HAL
open science

Quantile and Probability Curves without Crossing

Victor Chernozhukov, Ivan Fernandez-Val, Alfred Galichon

To cite this version:

Victor Chernozhukov, Ivan Fernandez-Val, Alfred Galichon. Quantile and Probability Curves without Crossing. Econometrica, 2010, 78 (3), pp.1093-1125. hal-01052958

HAL Id: hal-01052958

https://sciencespo.hal.science/hal-01052958

Submitted on 29 Jul 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:0704.3649v2 [stat.ME] 25 Jul 2009

QUANTILE AND PROBABILITY CURVES WITHOUT CROSSING

VICTOR CHERNOZHUKOV† IVÁN FERNÁNDEZ-VAL§ ALFRED GALICHON‡

Abstract. This paper proposes a method to address the longstanding problem of lack of monotonicity in estimation of conditional and structural quantile functions, also known as the quantile crossing problem. The method consists in sorting or monotone rearranging the original estimated non-monotone curve into a monotone rearranged curve. We show that the rearranged curve is closer to the true quantile curve in finite samples than the original curve, establish a functional delta method for rearrangement-related operators, and derive functional limit theory for the entire rearranged curve and its functionals. We also establish validity of the bootstrap for estimating the limit law of the the entire rearranged curve and its functionals. Our limit results are generic in that they apply to every estimator of a monotone econometric function, provided that the estimator satisfies a functional central limit theorem and the function satisfies some smoothness conditions. Consequently, our results apply to estimation of other econometric functions with monotonicity restrictions, such as demand, production, distribution, and structural distribution functions. We illustrate the results with an application to estimation of structural quantile functions using data on Vietnam veteran status and earnings.

JEL Classification: C10, C50 AMS Classification: 62J02; 62E20, 62P20

Date: This version of the paper is of July 25, 2009. Previous, more extended, versions (September 2006, April 2007) are available at www.mit.edu/∼vchern/www and www.ArXiv.org. The method developed in this paper has now been incorporated in the package `quantreg` (Koenker, 2007) in `R`. The title of this paper is (partially) borrowed from the work of Xuming He (1997), to whom we are grateful for the inspiration and formulation of the problem. We would like to thank the editor Oliver Linton, three anonymous referees, Alberto Abadie, Josh Angrist, Andrew Chesher, Phil Cross, James Durbin, Ivar Ekeland, Brigham Frandsen, Raymond Guiteras, Xuming He, Roger Koenker, Joonhwan Lee, Vadim Marmer, Ilya Molchanov, Francesca Molinari, Whitney Newey, Steve Portnoy, Shinichi Sakata, Art Shneyerov, Alp Simsek, and participants at BU, CEMFI, CEMMAP Measurement Matters Conference, Columbia Conference on Optimal Transportation, Columbia, Cornell, Cowles Foundation 75th Anniversary Conference, Duke-Triangle, Ecole Polytechnique, Frontiers of Microeconometrics in Tokyo, Georgetown, Harvard-MIT, MIT, Northwestern, UBC, UCL, UIUC, University of Alicante, and University of Gothenburg Conference "Nonsmooth Inference, Analysis, and Dependence," for comments that helped us to considerably improve the paper. We are grateful to Alberto Abadie for providing us the data for the empirical example. The authors gratefully acknowledge research support from the National Science Foundation and chaire X-Dauphine "Finance et Développement Durable".

† Massachusetts Institute of Technology, Department of Economics and Operations Research Center, University College London, CEMMAP. E-mail: vchern@mit.edu.

§ Boston University, Department of Economics. E-mail: ivanf@bu.edu.

‡ Ecole Polytechnique, Département d'Economie. E-mail: alfred.galichon@polytechnique.edu.

1. Introduction

This paper addresses the longstanding problem of lack of monotonicity in the estimation of conditional and structural quantile functions, also known as the quantile crossing problem (He, 1997). The most common approach to estimating quantile curves is to fit a curve, often linear, pointwise for each probability index.[1] Researchers use this approach for a number of reasons, including parsimony of the resulting approximations and excellent computational properties. The resulting fits, however, may not respect a logical monotonicity requirement – that the quantile curve should be increasing as a function of the probability index. This paper introduces a natural monotonization of these empirical curves by sampling from the estimated non-monotone model, and then taking the resulting conditional quantile curves which by construction are monotone in the probability index. This construction of the monotone curve may be seen as a bootstrap and as a sorting or monotone rearrangement of the original non-monotone function (see Hardy et al., 1952, and references given below). We show that the rearranged curve is closer to the true quantile curve in finite samples than the original curve is, and derive functional limit distribution theory for the rearranged curve to perform simultaneous inference on the entire quantile function. Our theory applies to both dependent and independent data, and to a wide variety of original estimators, with only the requirement that they satisfy a functional central limit theorem. Our results also apply to many other econometric problems with monotonicity restrictions, such as demand and production functions, option pricing functions, yield curves, distribution functions, and structural quantile functions (see Matzkin, 1994, for more examples and additional references). As an example, we provide an empirical application to estimation of structural distribution and quantile functions based on Abadie (2002) and Chernozhukov and Hansen (2005, 2006).

There exist other methods to obtain monotonic fits for conditional quantile functions. He (1997), for example, proposed to impose a location-scale regression model, which naturally satisfies monotonicity. This approach is fruitful for location-scale situations, but in numerous cases the data do not satisfy the location-scale paradigm, as discussed in Lehmann (1974), Doksum (1974), and Koenker (2005). Koenker and Ng (2005) developed a computational method for quantile regression that imposes the non-crossing constraints in simultaneous fitting of a finite number of quantile curves. The statistical properties of this method have yet to be studied, and the method does not immediately apply to other quantile estimation methods. Mammen (1991) proposed two-step estimators, with mean estimation in the first step followed

[1]This includes all principal approaches to estimation of conditional quantile functions, such as the canonical quantile regression of Koenker and Bassett (1978) and censored quantile regression of Powell (1986). This also includes principal approaches to estimation of structural quantile functions, such as the instrumental quantile regression methods via control functions of Imbens and Newey (2001), Blundell and Powell (2003), Chesher (2003), and Koenker and Ma (2006), and instrumental quantile regression estimators of Chernozhukov and Hansen (2005, 2006).

by isotonization in the second.[2] Similarly to Mammen (1991), we can employ quantile estimation in the first step followed by isotonization in the second, obtaining an interesting method whose properties have yet to be studied. In contrast, our method uses rearrangement rather than isotonization, and is much better suited for quantile applications. The reason is that isotonization is best suited for applications with (near) flat target functions, while rearrangement is best suited for applications with steep target functions, as in typical quantile applications. Indeed, in a numerical example closely matching our empirical application, presented in Section 3, rearrangement significantly outperforms isotonization. Finally, in an independent and contemporaneous work, Dette and Volgushev (2008) propose to obtain monotonic quantile curves by applying an integral transform to a local polynomial estimate of the conditional distribution function, and derive pointwise limit theory for this estimator. In contrast, we directly monotonize any generic estimate of a conditional quantile function and then derive generic functional limit theory for the entire monotonized curve.[3]

In addition to resolving the longstanding problem of estimating quantile curves that avoid crossing, this paper develops a number of original theoretical results on rearrangement estimators. It therefore makes both practical and theoretical contributions to econometrics and statistics. In order to discuss these contributions more specifically, it is helpful first to review some of the relevant literature and available results. We begin by noting that the idea of rearrangement goes back at least to Chebyshev (see Bronstein et al., 2003, p. 31, Hardy et al., 1952, and Lorentz, 1953, among others). Rearrangements have been extensively used in functional analysis and operations research (Villani, 2003, and Carlier and Dana, 2005), but not in econometrics or statistics until recently. Recent research on rearrangements in statistics include the work of Fougeres (1997), which used rearrangement to produce a monotonic kernel density estimator and derived its uniform rates of convergence; Davydov and Zitikis (2005), which considered tests of monotonicity based on rearranged kernel mean regression; Dette et al. (2006) and Dette and Scheder (2006), which introduced smoothed rearrangements for kernel mean regressions and derived pointwise limit theory for these estimators; and Chernozhukov et al. (2006), which used univariate and multivariate rearrangements on point and interval estimators of monotone functions based on series and kernel regression estimators. In the context of our problem, rearrangement is also connected to the quantile regression bootstrap of Koenker (1994). In fact, our research grew from the realization that we could use this bootstrap for the purpose of monotonizing quantile regressions, and we discovered the link to the classical procedure of rearrangement later, while reading Villani (2003).

[2]Isotonization is also known as the "pool-adjacent-violators algorithm" in statistics and "ironing" in economics. It amounts to projecting an initial estimate on the set of monotone functions.

[3]We refer to Dette and Volgushev (2008) for a nice, more detailed comparison of the two approaches.

The theoretical contributions of this paper are threefold. First, our paper derives functional limit theory for rearranged estimators and functional delta methods for rearrangement operators, both of which are important original results. Second, the paper derives functional limit results for estimators obtained by rearrangement-related operations, which are also original results. For example, our theory includes as a special case the asymptotics of the conditional distribution function estimator based on quantile regression, whose properties have long remained unknown. Moreover, our limit theory applies to functions, encompassing the pointwise results. An attractive feature of our theoretical results is that they do *not* rely on independence of data, the particular estimation method used, or any parametric assumptions. They only require that a functional central limit theorem applies to the original estimator of the curve, and the population curves have some smoothness properties. Our results therefore apply to any quantile model and quantile estimator that satisfy these requirements. Third, our results immediately yield validity of the bootstrap for rearranged estimators, which is an important result for practice.

We organize the rest of the paper as follows. In Section 2 we present some analytical results on rearrangement and then present all the main results; in Section 3 we provide an application and a numerical experiment that closely matches the application; and in Section 4 we give some concluding remarks.

2. Rearrangement: Analytical and Empirical Properties

In this section, we first describe rearrangement, then derive some basic analytical properties of the rearranged curves in the population, establish functional differentiability results, and finally establish functional limit theorems and other estimation properties.

2.1. **Rearrangement.** We consider a target function $u \mapsto Q_0(u|x)$ that, for each $x \in \mathcal{X}$, maps $(0,1)$ to the real line and is increasing in u. Suppose that $u \mapsto \widehat{Q}(u|x)$ is a parametric or nonparametric estimator of $Q_0(u|x)$. Throughout the paper, we use conditional and structural quantile estimation as the main application, where $u \mapsto Q_0(u|x)$ is the quantile function of a real response variable Y, given a vector of regressors $X = x$. Accordingly, we will usually refer to the functions $u \mapsto Q_0(u|x)$ as quantile functions throughout the paper. In other applications, such as estimation of conditional and structural distribution functions, other names would be appropriate and we need to accommodate different domains, as described in Remark 1 below.

Typical estimation methods fit the quantile function $\widehat{Q}(u|x)$ pointwise in $u \in (0,1)$.[4] A problem that might occur is that the map $u \mapsto \widehat{Q}(u|x)$ may not be increasing in u, which violates the logical monotonicity requirement. Another manifestation of this issue, known as

[4]See Koenker and Bassett (1978), Powell (1986), Chaudhuri (1991), Buchinsky and Hahn (1998), Yu and Jones (1998), Abadie et al. (2002), Honoré et al. (2002), and Chernozhukov and Hansen (2006), among others, for examples of exogenous, censored, endogenous, nonparametric, and other types of quantile regression estimators.

the quantile crossing problem, is that the conditional quantile curves $x \mapsto \widehat{Q}(u|x)$ may cross for different values of u (He, 1997). Similar issues also arise in estimation of conditional and structural distribution functions (Hall et al., 1999, and Abadie, 2002).

We can transform the possibly non-monotone function $u \mapsto \widehat{Q}(u|x)$ into a monotone function $u \mapsto \widehat{Q}^*(u|x)$ by quantile bootstrap or rearrangement. That is, we consider the random variable $Y_x := \widehat{Q}(U|x)$ where $U \sim \text{Uniform}(\mathcal{U})$ with $\mathcal{U} = (0,1)$, and take its quantile function denoted by $u \mapsto \widehat{Q}^*(u|x)$ instead of the original function $u \mapsto \widehat{Q}(u|x)$. This variable Y_x has a distribution function:

$$\widehat{F}(y|x) := \int_0^1 1\{\widehat{Q}(u|x) \leq y\} du, \tag{2.1}$$

which is naturally monotone in the level y, and quantile function:

$$\widehat{Q}^*(u|x) := \widehat{F}^{-1}(u|x) = \inf\{y : \widehat{F}(y|x) \geq u\}, \tag{2.2}$$

which is naturally monotone in the index u. Thus, starting with a possibly non-monotone original curve $u \mapsto \widehat{Q}(u|x)$, the rearrangement (2.1)-(2.2) produces a monotone quantile curve $u \mapsto \widehat{Q}^*(u|x)$. Of course, the rearranged quantile function $u \mapsto \widehat{Q}^*(u|x)$ coincides with the original function $u \mapsto \widehat{Q}(u|x)$ if the original function is non-decreasing in u, but differs from it otherwise.

The mechanism (2.1)-(2.2) and its name have a direct relation to the rearrangement operator from functional analysis (Hardy et al., 1952), since $u \mapsto \widehat{Q}^*(u|x)$ is the monotone rearrangement of $u \mapsto \widehat{Q}(u|x)$. Equivalently, as we stated earlier, rearrangement has a direct relation to the quantile bootstrap (Koenker, 1994), since the rearranged quantile curve is the quantile function of the bootstrap variable produced by the estimated quantile model. Moreover, we refer the reader to Dette et al. (2006, p. 470) who, using a closely related motivation, introduced the idea of smoothed rearrangement, which produces smoothed versions of (2.1) and (2.2), which can be valuable in applications. Finally, for practical and computational purposes, it is helpful to think of rearrangement as sorting. Indeed to compute the rearrangement of a continuous function $u \mapsto \widehat{Q}(u|x)$ we simply set $\widehat{Q}^*(u|x)$ as the u-th quantile of $\{\widehat{Q}(u_1|x), ..., \widehat{Q}(u_k|x)\}$, where $\{u_1, ...u_k\}$ is a sufficiently fine net of equidistant indices in $(0,1)$.

Remark 1.(Adjusting for domains other than the unit interval). Throughout the paper we assume that the domain of all the functions is the unit interval, $\mathcal{U} = (0,1)$, but in many applications we may have to deal with different domains. For example, in quantile estimation problems, we may consider a subinterval (a,b) of the unit interval as the domain, in order to avoid estimation of tail quantiles. In distribution estimation problems, we may consider the entire real line as the domain. In such cases we can first transform these functions to have the unit interval as the domain. Concretely, suppose we have $\bar{Q} : (a,b) \to \mathbb{R}$. Then using any increasing bijective mapping $\varphi : (a,b) \mapsto (0,1)$, we can define $Q := \bar{Q} \circ \varphi^{-1} : (0,1) \to \mathbb{R}$, and then proceed to obtain its rearrangement Q^*. In the case where $a \neq -\infty$ and $b \neq \infty$, we can

take φ to be an affine mapping. In order to obtain the rearrangement of the original function $\bar{Q}$, we then set $\bar{Q}^* = Q^* \circ \varphi$. □

Let Q denote the pointwise probability limit of $\widehat{Q}$, which we will refer to as the population curve. In the analysis we distinguish the following two cases:

(1) Monotonic Q: The population curve $u \mapsto Q(u|x)$ is increasing in u, and thus satisfies the monotonicity requirement.
(2) Non-monotonic Q: The population curve $u \mapsto Q(u|x)$ is non-monotone due to misspecification.

In case (1) the empirical curve $u \mapsto \widehat{Q}(u|x)$ may be non-monotone due to estimation error, while in case (2) it may be non-monotone due to both misspecification and estimation error. A leading example of case (1) is when the population curve Q is correctly specified, so that it equals the target quantile curve, namely $Q(u|x) = Q_0(u|x)$ for all $u \in (0,1)$. Case (1) also allows for some degree of misspecification, provided that the population curve, $Q \neq Q_0$, remains monotone. A leading example of case (2) is when the population curve Q is misspecified, $Q \neq Q_0$, to a degree that makes $u \mapsto Q(u|x)$ non-monotone. For example, the common linear specification $u \mapsto Q(u|x) = p(x)^\mathsf{T}\beta(u)$ can be non-monotone if the support of X is sufficiently rich, while the set of transformations of x, $p(x)$, is not (Koenker, 2005, Chap 2.5). Typically, by using a rich enough set $p(x)$ we can approximate the true function $Q_0(u|x)$ sufficiently well, and thus often avoid case (2). This is the strategy that we generally recommend, since inference and limit theory under case (1) is theoretically and practically simpler than under case (2). However, in what follows we analyze the behavior of rearranged estimates both in cases (1) and (2), since either of these cases could occur in practice.

In the rest of the section, we establish the empirical properties of the rearranged estimated quantile functions and the corresponding distribution functions:

$$u \mapsto \widehat{Q}^*(u|x) \quad \text{and} \quad y \mapsto \widehat{F}(y|x), \tag{2.3}$$

under cases (1) and (2).

2.2. **Basic Analytical Properties of Population Curves.** We start by characterizing certain analytical properties of the probability limits or population versions of empirical curves (2.3), namely

$$\begin{aligned} &y \mapsto F(y|x) = \textstyle\int_0^1 1\{Q(u|x) \leq y\}du, \\ &u \mapsto Q^*(u|x) := F^{-1}(u|x) = \inf\{y : F(y|x) \geq u\}. \end{aligned} \tag{2.4}$$

We need these properties to derive our main limit results stated in the following sections.

Recall first the following definitions from Milnor (1965). Let $g : \mathcal{U} \subset \mathbb{R} \mapsto \mathbb{R}$ be a continuously differentiable function. A point $u \in \mathcal{U}$ is called a regular point of g if the derivative of g at this point does not vanish, i.e., $\partial_u g(u) \neq 0$, where ∂_u denotes the partial derivative operator with respect to u . A point u which is not a regular point is called a critical point.

A value $y \in g(\mathcal{U})$ is called a regular value of g if $g^{-1}(y)$ contains only regular points, i.e., if $\forall u \in g^{-1}(y)$, $\partial_u g(u) \neq 0$. A value y which is not a regular value is called a critical value.

Define region $\mathcal{Y}_x$ as the support of Y_x, and regions $\mathcal{YX} := \{(y,x) : y \in \mathcal{Y}_x, x \in \mathcal{X}\}$ and $\mathcal{UX} := \mathcal{U} \times \mathcal{X}$. We assume throughout that $\mathcal{Y}_x \subset \mathcal{Y}$, a compact subset of $\mathbb{R}$, and that $x \in \mathcal{X}$, a compact subset of $\mathbb{R}^d$. In some applications the curves of interest are not functions of x, or we might be interested in a particular value x. In this case, we can take the set $\mathcal{X}$ to be a singleton $\mathcal{X} = \{x\}$.

Assumption 1. (Properties of Q). *We maintain the following assumptions on Q throughout the paper:*

(a) $Q : \mathcal{U} \times \mathcal{X} \mapsto \mathbb{R}$ *is a continuously differentiable function in both arguments.*
(b) *The number of elements of* $\{u \in \mathcal{U} \mid \partial_u Q(u|x) = 0\}$ *is uniformly bounded on* $x \in \mathcal{X}$.

Assumption 1(b) implies that, for each $x \in \mathcal{X}$, $\partial_u Q(u|x)$ is not zero almost everywhere on $\mathcal{U}$ and can switch sign only a bounded number of times. Further, we define $\mathcal{Y}_x^*$ be the subset of regular values of $u \mapsto Q(u|x)$ in $\mathcal{Y}_x$, and $\mathcal{YX}^* := \{(y,x) : y \in \mathcal{Y}_x^*, x \in \mathcal{X}\}$.

We use the following simple example to describe some basic analytical properties of (2.4), which we state more formally in the proposition given below. Consider the following pseudo-quantile function: $Q(u) = 5\{u + \sin(2\pi u)/\pi\}$, which is highly non-monotone in $(0,1)$ and therefore fails to be a proper quantile function. The left panel of Figure 1 shows Q together with its monotone rearrangement Q^*. We see that Q^* partially coincides with Q on the areas where Q behaves like a proper quantile function, and that Q^* is continuous and increasing. Note also that $1/3$ and $2/3$ are the critical points of Q, and 3.04 and 1.96 are the corresponding critical values. The right panel of Figure 1 shows the pseudo-distribution function Q^{-1}, which is multi-valued, and the distribution function $F = Q^{*-1}$ induced by sampling from Q. We see that F is continuous and does not have point masses. The left panel of Figure 2 shows $\partial_u Q^*$, the sparsity function for Q^*. We see that the sparsity function is continuous at the Q^{*-1}-image of the regular values of Q and has jumps at the Q^{*-1}-image of the critical values of Q. The right panel of Figure 2 shows $\partial_y F$, the density function for F. We see that $\partial_y F$ is continuous at the regular values of Q and has jumps at the critical values of Q.

The following proposition states more formally the properties of Q^* and F:

Proposition 1 (Basic properties of F and Q^*)**.** *The functions $y \mapsto F(y|x)$ and $u \mapsto Q^*(u|x)$ satisfy the following properties, for each $x \in \mathcal{X}$: (1) The set of critical values, $\mathcal{Y}_x \setminus \mathcal{Y}_x^*$, is finite, and $\int_{\mathcal{Y}_x \setminus \mathcal{Y}_x^*} dF(y|x) = 0$. (2) For any $y \in \mathcal{Y}_x^*$,*

$$F(y|x) = \sum_{k=1}^{K(y|x)} sign\{\partial_u Q(u_k(y|x)|x)\} u_k(y|x) + 1\{\partial_u Q(u_{K(y|x)}(y|x)|x) < 0\},$$

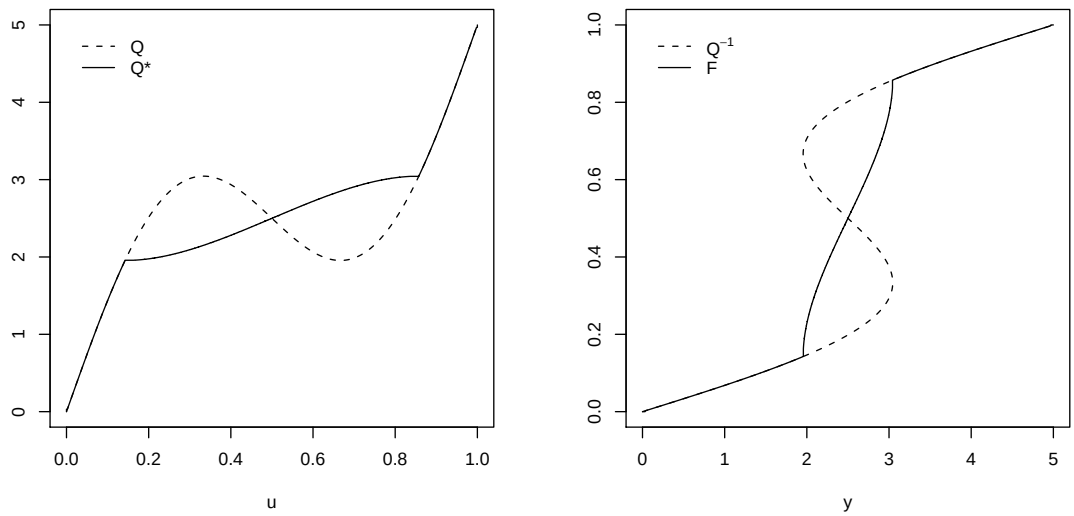

FIGURE 1. Left: The pseudo-quantile function Q and the rearranged quantile function Q^*. Right: The pseudo-distribution function Q^{-1} and the distribution function F induced by Q.

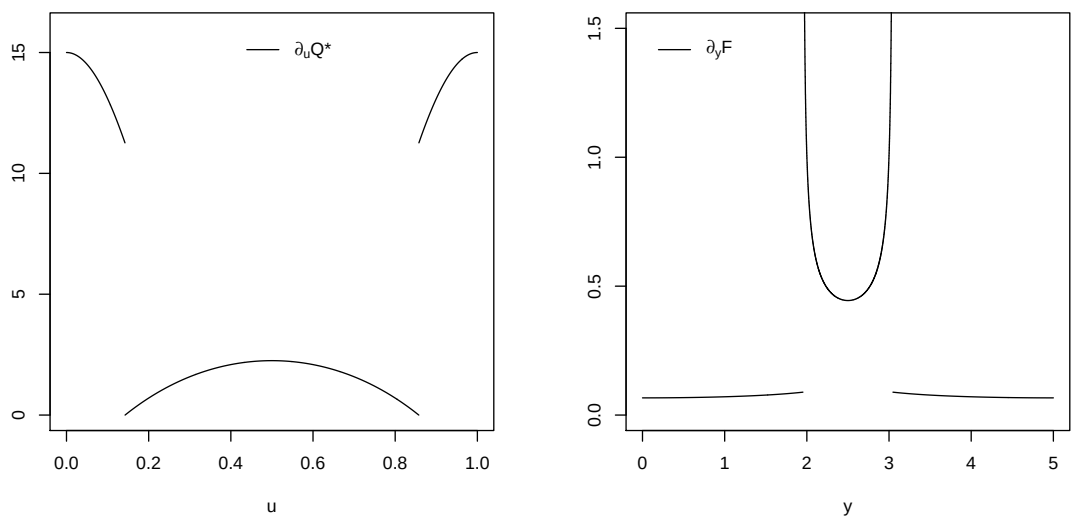

FIGURE 2. Left: The density (sparsity) function of the rearranged quantile function Q^*. Right: The density function of the distribution function F induced by Q.

where $\{u_k(y|x),\ \text{for } k = 1, 2, ..., K(y|x) < \infty\}$ *are the roots of* $Q(u|x) = y$ *in increasing order. (3) For any* $y \in \mathcal{Y}_x^*$*, the ordinary derivative* $f(y|x) := \partial_y F(y|x)$ *exists and takes the form*

$$f(y|x) = \sum_{k=1}^{K(y|x)} \frac{1}{|\partial_u Q(u_k(y|x)|x)|},$$

which is continuous at each $y \in \mathcal{Y}_x^*$*. For any* $y \in \mathcal{Y} \setminus \mathcal{Y}_x^*$*, we set* $f(y|x) := 0$*.* $F(y|x)$ *is absolutely continuous and strictly increasing in* $y \in \mathcal{Y}_x$*. Moreover,* $y \mapsto f(y|x)$ *is a Radon-Nikodym derivative of* $y \mapsto F(y|x)$ *with respect to the Lebesgue measure. (4) The quantile function* $u \mapsto Q^*(u|x)$ *partially coincides with* $u \mapsto Q(u|x)$*; namely* $Q^*(u|x) = Q(u|x)$*, provided that* $u \mapsto Q(u|x)$ *is increasing at* u*, and the preimage of* $Q^*(u|x)$ *under* Q *is unique. (5) The quantile function* $u \mapsto Q^*(u|x)$ *is equivariant to monotone transformations of* $u \mapsto Q(u|x)$*, in particular, to location and scale transformations. (6) The quantile function* $u \mapsto Q^*(u|x)$ *has an ordinary continuous derivative* $\partial_u Q^*(u|x) = 1/f(Q^*(u|x)|x)$*, when* $Q^*(u|x) \in \mathcal{Y}_x^*$*. This function is also a Radon-Nikodym derivative with respect to the Lebesgue measure. (7) The map* $(y, x) \mapsto F(y|x)$ *is continuous on* $\mathcal{YX}$ *and the map* $(u, x) \mapsto Q^*(u|x)$ *is continuous on* $\mathcal{UX}$*.*

2.3. **Functional Delta Method for Rearrangement-Related Operators.** Here we derive a functional delta method for the rearrangement operator $Q \mapsto Q^*$ and the pre-rearrangement operator $Q \mapsto F$ defined by equation (2.4). These results constitute the first set of original main theoretical results obtained in this paper. In the subsequent sections, these results allow us to establish a generic functional central limit theorem for the estimated functions $\widehat{Q}^*$ and $\widehat{F}$, as well as to establish validity of the bootstrap for estimating their limit laws.

In order to describe the results, let $\ell^\infty(\mathcal{UX})$ denote the set of bounded and measurable functions $h : \mathcal{UX} \mapsto \mathbb{R}$, $C(\mathcal{UX})$ the set of continuous functions $h : \mathcal{UX} \mapsto \mathbb{R}$, and $\ell^1(\mathcal{UX})$ the set of measurable functions $h : \mathcal{UX} \mapsto \mathbb{R}$ such that $\int_{\mathcal{U}} \int_{\mathcal{X}} |h(u|x)| du dx < \infty$, where du and dx denote the integration with respect to the Lebesgue measure on $\mathcal{U}$ and $\mathcal{X}$, respectively.

Proposition 2 (Hadamard derivatives of F and Q^* with respect to Q)**.** *(1) Define* $F(y|x, h_t) := \int_0^1 1\{Q(u|x) + t h_t(u|x) \le y\} du$*. As* $t \to 0$*,*

$$D_{h_t}(y|x, t) := \frac{F(y|x, h_t) - F(y|x)}{t} \to D_h(y|x), \tag{2.5}$$

$$D_h(y|x) := -\sum_{k=1}^{K(y|x)} \frac{h(u_k(y|x)|x)}{|\partial_u Q(u_k(y|x)|x)|}. \tag{2.6}$$

The convergence holds uniformly in any compact subset of $\mathcal{YX}^* := \{(y, x) : y \in \mathcal{Y}_x^*, x \in \mathcal{X}\}$*, for every* $|h_t - h|_\infty \to 0$*, where* $h_t \in \ell^\infty(\mathcal{UX})$*, and* $h \in C(\mathcal{UX})$*. (2) Define* $Q^*(u|x, h_t) :=$

$F^{-1}(y|x,h_t) = \inf\{y : F(y|x,h_t) \geq u\}$. *As* $t \to 0$,

$$\tilde{D}_{h_t}(u|x,t) := \frac{Q^*(u|x,h_t) - Q^*(u|x)}{t} \to \tilde{D}_h(u|x), \tag{2.7}$$

$$\tilde{D}_h(u|x) := -\frac{1}{f(Q^*(u|x)|x)}\, D_h(Q^*(u|x)|x). \tag{2.8}$$

The convergence holds uniformly in any compact subset of $\mathcal{UX}^* = \{(u,x) : (Q^*(u|x),x) \in \mathcal{YX}^*\}$, *for every* $|h_t - h|_\infty \to 0$, *where* $h_t \in \ell^\infty(\mathcal{UX})$, *and* $h \in C(\mathcal{UX})$.

This proposition establishes the Hadamard (compact) differentiability of the rearrangement operator $Q \mapsto Q^*$ and the pre-rearrangement operator $Q \mapsto F$ with respect to Q, tangentially to the subspace of continuous functions. Note that the convergence holds uniformly on regions that exclude the critical values of the mapping $u \mapsto Q(u|x)$. These results are new and could be of independent interest. Rearrangement operators include inverse (quantile) operators as a special case. In this sense, our results generalize the previous results of Gill and Johansen (1990), Doss and Gill (1992), and Dudley and Norvaisa (1999) on functional delta method (Hadamard differentiability) for the quantile operator. There are two main difficulties in establishing the Hadamard differentiability in our case: first, like in the quantile case, we allow the perturbations h_t to Q to be discontinuous functions, though converging to continuous functions; second, unlike in the quantile case, we allow the perturbed functions $Q + th_t$ to be non-monotone even when Q is monotone. We need to allow for such rich perturbations in order to match empirical applications, where empirical perturbations $h_t = (\widehat{Q} - Q)/t$ are discontinuous functions, though converging to continuous functions by the means of a functional central limit theorem; moreover, the empirical (pseudo) quantile functions $\widehat{Q} = Q + th_t$ are not monotone even when Q is monotone.

The following result deals with the monotonic case. It is worth emphasizing separately, because functional derivatives are particularly simple and we do not have to exclude any non-regular regions from the domains.

Corollary 1 (Hadamard derivatives of F and Q^* with respect to Q in the monotonic case)**.** *Suppose* $u \mapsto Q(u|x)$ *has* $\partial_u Q(u|x) > 0$, *for each* $(u,x) \in \mathcal{UX}$. *Then* $\mathcal{YX}^* = \mathcal{YX}$ *and* $\mathcal{UX}^* = \mathcal{UX}$. *Therefore, the convergence in Proposition 2 holds uniformly over the entire* $\mathcal{YX}$ *and* $\mathcal{UX}$, *respectively. Moreover,* $\tilde{D}_h(u|x) = h$, *i.e., the Hadamard derivative of the rearranged function with respect to the original function is the identity operator.*

Next we consider the following linear functionals obtained by integration:

$$(y',x) \mapsto \int_{\mathcal{Y}} g(y|x,y')F(y|x)dy, \quad (u',x) \mapsto \int_{\mathcal{U}} g(u|x,u')Q^*(u|x)du,$$

with the restrictions on g specified below. These functionals are of interest because they are useful building blocks for various statistics, for example, Lorenz curves with function

$g(u|x,u') = 1\{u \leq u'\}$, as discussed in the next section. The following proposition calculates the Hadamard derivative of these functionals.

Proposition 3 (Hadamard derivative of linear functionals of Q^* and F with respect to Q)**.** *The following results are true with the limits being continuous on the specified domains:*

$$1. \quad \int_{\mathcal{Y}} g(y|x,y')D_{h_t}(y|x,t)dy \to \int_{\mathcal{Y}} g(y|x,y')D_h(y|x)dy$$

uniformly in $(y',x) \in \mathcal{YX}$, for any measurable g that is is bounded uniformly in its arguments and such that $(x,y') \mapsto g(y|x,y')$ is continuous for a.e. y.

$$2. \quad \int_{\mathcal{U}} g(u|x,u')\tilde{D}_{h_t}(u|x,t)du \to \int_{\mathcal{U}} g(u|x,u')\tilde{D}_h(u|x)du \tag{2.9}$$

uniformly in $(u',x) \in \mathcal{UX}$, for any measurable g such that $\sup_{u',x} |g(u|x,u')| \in \ell^1(\mathcal{U})$ and such that $(x,u') \mapsto g(u|x,u')$ is continuous for a.e. u.

It is important to note that Proposition 3 applies to integrals defined over entire domains, unlike Proposition 2 which states uniform convergence of integrands over domains excluding non-regular neighborhoods. (Thus, Proposition 3 does not immediately follow from Proposition 2.) Here integration acts like a smoothing operation and allows us to ignore these non-regular neighborhoods. In order to prove convergence of integrals defined over entire domains, we couple the pointwise convergence implied by Proposition 2 with the uniform integrability of Lemma 3 in the Appendix, and then interchange limits and integrals. We should also note that an alternative way of proving result (2.9), but *not* other results in the paper, can be based on the convexity of the functional in (2.9) with respect to the underlying curve, following the approach of Mossino and Temam (1981), and Alvino et al. (1989). Due to this limitation, we do not pursue this approach in this paper.

It is also worth emphasizing the properties of the following smoothed functionals. For a measurable function $f : \mathbb{R} \mapsto \mathbb{R}$ define the smoothing operator S as

$$Sf(y') := \int k_\delta(y'-y)f(y)dy, \tag{2.10}$$

where $k_\delta(v) = 1\{|v| \leq \delta\}/2\delta$ and $\delta > 0$ is a fixed bandwidth. Accordingly, the smoothed curves SF and SQ^* are given by

$$SF(y'|x) := \int k_\delta(y'-y)F(y|x)dy, \qquad SQ^*(u'|x) := \int k_\delta(u'-u)Q^*(u|x)du.$$

Note that given the quantile function Q^*, the smoothed function SQ^* has a convenient interpretation of a local average quantile function or fractile. Since we form these curves as differences of the elementary functionals in Proposition 3 divided by 2δ, the following corollary is immediate:

Corollary 2 (Hadamard derivative of smoothed Q^* and F with respect to Q)**.** *We have that $SD_{h_t}(y'|x,t) \to SD_h(y'|x)$ uniformly in $(y',x) \in \mathcal{Y}\mathcal{X}$, and $S\tilde{D}_{h_t}(u'|x,t) \to S\tilde{D}_h(u'|x)$ uniformly in $(u',x) \in \mathcal{U}\mathcal{X}$. The results hold uniformly in the smoothing parameter $\delta \in [\delta_1, \delta_2]$, where δ_1 and δ_2 are positive constants.*

Note that smoothing allows us to achieve uniform convergence over the entire domain, without excluding non-regular neighborhoods.

2.4. **Empirical Properties and Functional Limit Theory for Rearranged Estimators.** Here we state a finite sample result and then derive functional limit laws for rearranged estimators. These results constitute the second set of original main theoretical results obtained in this paper.

The following proposition shows that the rearranged quantile curves have smaller estimation error than the original curves whenever the latter are not monotone.

Proposition 4 (Improvement in estimation property provided by rearrangement)**.** *Suppose that $\widehat{Q}$ is an estimator (not necessarily consistent) for some true quantile curve Q_0. Then, the rearranged curve $\widehat{Q}^*$ is closer to the true curve than $\widehat{Q}$ in the sense that, for each $x \in \mathcal{X}$,*

$$\|\widehat{Q}^* - Q_0\|_p \leq \|\widehat{Q} - Q_0\|_p, \ p \in [1, \infty],$$

where $\|\cdot\|_p$ denotes the L^p norm of a measurable function $Q : \mathcal{U} \mapsto \mathbb{R}$, namely $\|Q\|_p = \{\int_{\mathcal{U}} |Q(u)|^p du\}^{1/p}$. The inequality is strict for $p \in (1, \infty)$ whenever $u \mapsto \widehat{Q}(u|x)$ is strictly decreasing on a subset of $\mathcal{U}$ of positive Lebesgue measure, while $u \mapsto Q_0(u|x)$ is strictly increasing on $\mathcal{U}$. The above property is independent of the sample size and of the way the estimate of the curve is obtained, and thus continues to hold in the population.

This property suggests that the rearranged estimators should be preferred over the original estimators. Moreover, this property does not depend on the way the quantile model is estimated or any other specifics, and is thus applicable quite generally. Regarding the proof of this property, the weak reduction in estimation error follows from an application of a classical rearrangement inequality of Lorentz (1953) and the strict reduction follows from its appropriate strengthening (Chernozhukov et al., 2006).[5]

The following proposition derives functional limit laws for the rearranged quantile estimator $\widehat{Q}^*$ and the corresponding distribution estimator $\widehat{F}$, using the functional delta method for the rearrangement-related operators from the previous section. We maintain the following assumptions on $\widehat{Q}$ throughout the paper:

[5]Similar contractivity properties have been shown for the pool adjacent violators algorithm in different contexts. See, for example, Robertson et al. (1988) for isotonic regression, and Eggermont and LaRiccia (2000) for monotone density estimation. Glad et al. (2003) shows that a density estimator corrected to be a proper density satisfies a similar property.

Assumption 2. (Properties of $\widehat{Q}$). *The empirical curve $\widehat{Q}$ takes its values in the space of bounded measurable functions defined on $\mathcal{UX}$, and, in $\ell^\infty(\mathcal{UX})$,*

$$a_n(\widehat{Q}(u|x) - Q(u|x)) \Rightarrow G(u|x), \tag{2.11}$$

as a stochastic process indexed by $(u,x) \in \mathcal{UX}$, where $(u,x) \mapsto G(u|x)$ is a stochastic process (typically Gaussian) with continuous paths. Here a_n is a sequence of constants such that $a_n \to \infty$ as $n \to \infty$, where n is the sample size.

Condition (2.11) requires that the original quantile estimator satisfies a functional central limit theorem with a continuous limit stochastic process over the domain $\mathcal{U} = (0,1)$ for the index u. If (2.11) holds only over a subinterval of $(0,1)$, we can accommodate the reduced domain following Remark 1. This key condition is rather weak, and it holds for a wide variety of conditional and structural quantile estimators.[6] With an appropriate normalization rate and a fixed x, this assumption also holds for series and local-polynomial quantile regressions.[7]

Proposition 5 (Functional limit laws for $\widehat{F}$ and $\widehat{Q}^*$)**.** *In $\ell^\infty(K)$, where K is any compact subset of $\mathcal{YX}^*$,*

$$a_n(\widehat{F}(y|x) - F(y|x)) \Rightarrow D_G(y|x) \tag{2.12}$$

as a stochastic process indexed by $(y,x) \in \mathcal{YX}^*$; and in $\ell^\infty(\mathcal{UX}_K)$, with $\mathcal{UX}_K = \{(u,x) : (Q^*(u|x), x) \in K\}$,

$$a_n(\widehat{Q}^*(u|x) - Q^*(u|x)) \Rightarrow \widetilde{D}_G(u|x), \tag{2.13}$$

as a stochastic process indexed by $(u,x) \in \mathcal{UX}_K$.

This proposition provides the basis for inference using rearranged quantile estimators and corresponding distribution estimators. Let us first discuss inference for the case with a monotonic population curve Q. Proposition 5 enables us to perform uniform inference on Q and F based on the rearranged estimators $\widehat{Q}^*$ and $\widehat{F}$. It is useful to emphasize the following corollary of Proposition 5:

Corollary 3 (Functional limit laws for $\widehat{F}$ and $\widehat{Q}^*$ in the monotonic case)**.** *Suppose $u \mapsto Q(u|x)$ has $\partial_u Q(u|x) > 0$ for each $(u,x) \in \mathcal{UX}$. Then $\mathcal{YX}^* = \mathcal{YX}$ and $\mathcal{UX}^* = \mathcal{UX}$. Accordingly, the convergence in Proposition 5 holds uniformly over the entire $\mathcal{YX}$ and $\mathcal{UX}$. Moreover, $\widetilde{D}_G(u|x) = G(u|x)$, i.e., the rearranged quantile curves have the same first order asymptotic distribution as the original estimated quantile curves.*

[6] For sufficient conditions, see, for example, Gutenbrunner and Jurečková (1992), Portnoy (1991), Angrist et al. (2006), and Chernozhukov and Hansen (2006).

[7] See, for example, Chaudhuri (1991) and He and Shao (2000); Belloni and Chernozhukov (2007) have recently extended the results of He and Shao (2000) to the process case and established the functional central limit theorem for $a_n(\widehat{Q}(u|x) - Q(u|x))$ for a fixed x.

Thus, if the population curve is monotone, we can rearrange the original non-monotone quantile estimator to be monotonic *without* affecting its (first order) asymptotic properties. Hence, all the inference tools that apply to the original quantile estimator $\widehat{Q}$ also apply to the rearranged quantile estimator $\widehat{Q}^*$. In particular, if the bootstrap is valid for the original estimate, it is also valid for the rearranged estimate, by the functional delta method for the bootstrap.

Remark 2.(Detecting and avoiding cases with non-monotone Q.) Before discussing inference for the case with a non-monotonic population curve Q, let us first emphasize that since non-monotonicity of Q is a rather obvious sign of specification error, it is best to try to detect and avoid this case. For this purpose we should use sufficiently flexible functional forms and reject the ones that fail to pass monotonicity tests. For example, we can use the following generic test of monotonicity for Q: If Q is monotone, the first order behavior of $\widehat{Q}^*$ and $\widehat{Q}$ coincides, and if Q is not monotone, $\widehat{Q}^*$ and $\widehat{Q}$ converge to different probability limits Q^* and Q. Therefore, we can reject the hypothesis of monotone Q if a uniform confidence region for Q based on $\widehat{Q}$ does not contain $\widehat{Q}^*$, for at least one point $x \in \mathcal{X}$.[8] □

Let us now discuss inference for the case with a non-monotonic population curve Q. In this case, the large sample properties of the rearranged quantile estimators $\widehat{Q}^*$ substantially differ from those of the initial quantile estimators $\widehat{Q}$. Proposition 5 still enables us to perform uniform inference on the rearranged population curve Q^* based on the rearranged estimator $\widehat{Q}^*$, but only after excluding certain nonregular neighborhoods (for the distribution estimates, the neighborhoods of the critical values of the map $u \mapsto Q(u|x)$, and, for the rearranged quantile estimates, the image of these neighborhoods under F). These neighborhoods can be excluded by locating the points (u, x) where a consistent estimate of $|\partial_u Q(u|x)|$ is close to zero; see Hendricks and Koenker (1991) for a consistent estimator of $|\partial_u Q(u|x)|$.

Next we consider the following linear functionals of the rearranged quantile and distribution estimates:

$$(y', x) \mapsto \int_{\mathcal{Y}} g(y|x, y')\widehat{F}(y|x)dy, \quad (u', x) \mapsto \int_{\mathcal{U}} g(u|x, u')\widehat{Q}^*(u|x)du.$$

The following proposition derives functional limit laws for these functionals.[9] Here the convergence results hold without excluding any nonregular neighborhoods, which is convenient for practice in the non-monotonic case.

[8]This test is conservative, but it is generic and very inexpensive. In order to build non-conservative tests, we need to derive the limit laws for $\|\widehat{Q} - \widehat{Q}^*\|$ for suitable norms $\|\cdot\|$. These laws will depend on higher-order functional limit laws for quantile estimators, which appear to be non-generic and have to be dealt with on a case by case basis.

[9] Working with these functionals is equivalent to placing our empirical processes into the space L^p ($p = 1$ for rearranged distributions and $p = \infty$ for quantiles), equipped with weak* topology, instead of strong topology. Convergence in law of the integral functionals, shown in Proposition 6, is equivalent to the convergence in law of the rearranged estimated processes in such a metric space.

Proposition 6 (Functional limit laws for linear functionals of $\widehat{Q}^*$ and $\widehat{F}$)**.** *Under the same restrictions on the function g as in Proposition 3, the following results hold with the limits being continuous on the specified domains:*

$$1. \quad a_n \int_{\mathcal{Y}} g(y|x,y')(\widehat{F}(y|x) - F(y|x))dy \Rightarrow \int_{\mathcal{Y}} g(y|x,y')D_G(y|x)dy, \tag{2.14}$$

as a stochastic process indexed by $(y',x) \in \mathcal{Y}\mathcal{X}$, in $\ell^\infty(\mathcal{Y}\mathcal{X})$.

$$2. \quad a_n \int_{\mathcal{U}} g(u|x,u')(\widehat{Q}^*(u|x) - Q^*(u|x))du \Rightarrow \int_{\mathcal{U}} g(u|x,u')\tilde{D}_G(u|x)du, \tag{2.15}$$

as a stochastic process indexed by $(u',x) \in \mathcal{U}\mathcal{X}$, in $\ell^\infty(\mathcal{U}\mathcal{X})$.

The linear functionals defined above are useful building blocks for various statistics, such as partial means, various moments, and Lorenz curves. For example, the conditional Lorenz curve based on rearranged quantile functions is

$$\widehat{L}(u'|x) := \Big(\int_{\mathcal{U}} 1\{u \le u'\}\widehat{Q}^*(u|x)du\Big) / \Big(\int_{\mathcal{U}} \widehat{Q}^*(u|x)du\Big), \tag{2.16}$$

which is a ratio of partial and overall conditional means. Hadamard differentiability of the mapping

$$Q \mapsto L(u'|x) := \Big(\int_{\mathcal{U}} 1\{u \le u'\}Q^*(u|x)du\Big) / \Big(\int_{\mathcal{U}} Q^*(u|x)du\Big), \tag{2.17}$$

with respect to Q immediately follows from (a) the differentiability of a ratio β/γ with respect to its numerator β and denominator γ at $\gamma \neq 0$, (b) Hadamard differentiability of the numerator and denominator in (2.17) with respect to Q established in Proposition 6, and (c) the chain rule for the functional delta method. Hence, provided that $\int_{\mathcal{U}} Q^*(u|x)du \neq 0$, we have that in the metric space $\ell^\infty(\mathcal{U}\mathcal{X})$

$$a_n(\widehat{L}(u'|x) - L(u'|x)) \Rightarrow L(u'|x) \cdot \left(\frac{\int_{\mathcal{U}} 1\{u \le u'\}\tilde{D}_G(u|x)du}{\int_{\mathcal{U}} 1\{u \le u'\}Q^*(u|x)du} - \frac{\int_{\mathcal{U}} \tilde{D}_G(u|x)du}{\int_{\mathcal{U}} Q^*(u|x)du} \right), \tag{2.18}$$

as an empirical process indexed by $(u',x) \in \mathcal{U}\mathcal{X}$. In particular, validity of the bootstrap for estimating this functional limit law in (2.18) holds by the functional delta method for the bootstrap.

We next consider the empirical properties of the smoothed curves obtained by applying the linear smoothing operator S defined in (2.10) to $\widehat{F}$ and $\widehat{Q}^*$:

$$S\widehat{F}(y'|x) := \int k_\delta(y'-y)\widehat{F}(y|x)dy, \qquad S\widehat{Q}^*(u'|x) := \int k_\delta(u'-u)\widehat{Q}^*(u|x)du.$$

The following corollary immediately follows from Corollary 2 and the functional delta method.

Corollary 4 (Functional limit laws for smoothed $\widehat{Q}^*$ and $\widehat{F}$)**.** *In $\ell^\infty(\mathcal{Y}\mathcal{X})$,*

$$a_n(S\widehat{F}(y'|x) - SF(y'|x)) \Rightarrow SD_G(y'|x), \tag{2.19}$$

as a stochastic process indexed by $(y', x) \in \mathcal{YX}$*, and in* $\ell^\infty(\mathcal{UX})$*,*

$$a_n(S\widehat{Q}^*(u'|x) - SQ^*(u'|x)) \Rightarrow S\widetilde{D}_G(u'|x), \tag{2.20}$$

as a stochastic process indexed by $(u', x) \in \mathcal{UX}$*. The results hold uniformly in the smoothing parameter* $\delta \in [\delta_1, \delta_2]$*, where* δ_1 *and* δ_2 *are positive constants.*

Thus, as in the case of linear functionals, we can perform inference on SQ^* based on the smoothed rearranged estimates without excluding nonregular neighborhoods, which is convenient for practice in the non-monotonic case. Furthermore, validity of the bootstrap for the smoothed curves follows by the functional delta method for the bootstrap. Lastly, we note that it is not possible to simultaneously allow $\delta \to 0$ and preserve the uniform convergence stated in the corollary.

Our final corollary asserts validity of the bootstrap for inference on rearranged estimators and their functionals. This corollary follows from the functional delta method for the bootstrap (e.g., Theorem 13.9 in van der Vaart, 1998).

Corollary 5 (Validity of the bootstrap for estimating laws of rearranged estimators)**.** *If the bootstrap consistently estimates the functional limit law (2.11) of the empirical process* $\{a_n(\widehat{Q}(u|x) - Q(u|x), (u, x) \in \mathcal{UX}\}$*, then it also consistently estimates the functional limit laws (2.12), (2.13), (2.14), (2.15), (2.19), and (2.20).*

3. Examples

In this section we apply rearrangement to the estimation of structural quantile and distribution functions. We show how rearrangement monotonizes instrumental quantile and distribution function estimates, and demonstrate how to perform inference on the target functions using the results developed in this paper. Using a supporting numerical example, we show that rearranged estimators noticeably improve upon original estimators and also outperform isotonized estimators. Thus, rearrangement is necessarily preferable to the standard approach of simply ignoring non-monotonicity. Moreover, in quantile estimation problems, rearrangement is also preferable to the standard approach of isotonization used primarily in mean estimation problems.

3.1. **Empirical Example.** We consider estimation of the causal/structural effects of Vietnam veteran status $X \in \{0, 1\}$ in the quantiles and distribution of civilian earnings Y. Since veteran status is likely to be endogenous relative to potential civilian earnings, we employ an instrumental variables approach, using the U.S. draft lottery as an instrument for the Vietnam status (Angrist, 1990). We use the same data subset from the Current Population

Survey as in Abadie (2002).[10] We then estimate structural quantile and distribution functions with the instrumental quantile regression estimator of Chernozhukov and Hansen (2005, 2006) and the instrumental distribution regression estimator of Abadie (2002). Under some assumptions these procedures consistently estimate the structural quantile and distribution functions of interest.[11] However, like most estimation methods mentioned in the introduction, neither of these procedures explicitly imposes monotonicity of the distribution and quantile functions. Accordingly, they can produce estimates in finite samples that are nonmonotonic due to either sampling variation or violations of instrument independence or other modeling assumptions. We monotonize these estimates using rearrangement and perform inference on the target structural functions using uniform confidence bands constructed via bootstrap. We use the programming language R to implement the procedures (R Development Core Team, 2007). We present our estimation and inference results in Figures 3–5.

In Figure 3, we show Abadie's estimates of the structural distribution of earnings for veterans and non-veterans (left panel) as well as their rearrangements (right panel). For both veterans and non-veterans, the original estimates of the distributions exhibit clear local non-monotonicity. The rearrangement fixes this problem producing increasing estimated distribution functions. In Figure 4, we show Chernozhukov and Hansen's estimates of the structural quantile functions of earnings for veterans (left panel) as well as their rearrangements (right panel). For both veterans and non-veterans, the estimates of the quantile functions exhibit pronounced local non-monotonicity. The rearrangement fixes this problem producing increasing estimated quantile functions . In the case of quantile functions, the nonmonotonicity problem is specially acute for the small sample of veterans.

In Figure 5, we plot uniform 90% confidence bands for the structural quantile functions of earnings for veterans and non-veterans, together with uniform 90% confidence bands for the effect of Vietnam veteran status on the quantile functions for earnings, which measures the difference between the structural quantile functions for veterans and non-veterans. We construct the uniform confidence bands using both the original estimators and the rearranged estimators based on 500 bootstrap repetitions and a fine net of quantile indices $\{0.01, 0.02, ..., 0.99\}$. We obtain the bands for the rearranged functions assuming that the population structural quantile regression functions are monotonic, so that the first order behavior of the rearranged estimators

[10]These data consist of a sample of white men, born in 1950–1953, from the March Current Population Surveys of 1979 and 1981-1985. The data include annual labor earnings, the Vietnam veteran status and an indicator on the Vietnam era lottery. There are 11,637 men in the sample, with 2,461 Vietnam veterans and 3,234 eligible for U.S. military service according to the draft lottery indicator. Abadie (2002) gives additional information on the data and the construction of the variables.

[11]More specifically, Abadie's (2002) procedure consistently estimates these functions for the subpopulation of compliers under instrument independence and monotonicity. Chernozhukov and Hansen's (2005, 2006) approach consistently estimates these functions for the entire population under instrument independence and rank similarity.

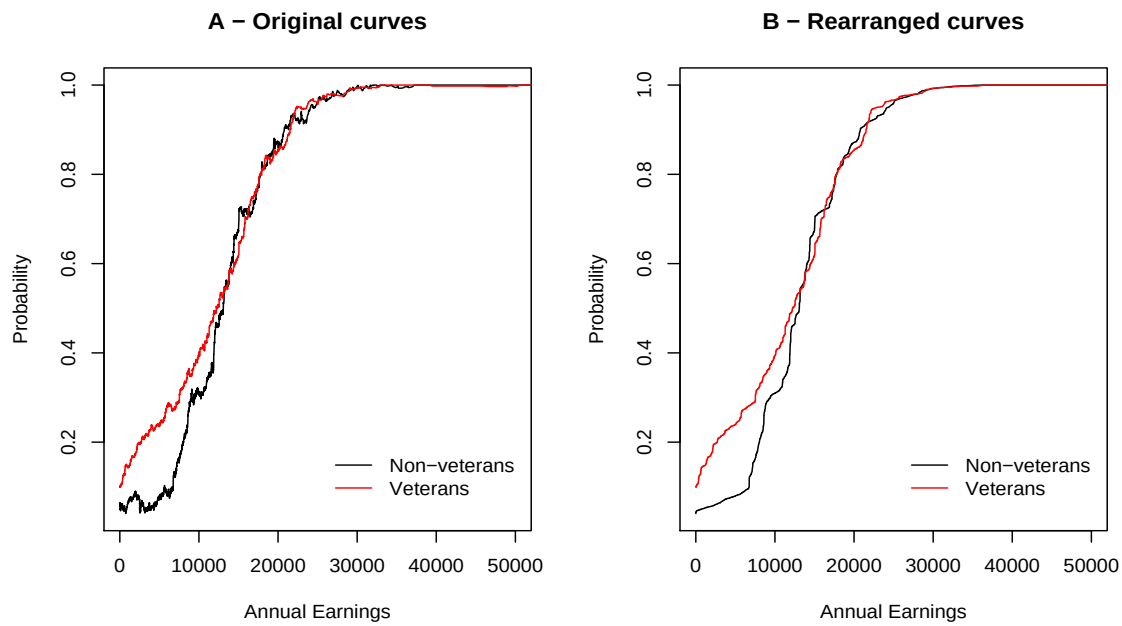

FIGURE 3. Abadie's estimates of the structural distributions of earnings for veteran and non-veterans (left panel), and their rearrangements (right panel).

coincides with the behavior of the original estimators. The figure shows that even for the large sample of non-veterans the rearranged estimates lie within the original bands, thus passing our automatic test of monotonicity specified in Remark 2. Thus, the lack of monotonicity of the estimated quantile functions in this case is likely caused by sampling error. From the figure, we conclude that veteran status has a statistically significant negative effect in the lower tail, with the bands for the rearranged estimates showing a wider range of quantile indices for which this holds.

3.2. **Monte Carlo.** We design a Monte Carlo experiment to closely match the previous empirical example. In particular, we consider a location model, where the outcome is $Y = [1, X]\alpha + \epsilon$, the endogenous regressor is $X = 1\{[1; Z]\pi + v \geq 0\}$, the instrument Z is a binary random variable, and the disturbances (ϵ, v) are jointly normal and independent of Z. The true structural quantile functions are $Q_0(u|x) = [1; x]\alpha + Q_\epsilon(u),\ x \in \{0, 1\}$, where Q_ϵ is the quantile function of the normal variable ϵ. The corresponding structural distribution functions are the inverse of the quantile functions with respect to u. We select the value of the parameters by estimating this location model parametrically by maximum likelihood, and then generate samples from the estimated model, holding the values of the instruments Z equal to those in the data set.[12]

[12]More specifically, after normalizing the standard deviation of v to one, we set $\pi = [-.92; .40]^\mathsf{T}$, $\alpha = [11,753; -911]^\mathsf{T}$, the standard deviation of ϵ to $8,100$, and the covariance between ϵ and v to 379. We draw

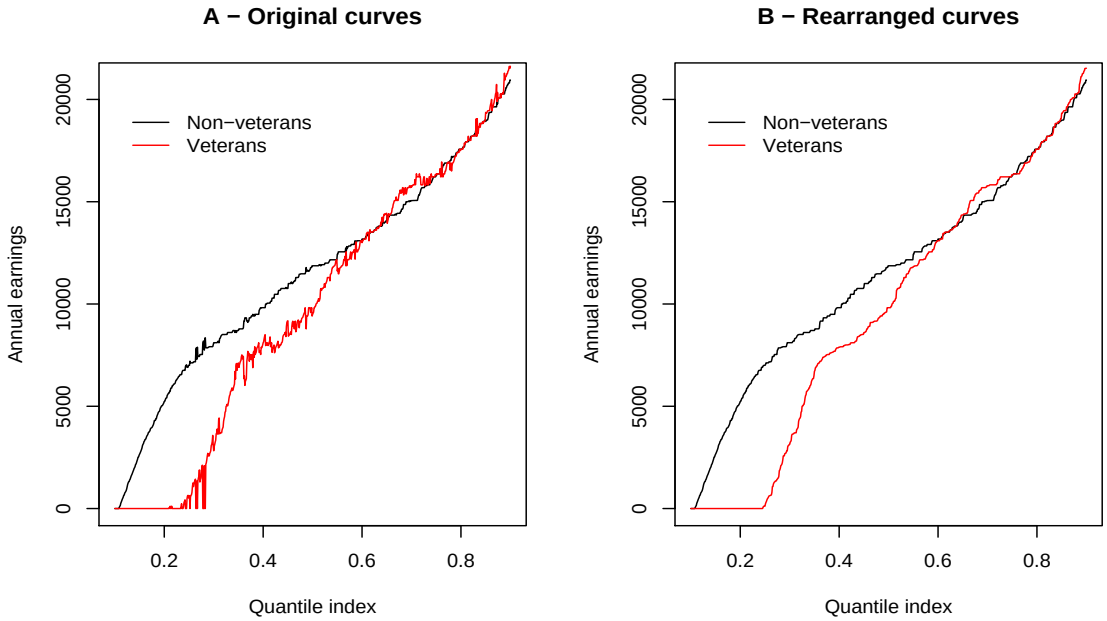

FIGURE 4. Chernozhukov and Hansen's estimates of the structural quantile functions of earnings for veterans (left panel), and their rearrangements (right panel).

We use the estimators for the structural distribution and quantile functions described in the previous section. We monotonize the estimates using either rearrangement or isotonization. We use isotonization as a benchmark since it is the standard approach in mean regression problems (Mammen, 1991); it amounts to projecting the estimated function on the set of monotone functions.

Table 1 reports ratios of estimation errors of the rearranged and isotonized estimates to those of the original estimates, recorded in percentage terms. The target functions are the structural distribution and quantile functions. We measure estimation errors using the average L^p norms $\|\cdot\|_p$ with $p = 1, 2$, and ∞, and we compute them as Monte Carlo averages of $\|f_0 - \tilde{f}\|_p$, where f_0 is the target function, and $\tilde{f}$ is either the original or rearranged or isotonized estimate of this function.

We find that the rearranged estimators noticeably outperform the original estimators, achieving a reduction in estimation error up to 14%, depending on the target function and the norm. Moreover, in this case the better approximation of the rearranged estimates to the structural

$5,000$ Monte Carlo samples of size $n = 11,627$. We generate the values of Y and X by drawing disturbances (ϵ, v) from a bivariate normal distribution with zero mean and the estimated covariance matrix.

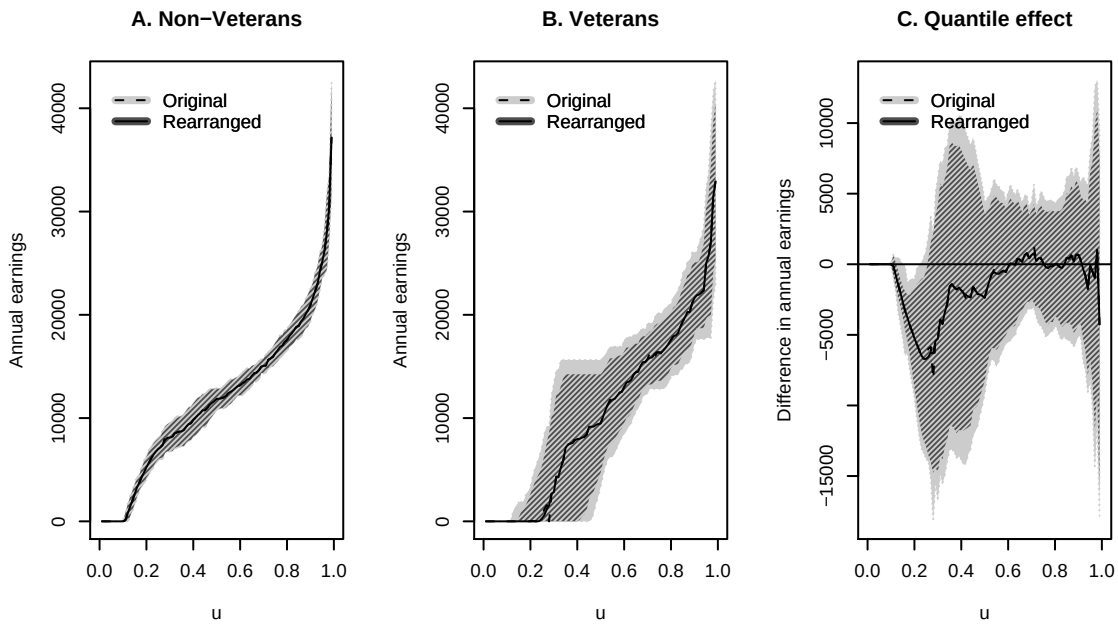

FIGURE 5. Simultaneous 90% confidence bands for structural quantile functions of earnings and structural quantile effect of Vietnam veteran status on earnings. The bands for the quantile functions are intersected with the class of monotone functions.

functions also produces more accurate estimates of the distribution and quantile effects, achieving a 3% to 9% reduction in estimation error for the distribution estimator and a 3% to 14% reduction in estimation error for the quantile estimator, depending on the norm.

We also find that the rearranged estimators noticeably outperform the isotonized estimators, achieving up to a further 4% reduction in estimation error, depending on the target function and the norm. The reason is that isotonization projects the original fitted function on the set of monotone functions, finding the flattest fit in this set. In contrast, rearrangement sorts the original fitted function, finding the steepest fit that preserves measure. In the context of estimating quantile and distribution functions, the target functions tend to be non-flat, suggesting that rearrangement should be typically preferred over isotonization.[13]

[13]To give some intuition about this point, it is instructive to consider a simple example with a two-point domain $\{0,1\}$. Suppose that the target function $f_0 : \{0,1\} \to \mathbb{R}$ is increasing, and steep, namely $f_0(0) > f_0(1)$, and the fitted function $\widehat{f} : \{0,1\} \to \mathbb{R}$ is decreasing, with $\widehat{f}(0) > \widehat{f}(1)$. In this case, isotonization produces a nondecreasing function $\bar{f} : \{0,1\} \to \mathbb{R}$, which is flat, with $\bar{f}(0) = \bar{f}(1) = [\widehat{f}(0) + \widehat{f}(1)]/2$, and is somewhat unsatisfactory. In such cases rearrangement can significantly outperform isotonization, since it produces the steepest fit, namely it produces $f^* : \{0,1\} \to \mathbb{R}$ with $f^*(0) = \widehat{f}(1) < f^*(1) = \widehat{f}(0)$. This observation provides a simple theoretical underpinning for the estimation results we see in Table 1.

Table 1. Ratios of estimation error of rearranged and isotonic estimators to those of original estimators, in percentage terms.

	Veterans		Non-Veterans		Effect	
	Rearranged	Isotonized	Rearranged	Isotonized	Rearranged	Isotonized
			Distribution function			
L^1	99	99	97	98	97	98
L^2	99	99	97	98	97	99
L^∞	96	98	90	94	91	95
			Quantile function			
L^1	97	98	100	100	97	98
L^2	96	97	100	100	96	98
L^∞	86	87	98	99	86	88

4. Conclusion

This paper develops a monotonization procedure for estimation of conditional and structural quantile and distribution functions based on rearrangement-related operations. Starting from a possibly non-monotone empirical curve, the procedure produces a rearranged curve that not only satisfies the natural monotonicity requirement, but also has smaller estimation error than the original curve. We derive asymptotic distribution theory for the rearranged curves, and illustrate the usefulness of the approach with an empirical application and a simulation example. There are many more potential applications of the results of the paper to other econometric problems with shape restrictions (see e.g. Matzkin, 1994, and Chernozhukov et al., 2006).

Appendix A. Proofs

A.1. **Proof of Proposition 1.** First, note that the distribution of Y_x has no atoms, i.e.,

$$\Pr[Y_x = y] = \Pr[Q(U|x) = y] = \Pr[U \in \{u \in \mathcal{U} : u \text{ is a root of } Q(u|x) = y\}] = 0,$$

since the number of roots of $Q(u|x) = y$ is finite under (a) - (b), and $U \sim \text{Uniform}(\mathcal{U})$. Next, by assumptions (a)-(b) the number of critical values of $Q(u|x)$ is finite, hence claim (1) follows.

Next, for any regular y, we can write $F(y|x)$ as

$$\int_0^1 1\{Q(u|x) \le y\}du = \sum_{k=0}^{K(y|x)-1} \int_{u_k(y|x)}^{u_{k+1}(y|x)} 1\{Q(u|x) \le y\}du + \int_{u_{K(y|x)}(y|x)}^1 1\{Q(u|x) \le y\}du,$$

where $u_0(y|x) := 0$ and $\{u_k(y|x), \text{ for } k = 1, 2, ..., K(y|x) < \infty\}$ are the roots of $Q(u|x) = y$ in increasing order. Note that the sign of $\partial_u Q(u|x)$ alternates over consecutive $u_k(y|x)$, determining whether $1\{Q(y|x) \le y\} = 1$ on the interval $[u_{k-1}(y|x), u_k(y|x)]$. Hence the first

term in the previous expression simplifies to $\sum_{k=0}^{K(y|x)-1} 1\{\partial_u Q(u_{k+1}(y|x)|x) \geq 0\}(u_{k+1}(y|x) - u_k(y|x))$; while the last term simplifies to $1\{\partial_u Q(u_{K(y|x)}(y|x)|x) \leq 0\}(1 - u_{K(y|x)}(y|x))$. An additional simplification yields the expression given in claim (2) of the proposition.

The proof of claim (3) follows by taking the derivative of expression in claim (2), noting that at any regular value y the number of solutions $K(y|x)$ and $\text{sign}(\partial_u Q(u_k(y|x)|x))$ are locally constant; moreover,

$$\partial_y u_k(y|x) = \frac{\text{sign}(\partial_u Q(u_k(y|x)|x))}{|\partial_u Q(u_k(y|x)|x)|}.$$

Combining these facts we get the expression for the derivative given in claim (3).

To show the absolute continuity of F with f being the Radon-Nykodym derivative, it suffices to show that for each $y' \in \mathcal{Y}_x$, $\int_{-\infty}^{y'} f(y|x)dy = \int_{-\infty}^{y'} dF(y|x)$, cf. Theorem 31.8 in Billingsley (1995). Let V_t^x be the union of closed balls of radius t centered on the critical points $\mathcal{Y}_x \backslash \mathcal{Y}_x^*$, and define $\mathcal{Y}_x^t = \mathcal{Y}_x \backslash V_t^x$. Then, $\int_{-\infty}^{y'} 1\{y \in \mathcal{Y}_x^t\} f(y|x)dy = \int_{-\infty}^{y'} 1\{y \in \mathcal{Y}_x^t\} dF(y|x)$. Since the set of critical points $\mathcal{Y}_x \backslash \mathcal{Y}_x^*$ is finite and has mass zero under F, $\int_{-\infty}^{y'} 1\{y \in \mathcal{Y}_x^t\} dF(y|x) \uparrow \int_{-\infty}^{y'} dF(y|x)$ as $t \to 0$. Therefore, $\int_{-\infty}^{y'} 1\{y \in \mathcal{Y}_x^t\} f(y|x)dy \uparrow \int_{-\infty}^{y'} f(y|x)dy = \int_{-\infty}^{y'} dF(y|x)$.

Claim (4) follows by noting that at the regions where $s \to Q(s|x)$ is increasing and one-to-one, we have that $F(y|x) = \int_{Q(s|x) \leq y} ds = \int_{s \leq Q^{-1}(y|x)} ds = Q^{-1}(y|x)$. Inverting the equation $u = F(Q^*(u|x)|x) = Q^{-1}(Q^*(u|x)|x)$ yields $Q^*(u|x) = Q(u|x)$.

Claim (5). We have $Y_x = Q(U|x)$ has quantile function Q^*. A quantile function is known to be equivariant to monotone increasing transformations, including location-scale transformations. Thus, this is true in particular for Q^*.

Claim (6) is immediate from claim (3).

Claim (7). The proof of continuity of F is subsumed in the step 1 of the proof of Proposition 3 (see below). Therefore, for any sequence $x_t \to x$ we have that $F(y|x_t) \to F(y|x)$ uniformly in y, and F is continuous. Let $u_t \to u$ and $x_t \to x$. Since $F(y|x) = u$ has a unique root $y = Q^*(u|x)$, the root of $F(y|x_t) = u_t$, i.e., $y_t = Q^*(u_t|x_t)$, converges to y by a standard argument, see, e.g., van der Vaart and Wellner (1997). □

A.2. **Proof of Propositions 2–6.** In the proofs that follow we will repeatedly use Lemma 1, which establishes the equivalence of continuous convergence and uniform convergence:

Lemma 1. *Let $\mathbb{D}$ and $\mathbb{D}'$ be complete separable metric spaces, with $\mathbb{D}$ compact. Suppose $f : \mathbb{D} \mapsto \mathbb{D}'$ is continuous. Then a sequence of functions $f_n : \mathbb{D} \mapsto \mathbb{D}'$ converges to f uniformly on $\mathbb{D}$ if and only if for any convergent sequence $x_n \to x$ in $\mathbb{D}$ we have that $f_n(x_n) \to f(x)$.*

Proof of Lemma 1: See, for example, Resnick (1987), page 2. □

Proof of Proposition 2.

Part 1. We have that for any $\delta > 0$, there exists $\epsilon > 0$ such that for $u \in B_\epsilon(u_k(y|x))$ and for small enough $t \geq 0$

$$1\{Q(u|x) + th_t(u|x) \leq y\} \leq 1\{Q(u|x) + t(h(u_k(y|x)|x) - \delta) \leq y\},$$

for all $k \in \{1, 2, ..., K(y|x)\}$; whereas for all $u \notin \cup_k B_\epsilon(u_k(y|x))$, as $t \to 0$,

$$1\{Q(u|x) + th_t(u|x) \leq y\} = 1\{Q(u|x) \leq y\}.$$

Therefore,

$$\begin{aligned} &\frac{\int_0^1 1\{Q(u|x) + th_t(u|x) \leq y\}du - \int_0^1 1\{Q(u|x) \leq y\}du}{t} \\ \leq \quad &\sum_{k=1}^{K(y|x)} \int_{B_\epsilon(u_k(y|x))} \frac{1\{Q(u|x) + t(h(u_k(y|x)|x) - \delta) \leq y\} - 1\{Q(u|x) \leq y\}}{t} du, \end{aligned} \tag{A.1}$$

which by the change of variable $y' = Q(u|x)$ is equal to

$$\frac{1}{t} \sum_{k=1}^{K(y|x)} \int_{J_k \cap [y, y - t(h(u_k(y|x)|x) - \delta)]} \frac{1}{|\partial_u Q(Q^{-1}(y'|x)|x)|} dy',$$

where J_k is the image of $B_\epsilon(u_k(y|x))$ under $u \mapsto Q(\cdot|x)$. The change of variable is possible because for ϵ small enough, $Q(\cdot|x)$ is one-to-one between $B_\epsilon(u_k(y|x))$ and J_k.

Fixing $\epsilon > 0$, for $t \to 0$, we have that $J_k \cap [y, y - t(h(u_k(y|x)|x) - \delta)] = [y, y - t(h(u_k(y|x)|x) - \delta)]$, and $|\partial_u Q(Q^{-1}(y'|x)|x)| \to |\partial_u Q(u_k(y|x)|x)|$ as $Q^{-1}(y'|x) \to u_k(y|x)$. Therefore, the right hand term in (A.1) is no greater than

$$\sum_{k=1}^{K(y|x)} \frac{-h(u_k(y|x)|x) + \delta}{|\partial_u Q(u_k(y|x)|x)|} + o(1).$$

Similarly $\sum_{k=1}^{K(y|x)} \frac{-h(u_k(y|x)|x) - \delta}{|\partial_u Q(u_k(y|x)|x)|} + o(1)$ bounds (A.1) from below. Since $\delta > 0$ can be made arbitrarily small, the result follows.

To show that the result holds uniformly in $(y, x) \in K$, a compact subset of $\mathcal{Y}\mathcal{X}^*$, we use Lemma 1. Take a sequence of (y_t, x_t) in K that converges to $(y, x) \in K$, then the preceding argument applies to this sequence, since (1) the function $(y, x) \mapsto -h(u_k(y|x)|x)/|\partial_u Q(u_k(y|x)|x)|$ is uniformly continuous on K, and (2) the function $(y, x) \mapsto K(y|x)$ is uniformly continuous on K. To see (2), note that K excludes a neighborhood of critical points $(\mathcal{Y} \setminus \mathcal{Y}_x^*, x \in \mathcal{X})$, and therefore can be expressed as the union of a finite number of compact sets $(K_1, ..., K_M)$ such that the function $K(y|x)$ is constant over each of these sets, i.e., $K(y|x) = k_j$ for some integer $k_j > 0$, for all $(y, x) \in K_j$ and $j \in \{1, ..., M\}$. Likewise, (1) follows by noting that the limit expression for the derivative is continuous on each of the sets $(K_1, ..., K_M)$ by the assumed continuity of $h(u|x)$ in both arguments, continuity of $u_k(y|x)$ (implied by the Implicit Function Theorem), and the assumed continuity of $\partial_u Q(u|x)$ in both arguments. □

Part 2. For a fixed x the result follows by Part 1 of Proposition 2, by step 1 of the proof below, and by an application of the Hadamard differentiability of the quantile operator shown by Doss and Gill (1992). Step 2 establishes uniformity over $x \in \mathcal{X}$.

Step 1. Let K be a compact subset of $\mathcal{Y}\mathcal{X}^*$. Let (y_t, x_t) be a sequence in K, convergent to a point, say (y, x). Then, for every such sequence, $\epsilon_t := t\|h_t\|_\infty + \|Q(\cdot|x_t) - Q(\cdot|x)\|_\infty + |y_t - y| \to 0$, and

$$\begin{aligned} |F(y_t|x_t, h_t) - F(y|x)| &\leq \Big| \int_0^1 [1\{Q(u|x_t) + th_t(u|x) \leq y_t\} - 1\{Q(u|x) \leq y\}]du \Big| \\ &\leq \Big| \int_0^1 1\{|Q(u|x) - y| \leq \epsilon_t\}du \Big| \to 0, \end{aligned} \tag{A.2}$$

where the last step follows from the absolute continuity of $y \mapsto F(y|x)$, the distribution function of $Q(U|x)$. By setting $h_t = 0$ the above argument also verifies that $F(y|x)$ is continuous in (y, x). Lemma 1 implies uniform convergence of $F(y|x, h_t)$ to $F(y|x)$, which in turn implies by a standard argument[14] the uniform convergence of quantiles $Q^*(u|x, h_t) \to Q^*(u|x)$, uniformly over K^*, where K^* is any compact subset of $\mathcal{U}\mathcal{X}^*$.

Step 2. We have that uniformly over K^*,

$$\begin{aligned} \frac{F(Q^*(u|x, h_t)|x, h_t) - F(Q^*(u|x, h_t)|x)}{t} &= D_h(Q^*(u|x, h_t)|x) + o(1), \\ &= D_h(Q^*(u|x)|x) + o(1), \end{aligned} \tag{A.3}$$

using Step 1, Proposition 2, and the continuity properties of $D_h(y|x)$. Further, uniformly over K^*, by Taylor expansion and Proposition 1, as $t \to 0$,

$$\frac{F(Q^*(u|x, h_t)|x) - F(Q^*(u|x)|x)}{t} = f(Q^*(u|x)|x)\frac{Q^*(u|x, h_t) - Q^*(u|x)}{t} + o(1), \tag{A.4}$$

and (as will be shown below)

$$\frac{F(Q^*(u|x, h_t)|x, h_t) - F(Q^*(u|x)|x)}{t} = o(1), \tag{A.5}$$

as $t \to 0$. Observe that the left hand side of (A.5) equals that of (A.4) plus that of (A.3). The result then follows.

It only remains to show that equation (A.5) holds uniformly in K^*. Note that for any right-continuous cdf F, we have that $u \leq F(Q^*(u)) \leq u + F(Q^*(u)) - F(Q^*(u)-)$, where $F(\cdot-)$ denotes the left limit of F, i.e., $F(x_0-) = \lim_{x \uparrow x_0} F(x)$. For any continuous, strictly increasing

[14] See, e.g., Lemma 1 in Chernozhukov and Fernandez-Val (2005).

cdf F, we have that $F(Q^*(u)) = u$. Therefore, write

$$\begin{aligned}
0 &\leq \frac{F(Q^*(u|x,h_t)|x,h_t) - F(Q^*(u|x)|x)}{t} \\
&\leq \frac{u + F(Q^*(u|x,h_t)|x,h_t) - F(Q^*(u|x,h_t) - |x,h_t) - u}{t} \\
&\leq \frac{F(Q^*(u|x,h_t)|x,h_t) - F(Q^*(u|x,h_t) - |x,h_t)}{t} \\
&\overset{(1)}{=} \frac{[F(Q^*(u|x,h_t)|x,h_t) - F(Q^*(u|x,h_t)|x)]}{t} \\
&- \frac{[F(Q^*(u|x,h_t) - |x,h_t) - F(Q^*(u|x,h_t) - |x)]}{t} \\
&\overset{(2)}{=} D_h(Q^*(u|x,h_t)|x) - D_h(Q^*(u|x,h_t) - |x) + o(1) = o(1),
\end{aligned}$$

as $t \to 0$, where in (1) we use that $F(Q^*(u|x,h_t)|x) = F(Q^*(u|x,h_t) - |x)$ since $F(y|x)$ is continuous and strictly increasing in y, and in (2) we use Proposition 2. □

The following lemma, due to Pratt (1960), will be very useful to prove Proposition 4.

Lemma 2. *Let $|f_n| \leq G_n$ and suppose that $f_n \to f$ and $G_n \to G$ almost everywhere, then if $\int G_n \to \int G$ finite, then $\int f_n \to \int f$.*

Proof of Lemma 2. See Pratt (1960). □

Lemma 3 (Boundedness and Integrability Properties)**.** *Under the hypotheses of Proposition 2, we have that, for all $(u,x) \in \mathcal{UX}$,*

$$|\widetilde{D}_{h_t}(u|x,t)| \leq \|h_t\|_\infty, \tag{A.6}$$

and, for all $(y,x) \in \mathcal{YX}$,

$$|D_{h_t}(y|x,t)| \leq \Delta(y|x,t) = \int_0^1 \frac{1\{|Q(u|x) - y| \leq t\|h_t\|_\infty\}}{t} du, \tag{A.7}$$

where for any $x_t \to x \in \mathcal{X}$, as $t \to 0$,

$$\Delta(y|x_t,t) \to 2\|h\|_\infty f(y|x) \text{ for a.e } y \in \mathcal{Y} \text{ and } \int_{\mathcal{Y}} \Delta(y|x_t,t)dy \to \int_{\mathcal{Y}} 2\|h\|_\infty f(y|x)dy.$$

Proof of Lemma 3. To show (A.6) note that

$$\sup_{(u,x)\in\mathcal{UX}} |\widetilde{D}_{h_t}(u|x,t)| \leq \|h_t\|_\infty \tag{A.8}$$

immediately follows from the equivariance property noted in Claim (5) of Proposition 1.

The inequality (A.7) is trivial. That for any $x_t \to x \in \mathcal{X}$, $\Delta(y|x_t,t) \to 2\|h\|_\infty f(y|x)$ for a.e $y \in \mathcal{Y}$ follows by applying Proposition 2 respectively with functions $h_t'(u|x) = \|h_t\|_\infty$ and $h_t'(u|x) = -\|h_t\|_\infty$ (for the case when $f(y|x) > 0$; and trivially otherwise). Similarly, that for any $y_t \to y \in \mathcal{Y}$, $\Delta(y_t|x,t) \to 2\|h\|_\infty f(y|x)$ for a.e $x \in \mathcal{X}$ follows by Proposition 2 (for the case when $f(y|x) > 0$; and trivially otherwise) .

Further, by Fubini's Theorem,

$$\int_{\mathcal{Y}} \Delta(y|x_t,t)dy \quad = \quad \int_0^1 \underbrace{\left(\int_{\mathcal{Y}} \frac{1\{|Q(u|x_t)-y| \leq t\|h_t\|_\infty\}}{t} dy\right)}_{=:\ f_t(u)} du. \tag{A.9}$$

Note that $f_t(u) \leq 2\|h_t\|_\infty$. Moreover, for almost every u, $f_t(u) = 2\|h_t\|_\infty$ for small enough t, and $2\|h_t\|_\infty$ converges to $2\|h\|_\infty$ as $t \to 0$. Then, trivially, $2\int_0^1 \|h_t\|_\infty du \to 2\|h\|_\infty$. By Lemma 2 the right hand side of (A.9) converges to $2\|h\|_\infty$. □

A.3. **Proof of Proposition 3.** Define $m_t(y|x,y') := g(y|x,y')D_{h_t}(y|x,t)$ and $m(y|x,y') := g(y|x,y')D_h(y|x)$. To show claim (1), we need to demonstrate that for any $y_t' \to y'$ and $x_t \to x$

$$\int_{\mathcal{Y}} m_t(y|x_t,y_t')dy \to \int_{\mathcal{Y}} m(y|x,y')dy, \tag{A.10}$$

and that the limit is continuous in (x,y'). We have that $|m_t(y|x_t,y_t)|$ is bounded, for some constant C, by $C\Delta(y|x_t,t)$ which converges a.e. and the integral of which converges to a finite number by Lemma 3. Moreover, by Proposition 2, for almost every y we have $m_t(y|x_t,y_t') \to m(y|x,y')$. We conclude that (A.10) holds by Lemma 2.

In order to check continuity, we need to show that for any $y_t' \to y'$ and $x_t \to x$

$$\int_{\mathcal{Y}} m(y|x_t,y_t')dy \to \int_{\mathcal{Y}} m(y|x,y')dy. \tag{A.11}$$

We have that $m(y|x_t,y_t') \to m(y|x,y')$ for almost every y. Moreover, $m(y|x_t,y_t)$ is dominated by $2\|g\|_\infty\|h\|_\infty f(y|x_t)$, which converges to $2\|g\|_\infty\|h\|_\infty f(y|x)$ for almost every y, and, moreover, $\int_{\mathcal{Y}} \|g\|_\infty\|h\|_\infty f(y|x)dy$ converges to $\|g\|_\infty\|h\|_\infty$. Conclude that (A.11) holds by Lemma 2.

To show claim (2), define $m_t(u|x,u') = g(u|x,u')\tilde{D}_{h_t}(u|x)$ and $m(u|x,u') = g(u|x,u')\tilde{D}_h(u|x)$. Here we need to show that for any $u_t' \to u'$ and $x_t \to x$

$$\int_{\mathcal{U}} m_t(u|x_t,u_t')du \to \int_{\mathcal{U}} m(u|x,u')du, \tag{A.12}$$

and that the limit is continuous in (u',x). We have that $m_t(u|x_t,u_t')$ is bounded by $g(u|x_t)\|h_t\|_\infty$, which converges to $g(u|x)\|h\|_\infty$ for a.e. u. Furthermore, the integral of $g(u|x_t)\|h_t\|_\infty$ converges to the integral of $g(u|x)\|h\|_\infty$ by the dominated convergence theorem. Moreover, by Proposition 2, we have that $m_t(u|x_t,u_t') \to m(u|x,u')$ for almost every u. We conclude that (A.12) holds by Lemma 2.

In order to check the continuity of the limit, we need to show that for any $u_t' \to u'$ and $x_t \to x$

$$\int_{\mathcal{U}} m(u|x_t,u_t')du \to \int_{\mathcal{U}} m(u|x,u')du. \tag{A.13}$$

We have that $m(u|x_t,u_t') \to m(u|x,u')$ for almost every u. Moreover, for small enough t, $m(u|x_t,u_t')$ is dominated by $|g(u|x_t,u_t')|\|h\|_\infty$, which converges for almost every value of u to

$|g(u|x,u')|\|h\|_\infty$ as $t \to 0$. Furthermore, the integral of $|g(u|x_t,u'_t)|\|h\|_\infty$ converges to the integral of $|g(u|x,u')|\|h\|_\infty$ by the dominated convergence theorem. We conclude that (A.13) holds by Lemma 2. □

A.4. **Proof of Proposition 5.** This proposition simply follows by the functional delta method (e.g., van der Vaart, 1998). Instead of restating what this method is, it takes less space to simply recall the proof in the current context.

To show the first part, consider the map $g_n(y,x|h) = a_n(F(y|x,h/a_n) - F(y|x))$. The sequence of maps satisfies $g_{n'}(y,x|h_{n'}) \to D_h(y|x)$ in $\ell^\infty(K)$ for every subsequence $h_{n'} \to h$ in $\ell^\infty(\mathcal{UX})$, where h is continuous. It follows by the extended continuous mapping theorem that, in $\ell^\infty(K)$, $g_n(y,x|a_n(\widehat{Q}(u|x) - Q(u|x))) \Rightarrow D_G(y|x)$ as a stochastic process indexed by (y,x), since $a_n(\widehat{Q}(u|x) - Q(u|x)) \Rightarrow G(u|x)$ in $\ell^\infty(\mathcal{UX})$.

Conclude similarly for the second part. □

A.5. **Proof of Proposition 6.** This follows by the functional delta method, similarly to the proof of Proposition 5. □

References

[1] Abadie, A. (2002):"Bootstrap tests for distributional treatment effects in instrumental variable models," *Journal of the American Statistical Association* 457, pp. 284–292.

[2] Abadie, A., Angrist, J., and G. Imbens (2002):"Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica* 70(1), pp. 91–117.

[3] Alvino, A., Lions, P. L., and Trombetti, G. (1989): "On Optimization Problems with Prescribed Rearrangements," *Nonlinear Analysis* 13 (2), pp. 185–220.

[4] Angrist, J. D. (1990): "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review* 80, pp. 313–336.

[5] Angrist, J. D., Chernozhukov, V., and I. Fernandez-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica* 74, pp. 539–563.

[6] Belloni, A., and V. Chernozhukov (2007): "Conditional Quantile and Probability Processes under Increasing Dimension," preprint. MIT.

[7] Billingsley, P. (1995): *Probability and measure*. Third edition. John Wiley & Sons, Inc., New York.

[8] Blundell, R., and J. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Models," in M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (ed.), *Advances in Econometrics, Eight World Congress,* Volume II. Cambridge University Press. Cambridge.

[9] Bronshtein, I. N., Semendyayev, K. A., Musiol, G., and H. Muehlig, H. (2003): *Handbook of Mathematics.* Fourth Edition. Springer-Verlag. Berlin.

[10] Buchinsky, M. (1994): "Changes in the US Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica* 62, pp. 405-458.

[11] Buchinsky, M., and J. Hahn (1998): "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica* 66, no. 3, pp. 653-671.

[12] Carlier, G., and R.-A. Dana (2005): "Rearrangement Inequalities in Non-Convex Insurance Problems," *Journal of Mathematical Economics* 41, 483–503.

[13] Chamberlain, G. (1994): "Quantile Regression, Censoring, and the Structure of Wages," in C. A. Sims (ed.), *Advances in Econometrics, Sixth World Congress,* Volume 1. Cambridge University Press. Cambridge.

[14] Chaudhuri, S. (1991): "Nonparametric estimates of regression quantiles and their local Bahadur representation," *Annals of Statistics* 19(2), 760–777.

[15] Chernozhukov, V., and I. Fernández-Val (2005): "Subsampling Inference on Quantile Regression Processes." *Sankhya* 67, pp. 253–276.

[16] Chernozhukov, V., Fernandez-Val, I., and A. Galichon (2006): "Improving Point and Interval Estimates of Monotone Functions by Rearrangement," forthcoming in *Biometrika*.

[17] Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica* 73, pp. 245–261.

[18] Chernozhukov, V., and C. Hansen (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics* 132(2), 491–525.

[19] Chesher, A., (2003): "Identification in Nonseparable Models," *Econometrica* 71(5), pp. 1405–1441.

[20] Davydov, Y., and R. Zitikis (2005): "An index of monotonicity and its estimation: a step beyond econometric applications of the Gini index," *Metron - International Journal of Statistics*, 63(3), 351–372.

[21] Dette, H., Neumeyer, N., and K. Pilz (2006): "A simple Nonparametric Estimator of a Strictly Monotone Regression Function," *Bernoulli*, 12, no. 3, pp 469-490.

[22] Dette, H., and R. Scheder (2006): "Strictly monotone and smooth nonparametric regression for two or more variables," *The Canadian Journal of Statistics,* 34(4), 535–561.

[23] Dette, H., and S. Volgushev (2008): "Non-Crossing Nonparametric Estimates of Quantile Curves," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70, pp. 609–627.

[24] Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *Annals of Statistics* 2, pp. 267–277.

[25] Doss, H., and R. D. Gill (1992): "An elementary approach to weak convergence for quantile processes, with applications to censored survival data." *Journal of the American Statistical Association* 87, no. 419, 869–877.

[26] Dudley, R. M., and R. Norvaisa (1999): *Differentiability of six operators on nonsmooth functions and p-variation*. Springer-Verlag, Berlin.

[27] Eggermont, P. P. B., and V. N. LaRiccia (2000): "Maximum likelihood estimation of smooth monotone and unimodal densities," *Annals of Statistics* 28(3) , 922–947.

[28] Fougères, A-L. (1997): "Estimation de densités unimodales," *Canad. J. Statist.* 25(3) , 375–387.

[29] Gill, R. D., and S. Johansen (1990): "A survey of product-integration with a view toward application in survival analysis." *Annals of Statistics* 18, no. 4, 1501–1555.

[30] Glad, I. K., Hjort, N. L., and N. G. Ushakov (2003): "Corrections of Density Estimators that are not Densities." *Scandinavian Journal of Statistics* 30, 415–427.

[31] Gutenbrunner, C., and J. Jurečková (1992): "Regression Quantile and Regression Rank Score Process in the Linear Model and Derived Statistics," *Annals of Statistics* 20, pp. 305-330.

[32] Hall, P., Wolff, R., and Yao, Q. (1999): "Methods for estimating a conditional distribution function," *Journal of the American Statistical Association* 94, pp. 154–163.

[33] Hardy, G., Littlewood, J., and G. Polya (1952): *Inequalities*. Cambridge: Cambridge University Press.

[34] He, X. (1997): "Quantile Curves Without Crossing," *American Statistician*, 51, pp. 186–192.

[35] He, X., and Q-M. Shao (2000): "On parameters of increasing dimensions," *Journal of Multivariate Analysis* 73(1) , 120–135.

[36] Hendricks, W., and R. Koenker (1991): "Hierchical Spline Models for Conditional Quantiles and the Demand for Electricity," *Journal of the American Statistical Association* 87 , 58–68.

[37] Honore, B., Khan, S., and J. Powell (2002): "Quantile Regression under Random Censoring," *Journal of Econometrics* 109(1) , 67–105.

[38] Imbens, G.W., and W. K. Newey (2001): "Identification and Estimation of Nonseparable Triangular Simultaneous Equations Models Without Additivity," *Econometrica,* forthcoming.

[39] Koenker, R. (1994): "Confidence Intervals for Regression Quantiles," in M.P. and M. Hušková (eds.), *Asymptotic Statistics: Proceeding of the 5th Prague Symposium on Asymptotic Statistics*. Physica-Verlag.

[40] Koenker, R. (2005): *Quantile Regression*. Econometric Society Monograph Series 38, Cambridge University Press.

[41] Koenker, R. (2007). *quantreg: Quantile Regression.* R package version 4.10. `http://www.r-project.org`.

[42] Koenker, R. and G. Bassett (1978): "Regression Quantiles," *Econometrica* 46, pp. 33-50.

[43] Koenker, R., and L. Ma (2006): "Quantile regression methods for recursive structural equation models." *Journal of Econometrics* 134, no. 2, 471–506.

[44] Koenker, R., and P. Ng (2005): "Inequality constrained quantile regression." *Sankhyā* 67, no. 2, 418–440.

[45] Lehmann, E. (1974): *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day.

[46] Lorentz, G. G. (1953): "An Inequality for Rearrangements," *The American Mathematical Monthly* 60, pp. 176–179.

[47] Mammen, E. (1991): "Nonparametric Regression Under Qualitative Smoothness Assumptions," *Annals of Statistics* 19, pp. 741–759.

[48] Matzkin, R. L. (1994): "Restrictions of economic theory in nonparametric methods," in *Handbook of econometrics*, Vol. IV, North-Holland, Amsterdam.

[49] Milnor, J. (1965): *Topology from the differential viewpoint*, Princeton University Press.

[50] Mossino J. and R. Temam (1981): "Directional derivative of the increasing rearrangement mapping and application to a queer differential equation in plasma physics," *Duke Math. J.* 48 (3), 475–495.

[51] Portnoy, S. (1991): "Asymptotic behavior of regression quantiles in nonstationary, dependent cases," *Journal of Multivariate Analysis* 38 , no. 1, 100–113.

[52] Powell, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics* 32, no. 1, pp. 143-155.

[53] Pratt, J.W. (1960): "On interchanging limits and integrals." *Annals of Mathematical Statistics* 31, 74–77.

[54] R Development Core Team (2007): *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`.

[55] Resnick, S. I. (1987): *Extreme values, regular variation, and point processes*, Applied Probability. A Series of the Applied Probability Trust, 4. Springer-Verlag, New York.

[56] Robertson, T., Wright, F.T., and R.L. Dykstra (1988): *Order restricted statistical inference*, Chichester: Wiley.

[57] Van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, 3.

[58] Van der Vaart, A., and J. Wellner (1996): *Weak convergence and empirical processes: with applications to statistics*, New York: Springer.

[59] Villani, C. (2003): *Topics in Optimal Transportation*, Providence: American Mathematical Society.