible to (24), the above inequality implies that the pair is optimal. By Lemma B.6, we have $\gamma_\infty = \gamma_*$. As the limit is independent of the choice of subsequence, the weak convergence of γ_n to $\gamma_\infty = \gamma_*$ occurs almost surely without taking a subsequence.

Lemma C.1. *Let D_n be the value of (21) and D_* be that of (7). Under the same setup as Lemma B.7,* $\liminf_{n\to\infty} D_n \geq D_*$ *holds almost surely.*

Proof. Let $\varepsilon > 0$. For each n, there exists Λ_n in the set (27) such that $\bar{\mathcal{A}}_f(\Lambda_n) + \delta \in K$ and

$$
\frac{1}{n}\sum_{i=1}^n \psi_n(X_i, Y_i) < D_n + \varepsilon
$$

where $\psi_n(x,y) \coloneqq \sup_{u\in[0,1]} \left(uy - x'\Lambda_n m^{J_n}(u)\right)$. Since $(\psi_n, \Lambda_n D_u m^{J_n}(\cdot))$ is feasible for (7), we have

$$
\int \psi_\beta dF_{XY} \leq \int \psi_n dF_{XY}.
$$

By the first part of the proof of Theorem 4.1, there exists a continuous function $\bar{\sigma} : (0,1) \to \mathbb{R}$ such that $\Lambda_n m^{J_n}(\cdot)$ converges to $\bar{\sigma}$ locally uniformly on $(0,1)$ up to a subsequence.

Observe

$$\psi_n(x,y) = \sup_{u\in(0,1)} \begin{pmatrix} y \\ x \end{pmatrix}' \begin{pmatrix} u \\ -\Lambda_n m^{J_n}(u) \end{pmatrix}.$$

Let

$$\bar{\psi}(x,y) \coloneqq \sup_{u\in(0,1)} (uy - x'\bar{\sigma}(u)) = \sup_{u\in(0,1)} \begin{pmatrix} y \\ x \end{pmatrix}' \begin{pmatrix} u \\ -\bar{\sigma}(u) \end{pmatrix}.$$

Lemma D.8 implies

$$\sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left|\psi_n(x,y) - \bar{\psi}(x,y)\right| \to 0$$

up to a subsequence. Since $\bar{\psi}$ is the limit of a uniformly convergent sequence of continuous functions ψ_n, it is continuous. The compactness of $\mathcal{X}\times\mathcal{Y}$ implies the boundedness of $\bar{\psi}$.

Then, we have

$$\begin{aligned}
\liminf_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^n \psi_n(X_i,Y_i)\right) - \int \psi_\beta dF_{XY} &= \liminf_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^n \psi_n(X_i,Y_i) - \int \psi_\beta dF_{XY}\right) \\
&\geq \liminf_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^n \psi_n(X_i,Y_i) - \int \psi_n dF_{XY}\right) \\
&= \liminf_{n\to\infty} \left(\frac{1}{n}\sum_{i=1}^n \bar{\psi}(X_i,Y_i) - \int \bar{\psi} dF_{XY}\right) \\
&= 0
\end{aligned}$$

by the dominated convergence theorem. Hence, we have $\liminf_{n\to\infty} D_n \geq D_*$. $\square$

D Useful Lemmas

Lemma D.1. *Let $\mathcal{L}\subset\mathbb{R}^{p\times J}$ be a nonempty closed set. Let $\mu\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ be a probability measure, where $\mathcal{X}\subset\mathbb{R}^p$ and $\mathcal{Y}\subset\mathbb{R}$ are bounded. Assume that $\int xx'd\nu$ is full-rank for some probability measure $\nu\in\mathcal{P}(\mathcal{X})$. Let $c:[0,1]\times\mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ be a continuous function. The LP*

$$\begin{aligned}
&\inf_{\substack{\psi\in L^1(F_{XY})\\ \Lambda\in\mathcal{L}}} \int \varphi d\mu \\
&\text{s.t.}\quad c(u,x,y) \leq \psi(x,y) + x'\Lambda m^J(u) \text{ for } (u,x,y)\in[0,1]\times\mathcal{X}\times\mathcal{Y}
\end{aligned} \tag{32}$$

is finite and attains the minimum.

Proof. For a $\Lambda\in\mathcal{L}$, the pair $(\psi(\cdot,\cdot):(x,y)\mapsto\sup_{u\in\mathcal{U}}(c(u,x,y)-x'\Lambda m^J(u)),\Lambda)$ is feasible and has weakly smaller value of the objective than any other feasible pair associated with the same Λ. Observe that for $x \in \mathcal{X}$, $\psi(x, \cdot)$ has the same modulus of continuity as c because for $y, \tilde{y} \in \mathcal{Y}$,

$$|\psi(x,y) - \psi(x,\tilde{y})| \leq \sup_{u \in \mathcal{U}} |c(u,x,y) - c(u,x,\tilde{y})| \leq \omega_c(|y - \tilde{y}|),$$

where $\omega_c : \mathbb{R}_+ \to \mathbb{R}_+$ is a modulus of continuity of c. Moreover, let $\tilde{\psi}(x,\tilde{y}) := \psi(x,\tilde{y}) - \min_{y \in \mathcal{Y}} \psi(x,y)$. Then ω_c is a modulus of continuity of $\tilde{\psi}(x,\cdot)$ as well. Hence, the stated LP has the same value as

$$\begin{aligned} &\inf_{\tilde{\psi} \in \mathcal{F}_c, \eta \in \mathbb{R}, \Lambda \in \mathcal{L}} \int \tilde{\psi} d\mu + \eta \\ \text{s.t.} \quad & c(u,x,y) \leq \tilde{\psi}(x,y) + \eta + x'\Lambda m^J(u) \text{ for } (u,x,y) \in [0,1] \times \mathcal{X} \times \mathcal{Y} \end{aligned} \tag{33}$$

where

$$\mathcal{F}_c := \left\{ \tilde{\psi} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \mid \text{For } x \in \mathcal{X}, \tilde{\psi}(x,\cdot) \text{ has a modulus of continuity } \omega_c \text{ and } \min_{y \in \mathcal{Y}} \tilde{\psi}(x,y) = 0 \right\}.$$

Observing that for a given pair $(\tilde{\psi}, \Lambda) \in \mathcal{F}_c \times \mathcal{L}$, the optimal η is given by

$$\eta = \sup_{(u,x,y) \in [0,1] \times \mathcal{X} \times \mathcal{Y}} \{c(u,x,y) - \tilde{\psi}(x,y) - x'\Lambda m^J(u)\},$$

the problem is further rewritten as

$$\inf_{\tilde{\psi} \in \mathcal{F}_c} \left[\int \tilde{\psi} d\mu + \inf_{\Lambda \in \mathcal{L}} \sup_{(u,x) \in [0,1] \times \mathcal{X}} \left\{ \sup_{y \in \mathcal{Y}} \Big(c(u,x,y) - \tilde{\psi}(x,y) \Big) - x'\Lambda m^J(u) \right\} \right].$$

Now, we fix some $\tilde{\psi} \in \mathcal{F}_c$ and consider the infimum of

$$V_c(\Lambda) := V_c(\Lambda; \tilde{\psi}) := \sup_{(u,x) \in [0,1] \times \mathcal{X}} \left\{ \sup_{y \in \mathcal{Y}} \Big(c(u,x,y) - \tilde{\psi}(x,y) \Big) - x'\Lambda m^J(u) \right\}$$

over $\Lambda \in \mathcal{L}$. Let $\bar{\Lambda} \in \mathcal{L} \neq \emptyset$. Observe that

$$V_c(\bar{\Lambda}) \leq C - \inf_{(u,x) \in [0,1] \times \mathcal{X}} x'\bar{\Lambda} m^J(u),$$

where

$$C := \max_{(u,x,y) \in [0,1] \times \mathcal{X} \times \mathcal{Y}} |c(u,x,y)| < \infty.$$

Consider the set

$$S := S_{p,J,C,\bar{\Lambda}} := \left\{ \Lambda \in \mathbb{R}^{p\times J} \mid \inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\Lambda m^J(u) \geq -2C + \inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\bar{\Lambda} m^J(u) \right\}.$$

If $\Lambda \notin S$, then it is suboptimal. Indeed, the inequality

$$\inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\Lambda m^J(u) < -2C + \inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\bar{\Lambda} m^J(u)$$

implies

$$V_c(\Lambda) \geq -C - \inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\Lambda m^J(u) > C - \inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\bar{\Lambda} m^J(u) \geq V_c(\bar{\Lambda}).$$

Hence, to find the infimum of $V_c(\Lambda)$ over $\Lambda \in \mathcal{L}$, we can focus on S.

We shall show that S is compact. Clearly, the set is closed. Suppose for a contradiction that S is unbounded. Since S is convex and $\bar{\Lambda} \in S$, there exists $V \in \mathbb{R}^{p\times J} \setminus \{0\}$ such that $\{\bar{\Lambda} + aV \mid a > 0\} \subset S$. The definition of S implies that for all $a > 0$ and $(u,x) \in [0,1] \times \mathcal{X}$,

$$x'\bar{\Lambda} m^J(u) + ax'Vm^J(u) \geq -2C + \inf_{(u,x)\in[0,1]\times\mathcal{X}} x'\bar{\Lambda} m^J(u),$$

and therefore, for all $a > 0$, we have

$$a \inf_{(u,x)\in[0,1]\times\mathcal{X}} x'Vm^J(u) \geq -2C - 2\left\|\bar{\Lambda}\right\| \cdot \sup_{u\in\mathcal{U}} \left\|m^J(u)\right\| \cdot \sup_{x\in\mathcal{X}} \|x\| > -\infty,$$

which implies $\inf_{(u,x)\in[0,1]\times\mathcal{X}} x'Vm^J(u) \geq 0$ by considering $a \nearrow \infty$. Since $\int_0^1 x'Vm^J(u)du = 0$ by the definition of m^J, we have $x'Vm^J(u) = 0$ for $(u,x) \in [0,1] \times \mathcal{X}$. Since nontrivial polynomials do not coincide with zero on an open interval, $V'x = 0$ holds for each $x \in \mathcal{X}$. By the invertibility of $\int xx'd\nu$, we have $V = 0$, but this contradicts the choice of V. Thus, S is bounded, and consequently, compact.

Observe that $S = S_{p,J,C,\bar{\Lambda}}$ does not depend on the choice of $\tilde{\psi} \in \mathcal{F}_c$. Hence, the original LP is transformed as follows:

$$\begin{aligned} & \inf_{\substack{\psi\in L^1(F_{XY})\\ \Lambda\in\mathcal{L}}} \int \psi d\mu \\ & \text{s.t.} \quad c(u,x,y) \leq \psi(x,y) + x'\Lambda m^J(u) \text{ for } (u,x,y) \in [0,1]\times\mathcal{X}\times\mathcal{Y} \\ = \quad & \inf_{\tilde{\psi}\in\mathcal{F}_c, \eta\in\mathbb{R}, \Lambda\in\mathcal{L}} \int \tilde{\psi} d\mu + \eta \\ & \text{s.t.} \quad c(u,x,y) \leq \tilde{\psi}(x,y) + \eta + x'\Lambda m^J(u) \text{ for } (u,x,y) \in [0,1]\times\mathcal{X}\times\mathcal{Y} \end{aligned}$$

$$
\begin{aligned}
&= \inf_{\tilde{\psi}\in\mathcal{F}_c}\left[\int \tilde{\psi}d\mu + \inf_{\Lambda\in\mathcal{L}} V_c(\Lambda;\tilde{\psi})\right] \\
&= \inf_{\tilde{\psi}\in\mathcal{F}_c}\left[\int \tilde{\psi}d\mu + \inf_{\Lambda\in\mathcal{L}\cap S} V_c(\Lambda;\tilde{\psi})\right] \\
&= \begin{array}{cc} & \displaystyle\inf_{\tilde{\psi}\in\mathcal{F}_c,\eta\in\mathbb{R},\Lambda\in\mathcal{L}\cap S} \int \tilde{\psi}d\mu + \eta \\ \text{s.t.} & c(u,x,y) \le \tilde{\psi}(x,y) + \eta + x'\Lambda m^J(u) \text{ for } (u,x,y)\in[0,1]\times\mathcal{X}\times\mathcal{Y} \end{array} \\
&= \begin{array}{cc} & \displaystyle\inf_{\substack{\psi\in L^1(F_{XY})\\ \Lambda\in\mathcal{L}\cap S}} \int \psi d\mu \\ \text{s.t.} & c(u,x,y) \le \psi(x,y) + x'\Lambda m^J(u) \text{ for } (u,x,y)\in[0,1]\times\mathcal{X}\times\mathcal{Y} \end{array} \\
&= \inf_{\Lambda\in\mathcal{L}\cap S}\int \sup_{u\in[0,1]}\left\{c(u,x,y) - x'\Lambda m^J(u)\right\} d\mu.
\end{aligned}
$$

Since the objective function of the last expression is continuous in Λ and $\mathcal{L}\cap S$ is compact, it admits a minimizer $\Lambda^{\text{opt}}\in\mathcal{L}\cap S$. For this Λ^{opt}, define

$$
\psi^{\text{opt}}(x,y) \coloneqq \sup_{(u,x,y)\in[0,1]\times\mathcal{X}\times\mathcal{Y}} \left\{c(u,x,y) - x'\Lambda m^J(u)\right\}.
$$

Since it is easily seen that the pair $(\psi^{\text{opt}},\Lambda^{\text{opt}})$ is feasible to the original problem LP, the pair is a solution. $\square$

Lemma D.2. *For any $\varepsilon\in(0,1/2)$,*

$$
\max_{x\in[\varepsilon,1-\varepsilon]} f(x) \le \frac{1}{2\varepsilon}\int_0^1 |f(x)|dx
$$

holds for all convex function $f:[0,1]\to\mathbb{R}$.

Proof. Let $x_0\in[\varepsilon,1-\varepsilon]$ be such that $f(x_0)=\max_{x\in[\varepsilon,1-\varepsilon]} f(x)$. Then, we have

$$
f(x_0) \le \frac{1}{2\varepsilon}\int_{x_0-\varepsilon}^{x_0+\varepsilon} f(x)dx \le \frac{1}{2\varepsilon}\int_0^1 |f(x)|dx
$$

where the first inequality holds by the Hermite-Hadamard inequality. $\square$

Lemma D.3. *For all continuous functions $f:[0,1]\to\mathbb{R}$ such that $\int_0^1 f = 0$,*

$$
\int_0^1 |f(x)|dx \le 2\left(-\min_{x\in[0,1]} f(x)\right)
$$

holds.

Proof. Since $\int_0^1 f = 0$, we have $-\min_{x\in[0,1]} f(x) \geq 0$. Let $a \in (0,1)$ be the Lebesgue measure of the set $\{x \in [0,1] \mid f(x) \leq 0\}$. Then,

$$-\min_{x\in[0,1]} f(x) \geq a\left(-\min_{x\in[0,1]} f(x)\right) \geq \int_0^1 f_- = \frac{1}{2}\int_0^1 |f(x)|dx$$

where $f_-(x) \coloneqq -(f(x) \wedge 0)$. □

Proposition D.4. *Let $\varepsilon \in (0, 1/2)$. For $M > 0$, let*

$$\mathcal{F}_M \coloneqq \left\{f : [0,1] \to \mathbb{R} \mid f \textit{ is differentiable and convex}, \int_0^1 f = 0, -\min_{x\in[0,1]} f(x) \leq M\right\}.$$

Then,

$$\sup_{f\in\mathcal{F}_M} \sup_{x\in[\varepsilon,1-\varepsilon]} |f(x)| < \infty \textit{ and } \sup_{f\in\mathcal{F}_M} \sup_{x\in[\varepsilon,1-\varepsilon]} |f'(x)| < \infty$$

hold for any M. That is, $\mathcal{F}_M \mid_{[\varepsilon,1-\varepsilon]}$ is uniformly bounded and equicontinuous. In particular, $\mathcal{F}_M \mid_{[\varepsilon,1-\varepsilon]}$ is relatively compact in $C[\varepsilon, 1-\varepsilon]$.

Proof. Let $f \in \mathcal{F}_M$. It is clear from the definition of $\mathcal{F}_M$ that

$$-\min_{x\in[\varepsilon,1-\varepsilon]} f(x) \leq -\min_{x\in[0,1]} f(x) \leq M.$$

By Lemmas D.2, D.3, we also have

$$\max_{x\in[\varepsilon,1-\varepsilon]} f(x) \leq \frac{1}{2\varepsilon}\int_0^1 |f(x)|dx \leq \frac{1}{\varepsilon}\left(-\min_{x\in[0,1]} f(x)\right) \leq \frac{M}{\varepsilon}.$$

Combining these inequalities completes the proof of the uniform boundedness. The proposition holds since Theorem 10.6 of Rockafellar (1970) implies that $\mathcal{F}_M \mid_{[\varepsilon,1-\varepsilon]}$ is uniformly Lipschitz. □

Lemma D.5. *Let $f_n, f : \mathbb{R} \to \bar{\mathbb{R}}$ be proper lsc convex functions such that* $\mathrm{dom} f$ *has nonempty interior. If f_n converges to f locally uniformly on the interior of* $\mathrm{dom} f$, *then ∂f_n converges to ∂f graphically.*

Proof. By Theorem 7.17 of Rockafellar and Wets (2009), f_n epi-converges to f. Their Theorem 12.35 concludes the convergence of the subdifferentials. □

Lemma D.6 (Example 3.2.12 of van der Vaart and Wellner (1996))**.** *Let Θ be an open set of $\mathbb{R}^d$, and $\theta_0 \in \Theta$. Let $X_1, \ldots, X_n$ be i.i.d. random variables with common law P, and let*

$g_\theta : \mathcal{X} \to \mathbb{R}$ be a measurable function such that

$$|g_{\theta_1}(x) - g_{\theta_2}(x)| \leq \dot{g}(x) \|\theta_1 - \theta_2\|$$

holds for any $\theta_1, \theta_2 \in \Theta$ in a neighborhood of θ_0, where $\dot{g} \in L^2(P)$. Let $\hat{\theta}_n$ be a maximizer of $\theta \mapsto n^{-1} \sum_{i=1}^n g_\theta(X_i)$, and suppose $\hat{\theta}_n \to \theta_0$ in probability. Moreover, assume that the map $\theta \mapsto \int g_\theta dP$ is maximized at $\theta = \theta_0$ and twice continuously differentiable at $\theta = \theta_0$ with a non-singular second-derivative V. If there exists a measurable function $\dot{g}_{\theta_0} : \mathcal{X} \to \mathbb{R}^d$ such that

$$\int \left(\frac{g_{\theta_0 + \delta h} - g_{\theta_0}}{\delta} - h' \dot{g}_{\theta_0} \right)^2 dP \to 0$$

as $\delta \to 0$ for any $h \in \mathbb{R}^d$, then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges weakly to $N(0, V^{-1} W V^{-1})$, where $W := \int \dot{g}_{\theta_0} \dot{g}'_{\theta_0} dP$.

Lemma D.7. *Let (X, μ) be a measure space. Let $\Theta \subset \mathbb{R}^k$ be an open set and $\theta_0 \in \Theta$. Let $f : X \times \Theta \to \mathbb{R}$ be such that (i) $f(\cdot, \theta)$ is in $L^1(\mu)$ for all $\theta \in \Theta$ and (ii) $f(x, \cdot)$ is differentiable at θ_0 for μ-almost all x. Furthermore, assume that $\nabla_\theta f(\cdot, \theta_0) \in L^1(\mu)$, and that there exist $\varepsilon > 0$ and $g \in L^1(\mu)$ such that*

$$\sup_{\|h\| \leq \varepsilon} \frac{1}{\|h\|} |f(x, \theta_0 + h) - f(x, \theta_0) - h' \nabla_\theta f(x, \theta_0)| \leq g(x) \quad \text{for } \mu\text{-almost all } x.$$

Then, the function $F : \Theta \to \mathbb{R}$, $F(\theta) := \int_X f(x, \theta) d\mu$ is differentiable at θ_0 and the derivative is given by

$$\nabla_\theta F(\theta_0) = \int_X \nabla_\theta f(x, \theta_0) d\mu.$$

Proof. Since

$$\frac{1}{\|h\|} \left| F(\theta_0 + h) - F(\theta_0) - h' \int_X \nabla_\theta f(x, \theta_0) d\mu \right| \leq \int_X \frac{1}{\|h\|} |f(x, \theta_0 + h) - f(x, \theta_0) - h' \nabla_\theta f(x, \theta_0)| \, d\mu$$

By the differentiability of $f(x, \cdot)$ at θ_0, the integrand converges to zero μ-almost surely. By the uniformity condition, Lebesgue's dominated convergence theorem implies that the right-hand side converges to zero as $\|h\| \to 0$. $\square$

Lemma D.8. *Let $V \subset \mathbb{R}^d$ be a compact set. Suppose that $h_n, h : (0,1) \to \mathbb{R}^d$ are continuous functions such that h_n converges to h locally uniformly on $(0,1)$, and the maps $u \mapsto z' h_n(u)$ and $u \mapsto z' h(u)$ are both concave on $(0,1)$ for all $z \in V$. Let $f_n(z) := \sup_{u \in (0,1)} z' h_n(u)$ and $f(z) := \sup_{u \in (0,1)} z' h(u)$ for $z \in V$. Then, f_n converges to f uniformly on V.*

Proof. Fix $\varepsilon > 0$. For each $z \in V$, there exists $u_z \in (0,1)$ such that $f(z) < z'h(u_z) + \varepsilon$. Define $V_z := \{w \in V \mid f(w) < w'h(u_z) + \varepsilon\}$. Since f is continuous by Corollary 2a of Pinelis (2013), $(V_z)_{z\in V}$ is an open cover of V. By the compactness of V, there exist $z_1, \dots z_k \in V$ such that $V = \bigcup_{i=1}^k V_{z_i}$. Letting $\underline{u}_\varepsilon := (\min_{i=1,\dots,k} u_{z_i}) \wedge \varepsilon$ and $\overline{u}_\varepsilon := (\max_{i=1,\dots,k} u_{z_i}) \vee (1-\varepsilon)$, we have that for any $z \in V$,

$$f(z) < \max_{i=1,\dots,k} z'h(u_{z_i}) + \varepsilon \le \sup_{u\in[\underline{u}_\varepsilon,\overline{u}_\varepsilon]} z'h(u) + \varepsilon,$$

and therefore,

$$\sup_{z\in V}\left(f(z) - \sup_{u\in[\underline{u}_\varepsilon,\overline{u}_\varepsilon]} z'h(u)\right) \le \varepsilon \tag{34}$$

holds. We also observe that since h_n converges to h uniformly on $[\underline{u}_\varepsilon, \overline{u}_\varepsilon]$, there exists N_ε such that

$$n \ge N_\varepsilon \Rightarrow \sup_{z\in V} \sup_{u\in[\underline{u}_\varepsilon,\overline{u}_\varepsilon]} |z'h_n(u) - z'h(u)| < \varepsilon. \tag{35}$$

For $z \in V$, $u \in [\underline{u}_\varepsilon, \overline{u}_\varepsilon]$ and $n \ge N_\varepsilon$, we have

$$z'h_n(u) \le z'h(u) + \varepsilon \le f(z) + \varepsilon$$

by (35). For $z \in V$, $u \in (\overline{u}_\varepsilon, 1)$ and $n \ge N_\varepsilon$, we have

$$\begin{aligned}
z'h_n(u) &= z'h_n(\overline{u}_\varepsilon) + (u - \overline{u}_\varepsilon)\frac{z'h_n(u) - z'h_n(\overline{u}_\varepsilon)}{u - \overline{u}_\varepsilon} \\
&\le z'h_n(\overline{u}_\varepsilon) + (u - \overline{u}_\varepsilon)\frac{z'h_n(\overline{u}_\varepsilon) - z'h_n(1/2)}{\overline{u}_\varepsilon - 1/2} \\
&\le z'h(\overline{u}_\varepsilon) + \varepsilon + (1 - \overline{u}_\varepsilon)\left(\frac{z'h_n(\overline{u}_\varepsilon) - z'h_n(1/2)}{\overline{u}_\varepsilon - 1/2} \vee 0\right) \\
&\le f(z) + \varepsilon + (1 - \overline{u}_\varepsilon)\left(\frac{z'h_n(\overline{u}_\varepsilon) - z'h_n(1/2)}{\overline{u}_\varepsilon - 1/2} \vee 0\right)
\end{aligned}$$

by the concavity and (35). For $z \in V$, $u \in (0, \underline{u}_\varepsilon)$ and $n \ge N_\varepsilon$, we have

$$\begin{aligned}
z'h_n(u) &= z'h_n(\underline{u}_\varepsilon) + (u - \underline{u}_\varepsilon)\frac{z'h_n(u) - z'h_n(\underline{u}_\varepsilon)}{u - \underline{u}_\varepsilon} \\
&\le z'h_n(\underline{u}_\varepsilon) + (u - \underline{u}_\varepsilon)\frac{z'h_n(\underline{u}_\varepsilon) - z'h_n(1/2)}{\underline{u}_\varepsilon - 1/2} \\
&\le z'h(\underline{u}_\varepsilon) + \varepsilon - \underline{u}_\varepsilon\left(\frac{z'h_n(\underline{u}_\varepsilon) - z'h_n(1/2)}{\underline{u}_\varepsilon - 1/2} \wedge 0\right)
\end{aligned}$$

$$\leq f(z) + \varepsilon - \underline{u}_\varepsilon \left(\frac{z'h_n(\underline{u}_\varepsilon) - z'h_n(1/2)}{\underline{u}_\varepsilon - 1/2} \wedge 0 \right)$$

by the concavity and (35). For small $\Delta > 0$ and $n \geq N_\varepsilon$, we get

$$\begin{aligned}
\frac{z'h_n(\overline{u}_\varepsilon) - z'h_n(1/2)}{\overline{u}_\varepsilon - 1/2} &= \frac{z'h(\overline{u}_\varepsilon) - z'h(1/2)}{\overline{u}_\varepsilon - 1/2} + \frac{z'h_n(\overline{u}_\varepsilon) - z'h(\overline{u}_\varepsilon)}{\overline{u}_\varepsilon - 1/2} + \frac{z'h(1/2) - z'h_n(1/2)}{\overline{u}_\varepsilon - 1/2} \\
&\leq \frac{z'h(1/2+\Delta) - z'h(1/2)}{\Delta} + \frac{2\varepsilon}{\overline{u}_\varepsilon - 1/2}
\end{aligned}$$

by (35). Similarly, it holds that

$$\frac{z'h_n(\underline{u}_\varepsilon) - z'h_n(1/2)}{\underline{u}_\varepsilon - 1/2} \geq \frac{z'h(1/2+\Delta) - z'h(1/2)}{\Delta} - \frac{2\varepsilon}{1/2 - \underline{u}_\varepsilon}$$

Therefore, if $D_{1/2} \coloneqq \Delta^{1/2}(z'h(1/2+\Delta) - z'h(1/2))$, we obtain

$$\begin{aligned}
f_n(z) &= \sup_{u \in (0,1)} z'h_n(u) \\
&\leq f(z) + \varepsilon + \max\left(0, (1-\overline{u}_\varepsilon)\left(D_{1/2} + \frac{2\varepsilon}{\overline{u}_\varepsilon - 1/2}\right), -\underline{u}_\varepsilon \left(D_{1/2} - \frac{2\varepsilon}{1/2 - \underline{u}_\varepsilon}\right)\right)
\end{aligned}$$

for $n \geq N_\varepsilon$, which implies

$$\sup_{n \geq N_\varepsilon} \sup_{z \in V} (f_n(z) - f(z)) \leq \varepsilon + \max\left(0, (1-\overline{u}_\varepsilon)\left(D_{1/2} + \frac{2\varepsilon}{\overline{u}_\varepsilon - 1/2}\right), -\underline{u}_\varepsilon \left(D_{1/2} - \frac{2\varepsilon}{1/2 - \underline{u}_\varepsilon}\right)\right).$$

By letting $\varepsilon \to 0$, we have

$$\limsup_{n\to\infty} \sup_{z \in V} (f_n(z) - f(z)) \leq 0. \tag{36}$$

On other other hand, for $z \in V$ and $n \geq N_\varepsilon$, we have

$$\begin{aligned}
f_n(z) &\geq \sup_{u \in [\underline{u}_\varepsilon, \overline{u}_\varepsilon]} z'h_n(u) \\
&\geq \sup_{u \in [\underline{u}_\varepsilon, \overline{u}_\varepsilon]} z'h(u) - \varepsilon \\
&\geq f(z) - 2\varepsilon
\end{aligned}$$

by (34) and (35), which implies

$$\liminf_{n\to\infty} \inf_{z \in V} (f_n(z) - f(z)) \geq 0. \tag{37}$$

Combining (36) and (37) yields

$$\lim_{n\to\infty} \sup_{z\in V} |f_n(z) - f(z)| = 0.$$

□

E Computational Efficiency of $\hat{\beta}_n$ and $\hat{\beta}^{\mathbf{BRW}}$

In this section, we compare the proposed estimator $\hat{\beta}_n$ and the estimator $\hat{\beta}^{\mathrm{BRW}}$ proposed by Bondell et al. (2010) from the perspective of computation. First, we transform them into the non-redundant canonical form. Using the equality $\rho_u(\varepsilon) = \varepsilon_+ - (1-u)\varepsilon$, the convex problem (12) is rewritten as the following linear program:

$$\begin{aligned}
&\min_{\substack{(\beta_\ell^+)_{\ell=1,\dots,L}\in\mathbb{R}_+^{p\times L}\\ (\beta_\ell^-)_{\ell=1,\dots,L}\in\mathbb{R}_+^{p\times L}\\ (\gamma_{i\ell}^+)_{i=1,\dots,n,\ell=1,\dots,L}\in\mathbb{R}_+^{n\times L}\\ (\gamma_{i\ell}^-)_{i=1,\dots,n,\ell=1,\dots,L}\in\mathbb{R}_+^{n\times L}\\ (s_{x\ell})_{x,\ell}\in\mathbb{R}_+^{|\mathcal{X}_k|\times L}}} \sum_{\ell=1}^{L}\sum_{i=1}^{n}\left(u\gamma_{i\ell}^+ + (1-u)\gamma_{i\ell}^-\right) \\
&\text{s.t.} \quad \begin{cases} \gamma_{i\ell}^+ - \gamma_{i\ell}^- = Y_i - X_i'(\beta_\ell^+ - \beta_\ell^-) & \text{for} \quad (i,\ell)\in\{1,\dots,n\}\times\{1,\dots,L\} \\ x'(\beta_{\ell+1}^+ - \beta_{\ell+1}^-) = x'(\beta_\ell^+ - \beta_\ell^-) + s_{x\ell} & \text{for} \quad (x,\ell)\in\mathcal{X}_k\times\{1,\dots,L-1\} \end{cases}
\end{aligned} \quad . \tag{38}$$

The canonical form of the LP (11) with the non-crossing condition is

$$\begin{aligned}
&\min_{\substack{(\psi_i^+)_{i=1,\dots n}\in\mathbb{R}_+^n\\ (\psi_i^-)_{i=1,\dots n}\in\mathbb{R}_+^n\\ \Lambda^+\in\mathbb{R}_+^{p\times J_n}\\ \Lambda^-\in\mathbb{R}_+^{p\times J_n}\\ (s_{i\ell}^{(1)})_{i,\ell}\in\mathbb{R}_+^{n\times L}\\ (s_{x\ell}^{(2)})_{x,\ell}\in\mathbb{R}_+^{|\mathcal{X}_k|\times L}}} \sum_{i=1}^{n}\left(\psi_i^+ - \psi_i^-\right) \\
&\text{s.t.} \quad \begin{cases} u_\ell Y_i + s_{i\ell}^{(1)} = \psi_i^+ - \psi_i^- + X_i'(\Lambda^+ - \Lambda^-)m^{J_n}(u_\ell) & \text{for} \quad (i,\ell)\in\{1,\dots,n\}\times\{1,\dots,L\} \\ x'(\Lambda^+ - \Lambda^-)D_u^2 m^{J_n}(u_\ell) = s_{x\ell}^{(2)} & \text{for} \quad (x,\ell)\in\mathcal{X}_k\times\{1,\dots,L-1\} \end{cases}
\end{aligned} , \tag{39}$$

where the constraints are imposed for $u \in \{u_1,\dots,u_L\}$ and $x \in \mathcal{X}_k$ for a fair comparison.

While both problems have the same number of constraints, $C := (n + |\mathcal{X}_k|)L$, they have different number of variables. It is easy to check that (38) has $V_{\mathrm{BRW}} := (2p + 2n + |\mathcal{X}_k|)L$

variables whereas there are $V_{\text{SR}} := 2n + 2pJ_n + (n + |\mathcal{X}_k|)L$ variables in (39). The fact that

$$V_{\text{BRW}} - V_{\text{SR}} = n(L-2) + 2p(L - J_n)$$

implies that whenever the number L of quantiles is greater than the degree J_n of the polynomial basis, which is always true in our numerical studies and empirical applications below, our formulation has less variables than the approach based on the check function minimization does. The difference is not negligible especially when the sample size n is large.

Just because an LP involves fewer variables and constraints than another, it does not necessarily mean the former is easier to solve than the latter. LP solvers often exploits the sparsity of the coefficient matrix of LPs to improve the computational efficiency. We calculate the sparsity index of the coefficient matrix, defined as the number of zero entries divided by the number of all entries in the matrix, for each problem, although it does not characterize the computational efficiency perfectly. The index for the LP (38) is

$$S_{\text{BRW}} := 1 - \frac{\#\text{non-zeros}}{\#\text{variables} \times \#\text{constraints}} = 1 - \frac{2(1+p)nL + (4p+1)|\mathcal{X}_k|L}{V_{\text{BRW}}C},$$

and the index of (39) is

$$S_{\text{SR}} := 1 - \frac{(3 + pJ_n)nL + (pJ_n + 1)|\mathcal{X}_k|L}{V_{\text{SR}}C}.$$

Then, we have

$$\begin{aligned} \frac{1 - S_{\text{BRW}}}{1 - S_{\text{SR}}} &= \frac{V_{\text{SR}}}{V_{\text{BRW}}} \cdot \frac{2(1+p)nL + (4p+1)|\mathcal{X}_k|L}{(3 + pJ_n)nL + (pJ_n + 1)|\mathcal{X}_k|L} \\ &\asymp \frac{J_n + (n + |\mathcal{X}_k|)L}{(n + |\mathcal{X}_k|)L} \cdot \frac{(n + |\mathcal{X}_k|)L}{J_n(n + |\mathcal{X}_k|)L} \\ &= \frac{J_n + (n + |\mathcal{X}_k|)L}{J_n(n + |\mathcal{X}_k|)L}, \end{aligned}$$

where the approximation is up to constant scale with p fixed. In most relevant cases, this ratio is less than one, which means that the coefficient matrix of the LP (38) is more sparse than that of (39). In particular, when the degree J_n of the polynomial basis is large, it becomes close to zero, and solving the LP (39) to obtain our estimator is computationally hard. A good news is that our numerical simulations in Section 5 and Appendix G show that polynomials of degree at most 10 can approximate the underlying truth well. It turns out that common LP solvers can compute our estimator efficiently in the cases of $J_n \leq 10$.

It must be noted that the sparsity index is not a perfect measure, and it is possible that other special structures that could be exploited for computation can exist in the coefficient matrix. The construction of more efficient algorithms and the investigation of their computation complexity for the two estimators are left as future work.

F Validity of Assumptions IF, BB

We consider the validity of Assumptions IF, BB for cases where $x = f(z) = (1, z, \ldots, z^{p-1})'$ for $p \geq 3$ and $\mathcal{A}_f = (\mathcal{A}_f^{\mathrm{NC}}, \mathcal{A}_f^{\mathrm{LB}}, \mathcal{A}_f^{\uparrow}, \mathcal{A}_f^{\mathrm{CVX}}, \mathcal{A}_f^{\uparrow\xi})$. Similar restrictions, such as upper bounds, decreasingness, and concavity, can be considered in the same way. We focus on cases where Z is one-dimensional, but the analysis below can be extended to multidimensional covariates.

First, we consider Assumption IF. Suppose that $\mathrm{supp}(F_Z) \subset [-B, B]$ for some $B > 0$. Take $A_1, A_2 > 0$ such that $A_1 > A_2 B$ and $A_2 > 2B$. Let

$$\Lambda_n^{\mathrm{IF}} = \begin{pmatrix} A_1 & 1 & 0 & \ldots & 0 \\ A_2 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix} \in \mathbb{R}^{p \times J_n}.$$

Then, we first have

$$(\mathcal{A}_f^{\mathrm{NC}}(\Lambda_n^{\mathrm{IF}} D_u m^{J_n}(\cdot)))(u, x) = x' \Lambda_n^{\mathrm{IF}} D_u^2 m^{J_n}(u) = 1 > 0$$

for all $(u, x) \in [0, 1] \times \mathcal{X}$. For $(u, z, v) \in [0, 1] \times \mathcal{Z} \times \{\pm 1\}$, we have

$$\begin{aligned} (\mathcal{A}_f^{\mathrm{LB}}(\Lambda_n^{\mathrm{IF}} D_u m^{J_n}(\cdot)))(u, z) &= f(z)' \Lambda_n^{\mathrm{IF}} D_u m^{J_n}(u) \\ &= A_1 + A_2 z + z^2 + u \\ &\geq A_1 - A_2 B \\ &> 0, \\ (\mathcal{A}_f^{\uparrow}(\Lambda_n^{\mathrm{IF}} D_u m^{J_n}(\cdot)))(u, z) &= (D_z f(z))' \Lambda_n^{\mathrm{IF}} D_u m^{J_n}(u) \\ &= A_2 + 2z \\ &\geq A_2 - 2B \\ &> 0, \end{aligned}$$

and

$$(\mathcal{A}_f^{\mathrm{CVX}}(\Lambda_n^{\mathrm{IF}} D_u m^{J_n}(\cdot)))(u,z,v) = D_z^2 f(z)' \Lambda_n^{\mathrm{IF}} D_u m^{J_n}(u) = 2 > 0.$$

Moreover, it can be verified that

$$(\mathcal{A}_f^{\uparrow\xi}(\Lambda_n^{\mathrm{IF}} D_u m^{J_n}(\cdot)))(u,x) = \frac{I}{I-1} > 0.$$

These (in)equalities imply $\mathcal{A}_f(\Lambda_n^{\mathrm{IF}} D_u m^{J_n}(\cdot)) + \delta \in \mathrm{Int}_{\mathbb{D}}(K)$.

From the same analysis, it can be shown that Assumption BB is satisfied for $\Lambda_n^{\mathrm{BB}} = \Lambda_n^{\mathrm{IF}}$.

G Choice of J_n

In Section 3, we show that the growth rate of the approximation degree J_n does not affect the consistency of our estimator. Here, we examine how different choices of J_n influence its finite-sample performance.

We first consider the following DGP:

$$Y_i = X_i'(\alpha + \gamma\Phi^{-1}(U_i)), \quad X_i = (1, Z_i)', \quad (Z_i, U_i) \sim U[0,1] \otimes U[0,1]$$

for $i = 1, \ldots, n$, where Φ^{-1} is the quantile function of the standard Gaussian distribution, $\alpha = (1,1)'$ and $\gamma = (1,1)'$. In this DGP, we have

$$\beta(u) = \begin{pmatrix} 1 + \Phi^{-1}(u) \\ 1 + \Phi^{-1}(u) \end{pmatrix}.$$

We generate data from the DGP, estimate β using the proposed estimator with the non-crossing constraint for $J_n \in \{2, 4, \ldots, 14, 16\}$, and assess its performance using the mean-squared loss defined in (16). This procedure is repeated 1000 times.

Figure 15 presents boxplots of the loss distributions across different values of J_n and quantile levels. In the middle range of quantiles ($0.1 \leq u \leq 0.9$), the differences across J_n are negligible. For extreme quantiles, the estimator with $J_n = 2$ performs noticeably worse because it lacks sufficient flexibility to approximate the true function over the entire unit interval. For larger degrees (e.g., $J_n \geq 6$), the performance differences become minimal.

Next, we revisit the auction model discussed in Section 6 to examine how the choice of J_n influences the estimates. Figure 16 displays the estimated bid and valuation quantile regression coefficients for different values of J_n. While the estimators with $J_n = 2, 4, 6$ behave somewhat differently due to limited approximation capacity, the estimates become similar

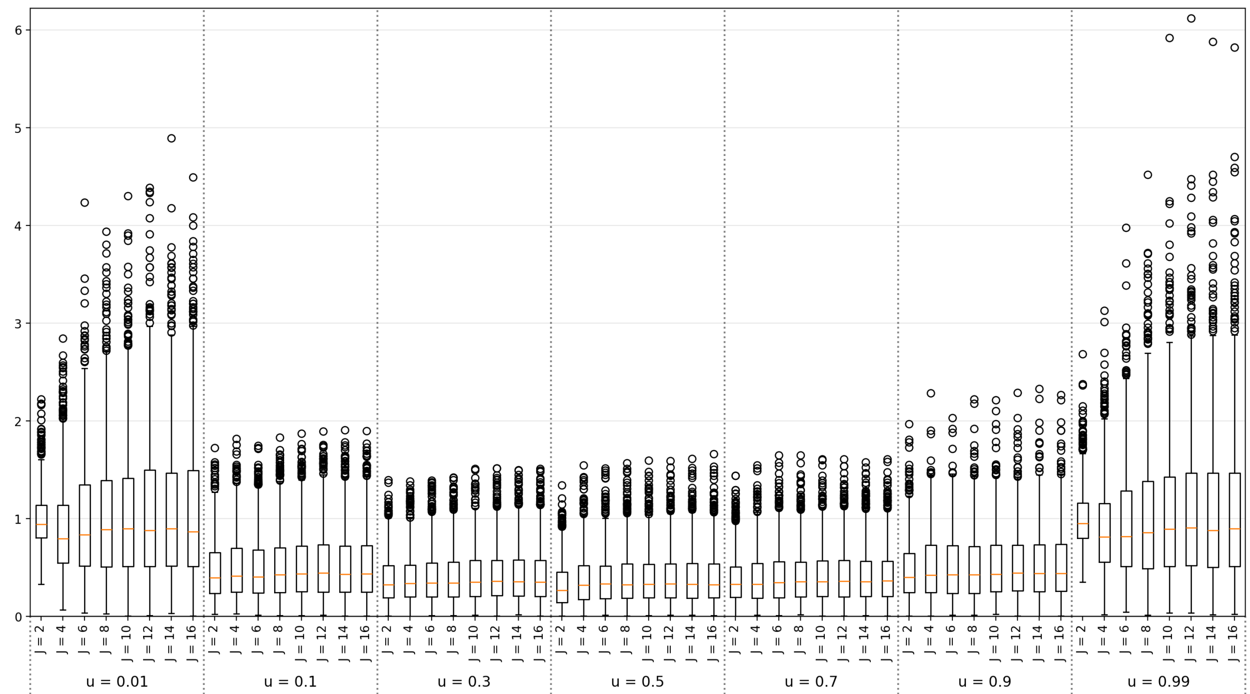

Figure 15: Boxplot of the loss values for the estimators with $J_n = 2, 4, \cdots, 14, 16$ at quantile levels $u = 0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99$. The horizontal axis represents different pairs of an estimator and a quantile level.

once J_n exceeds 6. This result indicates that the qualitative conclusions remain largely unaffected as long as J_n is sufficiently large to approximate the true quantile regression coefficient.

H Estimation of Production Functions

We apply the proposed method to a real dataset on production in Japan, using the 2020 Census of Manufacture provided by the Ministry of Economy, Trade and Industry. Our analysis focuses on establishments with thirty or more persons engaged in the production of textile products. Value added is used as the measure of output, the number of employees as labor, and the book value of tangible fixed assets at the end of the year as capital. The dataset provides these variables at the prefecture level. Let Y_i, L_i, and K_i denote output, labor, and capital, respectively. The sample consists of 43 observations; although Japan has 47 prefectures, data for the remaining four are unavailable because the number of target establishments there is insufficient for disclosure.

This analysis is motivated by Doty and Song (2021), who extend Ackerberg et al. (2015) to cases where ex-post production shocks enter the production function non-additively. They

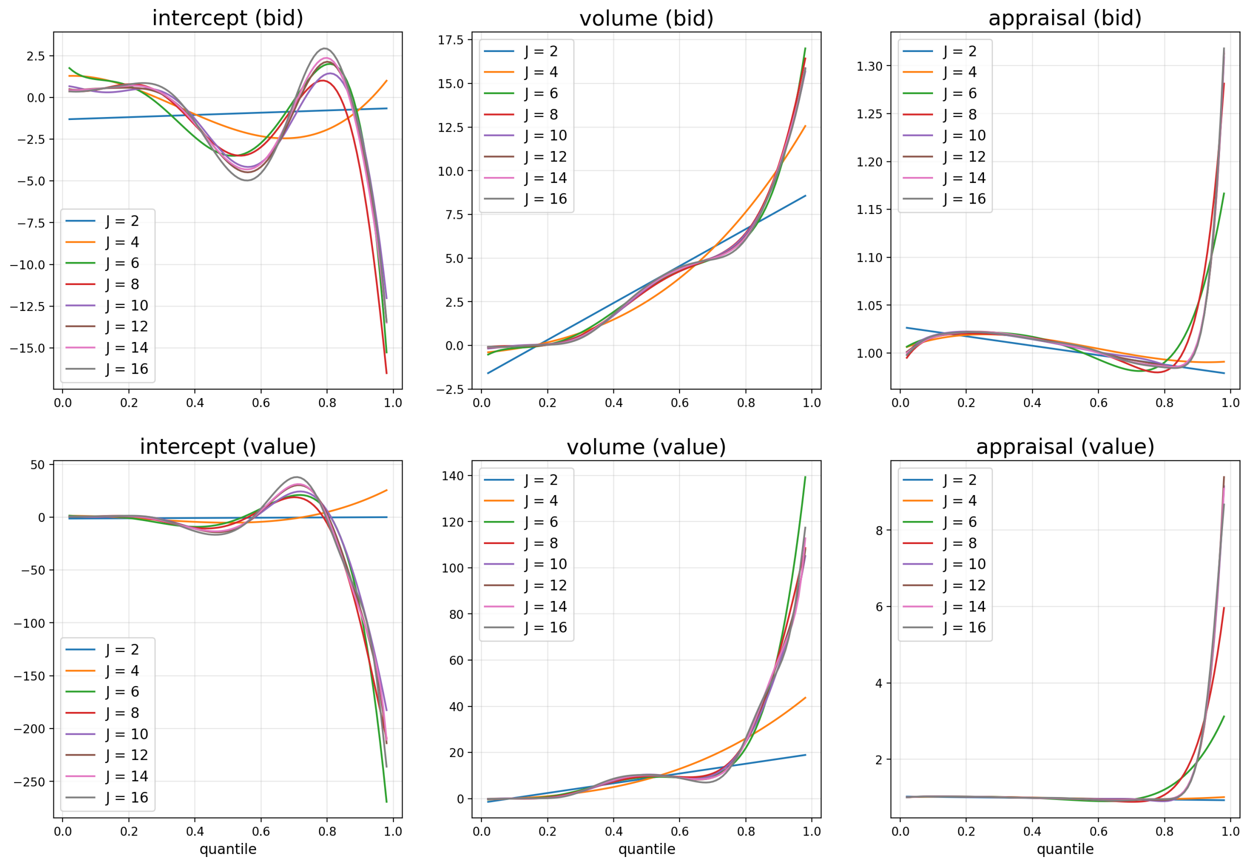

Figure 16: Bid (top) and valuation (bottom) quantile regression coefficient estimates. Estimates with $J_n = 2, 4, \cdots, 14, 16$ are shown.

develop an identification and estimation framework based on quantile regression, proposing a two-step procedure: first, they remove simultaneity bias using instruments as in Ackerberg et al. (2015); second, they estimate the quantile production function using the debiased outcomes. For simplicity, we focus on the second step, assuming that simultaneity bias is not severe in our application.

Cobb-Douglas specification. First, we consider the Cobb-Douglas production function with heterogeneous coefficients. The model is specified as follows:

$$\log Y_i = h_1(\log L_i, \log K_i, \bar{U}_i) := \beta_0(\bar{U}_i) + \beta_1(\bar{U}_i)\log L_i + \beta_2(\bar{U}_i)\log K_i,$$

where $\bar{U}_i \sim U[0,1]$ satisfies the mean independence condition:

$$\mathbb{E}[\log L_i \mid \bar{U}_i] = \mathbb{E}[\log L_i] \quad \text{and} \quad \mathbb{E}[\log K_i \mid \bar{U}_i] = \mathbb{E}[\log K_i].$$

To interpret $\bar{U}_i$ as a latent index of productivity, we assume that $u \mapsto \beta_0(u) + \beta_1(u)\log \ell +$

$\beta_2(u) \log k$ is strictly increasing. Under these assumptions, the quasi-linearity (Assumption QL) holds by Lemma 2.2, and hence, $\beta = (\beta_0, \beta_1, \beta_2)'$ is characterized as the solution to the check loss function (4). Notice also that the classical Cobb-Douglas production function (e.g., Zellner et al. (1966)), where β_1 and β_2 are constant and $\bar{U}_i$ is independent of the input variables, is a special case of this specification.

As a benchmark, we first estimate β using the estimator of Koenker and Bassett (1978). Figure 17 reports the contour plots of the estimated production functions $(\log \ell, \log k) \mapsto \beta_0(u) + \beta_1(u) \log \ell + \beta_2(u) \log k$ for the five quantile levels $u = 0.1, 0.3, 0.5, 0.7, 0.9$. The contour lines are shown only on a neighborhood of the region where the data are observed. The contour lines in the panel for $u = 0.1$ indicate that for a fixed labor input, an increase in capital input decreases output, which is highly counter-intuitive and makes little sense. Moreover, it can be inferred that the quantile crossing problem occurs by comparing the panels for $u = 0.1$ and $u = 0.3$. This is incompatible with the interpretation that $\bar{U}_i$ is an index for productivity.

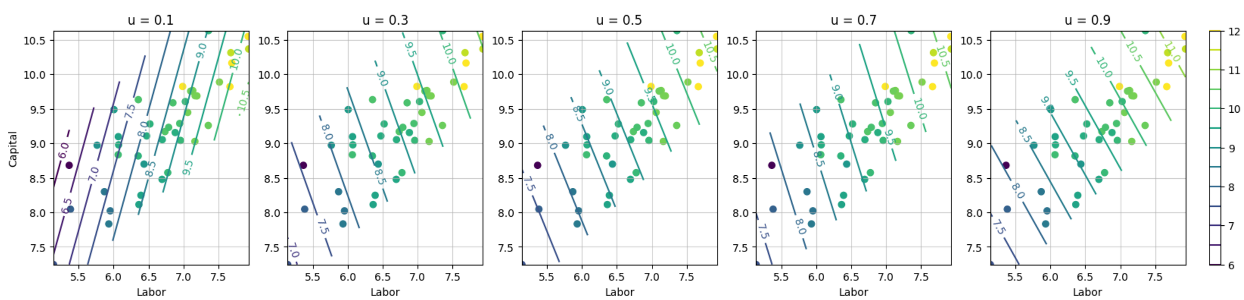

Figure 17: Koenker and Bassett (1978) under the Cobb-Douglas specification.

Next, we estimate β using the proposed estimator with the non-crossing condition and the monotonicity in the input variables. The same contour plots are provided in Figure 18. We can observe that increasing capital always increases output even in the panel for $u = 0.1$, unlike the estimates by Koenker and Bassett (1978).

More interestingly, our estimates have some implications for the output elasticities of the input variables, recalling that the output elasticities of labor and capital are given by

$$\frac{\partial h_1(\log \ell, \log k, u)}{\partial \log \ell} = \beta_1(u) \quad \text{and} \quad \frac{\partial h_1(\log \ell, \log k, u)}{\partial \log k} = \beta_2(u),$$

respectively. It is observed that $\hat{\beta}_{n,1}$ declines while $\hat{\beta}_{n,2}$ is rises. In other words, as the productivity increases, additional capital becomes relatively more beneficial and additional labor relatively less so.

Third-order translog specification. In order to model the elasticity of substitution

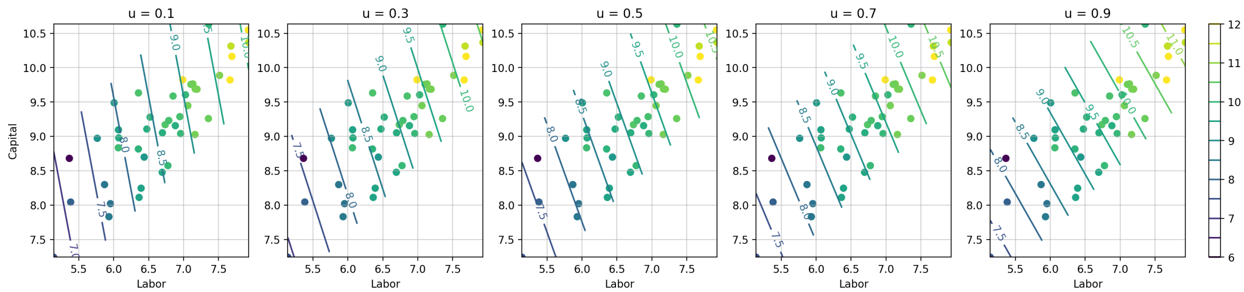

Figure 18: Our estimator under the Cobb-Douglas specification.

more flexibly, researchers often choose the translog production function instead of the Cobb-Douglas specification. See, for example, Kmenta (1967), León-Ledesma et al. (2010), and Gechert et al. (2022). Allowing for heterogeneous coefficients as the Cobb-Douglas specification above, we consider

$$\log Y_i = h_3(\log L_i, \log K_i, \bar{U}_i) \coloneqq \sum_{\substack{p,q \geq 0 \\ 0 \leq p+q \leq r}} \beta_{p,q}(\bar{U}_i)(\log L_i)^p (\log K_i)^q.$$

where $\bar{U}_i \sim U[0,1]$ satisfies the mean independence condition

$$\mathbb{E}\left[(\log L_i)^p (\log K_i)^q \mid \bar{U}_i\right] = \mathbb{E}\left[(\log L_i)^p (\log K_i)^q\right]$$

holds for all (p, q) such that $0 \leq p + q \leq r$. In the following analysis, we choose $r = 3$. As above, under the assumption that the map $u \mapsto h_3(\log \ell, \log k, u)$ is strictly increasing, the Assumption QL is satisfied.

Figure 19 shows the contour plots estimated using the method of Koenker and Bassett (1978). The non-monotonicity of the estimated production function with respect to the input variables is more pronounced than in the Cobb–Douglas specification presented above. Near the boundaries of, or outside, the data support, the estimates become highly unstable. As a result, it is difficult to draw meaningful or policy-relevant implications from these estimates.

Next, we consider the proposed estimator. Since it is reasonable to presume that the output is an increasing function of labor and capital, respectively, we impose the restriction

$$\text{the map } (\log \ell, \log k) \mapsto h_3(\log \ell, \log k, u) \text{ is increasing in both variables}$$

as well as the non-crossing condition on the estimator.

Figure 20 presents the contour plots obtained using the proposed estimator. In contrast

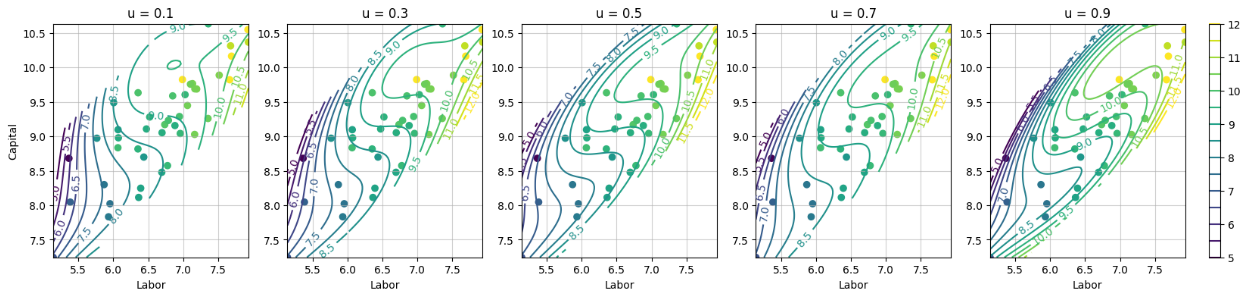

Figure 19: Koenker and Bassett (1978) under third-order translog specification.

to the results of Koenker and Bassett (1978), the estimated production function satisfies the monotonicity in both labor and capital inputs. Moreover, the nonlinear specification allows the substitution pattern to vary with the levels of the inputs. In particular, prefectures with low labor input tend to exhibit lower output elasticity of capital, indicating that capital is less productive when labor is scarce.

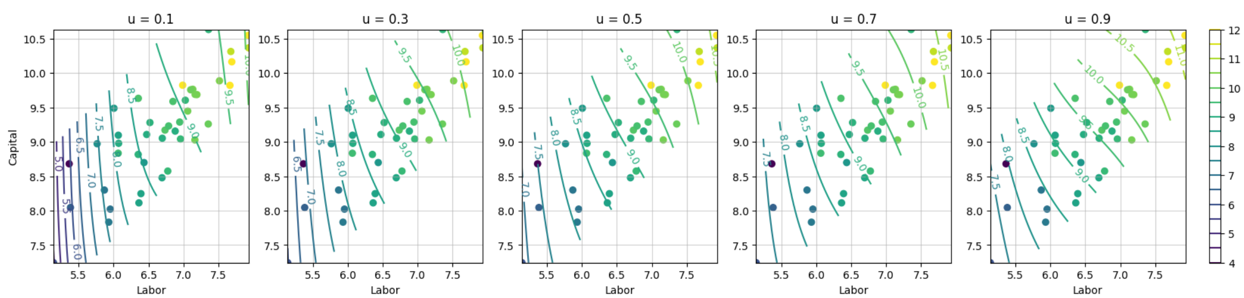

Figure 20: Our estimator under third-order translog specification.

References

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer (2015). “Identification properties of recent production function estimators”. In: *Econometrica* 83(6), pp. 2411–2451.

Ando, Tomohiro and Ker-Chau Li (2025). “Simplex quantile regression without crossing”. In: *The Annals of Statistics* 53(1), pp. 144–169.

Angrist, Joshua, Victor Chernozhukov, and Iván Fernández-Val (2006). “Quantile regression under misspecification, with an application to the US wage structure”. In: *Econometrica* 74(2), pp. 539–563.

Athey, Susan and Philip A Haile (2007). “Nonparametric approaches to auctions”. In: *Handbook of econometrics* 6, pp. 3847–3965.

Bondell, Howard D, Brian J Reich, and Huixia Wang (2010). "Noncrossing quantile regression curve estimation". In: *Biometrika* 97(4), pp. 825–838.

Carlier, Guillaume, Victor Chernozhukov, Gwendoline De Bie, and Alfred Galichon (2020). "Vector quantile regression and optimal transport, from theory to numerics". In: *Empirical Economics*, pp. 1–28.

Carlier, Guillaume, Victor Chernozhukov, and Alfred Galichon (2016). "Vector Quantile Regression: An Optimal Transport Approach". In: *Annals of Statistics* 44(3), pp. 1165–1192.

Carlier, Guillaume, Victor Chernozhukov, and Alfred Galichon (2017). "Vector quantile regression beyond the specified case". In: *Journal of Multivariate Analysis* 161, pp. 96–102.

Chen, Xiaohong (2007). "Large sample sieve estimation of semi-nonparametric models". In: *Handbook of econometrics* 6, pp. 5549–5632.

Chernozhukov, Victor, Iván Fernández-Val, and Alfred Galichon (2010). "Quantile and probability curves without crossing". In: *Econometrica* 78(3), pp. 1093–1125.

Dai, Sheng, Timo Kuosmanen, and Xun Zhou (2023). "Non-crossing convex quantile regression". In: *Economics letters* 233, p. 111396.

del Barrio, Eustasio and Jean-Michel Loubes (2019). "Central Limit Theorem for empirical transportation cost in general dimension". In: *Annals of Probability* 2.

Dette, Holger and Stanislav Volgushev (2008). "Non-crossing non-parametric estimates of quantile curves". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 70(3), pp. 609–627.

Doty, Justin and Suyong Song (2021). "Nonparametric Identification and Estimation of Quantile Production Functions". In: URL: https://github.com/justdoty/Quantile-Production-Functions/blob/master/Paper.pdf.

Dudley, R. M. (2002). *Real Analysis and Probability*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Feng, Oliver Y, Yining Chen, Qiyang Han, Raymond J Carroll, and Richard J Samworth (2022). "Nonparametric, tuning-free estimation of S-shaped functions". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84(4), pp. 1324–1352.

Firpo, Sergio, Antonio F Galvao, Cristine Pinto, Alexandre Poirier, and Graciela Sanroman (2022). "GMM quantile regression". In: *Journal of Econometrics* 230(2), pp. 432–452.

Frumento, Paolo and Matteo Bottai (2016). "Parametric modeling of quantile regression coefficient functions". In: *Biometrics* 72(1), pp. 74–84.

Gechert, Sebastian, Tomas Havranek, Zuzana Irsova, and Dominika Kolcunova (2022). "Measuring capital-labor substitution: The importance of method choices and publication bias". In: *Review of Economic Dynamics* 45, pp. 55–82.

Gimenes, Nathalie (2017). "Econometrics of ascending auctions by quantile regression". In: *Review of Economics and Statistics* 99(5), pp. 944–953.

Gimenes, Nathalie and Emmanuel Guerre (2022). "Quantile regression methods for first-price auctions". In: *Journal of Econometrics* 226(2), pp. 224–247.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong (2000). "Optimal nonparametric estimation of first-price auctions". In: *Econometrica* 68(3), pp. 525–574.

Hall, Peter and Li-Shan Huang (2001). "Nonparametric kernel regression subject to monotonicity constraints". In: *The Annals of Statistics* 29(3), pp. 624–647.

He, Xuming (1997). "Quantile curves without crossing". In: *The American Statistician* 51(2), pp. 186–192.

Henderson, Daniel J, John A List, Daniel L Millimet, Christopher F Parmeter, and Michael K Price (2012). "Empirical implementation of nonparametric first-price auction models". In: *Journal of Econometrics* 168(1), pp. 17–28.

Hoderlein, Stefan and Enno Mammen (2007). "Identification of marginal effects in nonseparable models without monotonicity". In: *Econometrica* 75(5), pp. 1513–1518.

Kim, Tae-Hwan and Halbert White (2003). "Estimation, inference, and specification testing for possibly misspecified quantile regression". In: *Maximum likelihood estimation of misspecified models: twenty years later*. Emerald Group Publishing Limited, pp. 107–132.

Kitahara, Daichi, Ke Leng, Yuji Tezuka, and Akira Hirabayashi (2021). "Simultaneous spline quantile regression under shape constraints". In: *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, pp. 2423–2427.

Kmenta, Jan (1967). "On estimation of the CES production function". In: *International Economic Review* 8(2), pp. 180–189.

Koenker, Roger and Gilbert Bassett (1978). "Regression quantiles". In: *Econometrica: journal of the Econometric Society*, pp. 33–50.

Koenker, Roger and Pin Ng (2005). "Inequality constrained quantile regression". In: *Sankhyā: The Indian Journal of Statistics*, pp. 418–440.

León-Ledesma, Miguel A, Peter McAdam, and Alpo Willman (2010). "Identifying the elasticity of substitution with biased technical change". In: *American Economic Review* 100(4), pp. 1330–1357.

López, Marco and Georg Still (2007). "Semi-infinite programming". In: *European journal of operational research* 180(2), pp. 491–518.

Lu, Jingfeng and Isabelle Perrigne (2008). “Estimating risk aversion from ascending and sealed-bid auctions: the case of timber auction data”. In: *Journal of Applied Econometrics* 23(7), pp. 871–896.

Luo, Yao and Yuanyuan Wan (2018). “Integrated-quantile-based estimation for first-price auction models”. In: *Journal of Business & Economic Statistics* 36(1), pp. 173–180.

Ma, Jun, Vadim Marmer, Artyom Shneyerov, and Pai Xu (2021). “Monotonicity-constrained nonparametric estimation and inference for first-price auctions”. In: *Econometric Reviews* 40(10), pp. 944–982.

Marmer, Vadim and Artyom Shneyerov (2012). “Quantile-based nonparametric inference for first-price auctions”. In: *Journal of Econometrics* 167(2), pp. 345–357.

Munkres, J.R. (2000). *Topology*. Featured Titles for Topology. Prentice Hall, Incorporated.

Newey, Whitney K (1997). “Convergence rates and asymptotic normality for series estimators”. In: *Journal of econometrics* 79(1), pp. 147–168.

Parker, Thomas (2019). “Asymptotic inference for the constrained quantile regression process”. In: *Journal of Econometrics* 213(1), pp. 174–189.

Pinelis, Iosif (2013). “A necessary and sufficient condition on the stability of the infimum of convex functions”. In: *arXiv preprint arXiv:1307.3806*.

Pinkse, Joris and Karl Schurter (2019). “Estimation of auction models with shape restrictions”. In: *arXiv preprint arXiv:1912.07466*.

Rockafellar, R Tyrrell (1970). *Convex Analysis*. Princeton University Press.

Rockafellar, R Tyrrell and Roger J-B Wets (2009). *Variational analysis*. Vol. 317. Springer Science & Business Media.

Shapiro, Alexander (2001). “On duality theory of conic linear problems”. In: *Nonconvex Optimization and its Applications* 57, pp. 135–155.

Shapiro, Alexander (2009). “Semi-infinite programming, duality, discretization and optimality conditions”. In: *Optimization* 58(2), pp. 133–161.

Spady, Richard Henry and Sami Stouli (2018). “Dual regression”. In: *Biometrika* 105(1), pp. 1–18.

van der Vaart, Aad W and Jon A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer. ISBN: 9780387946405.

Villani, Cédric (2003). *Topics in Optimal Transportation*. 58. American Mathematical Soc.

Villani, Cédric (2009). *Optimal transport: old and new*. Vol. 338. Springer.

Wang, Yongqiao, Shouyang Wang, Chuangyin Dang, and Wenxiu Ge (2014). “Nonparametric quantile frontier estimation under shape restriction”. In: *European Journal of Operational Research* 232(3), pp. 671–678.

Zellner, Arnold, Jan Kmenta, and Jacques Dreze (1966). "Specification and estimation of Cobb-Douglas production function models". In: *Econometrica: Journal of the Econometric Society*, pp. 784–795.