Chernozhukov, Victor; Fernández-Val, Iván; Galichon, Alfred

Working Paper

Improving point and interval estimates of monotone functions by rearrangement

cemmap working paper, No. CWP17/08

Provided in Cooperation with:
The Institute for Fiscal Studies (IFS), London

Suggested Citation: Chernozhukov, Victor; Fernández-Val, Iván; Galichon, Alfred (2008) : Improving point and interval estimates of monotone functions by rearrangement, cemmap working paper, No. CWP17/08, Centre for Microdata Methods and Practice (cemmap), London, https://doi.org/10.1920/wp.cem.2008.1708

This Version is available at:
https://hdl.handle.net/10419/64669

cemmap
centre for microdata methods and practice

Improving point and interval estimates of monotone functions by rearrangement

Victor Chernozhukov
Iván Fernández-Val
Alfred Galichon

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP17/08

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

IMPROVING POINT AND INTERVAL ESTIMATES OF MONOTONE FUNCTIONS BY REARRANGEMENT

VICTOR CHERNOZHUKOV[†] IVÁN FERNÁNDEZ-VAL[§] ALFRED GALICHON[‡]

ABSTRACT. Suppose that a target function $f_0 : \mathbb{R}^d \to \mathbb{R}$ is monotonic, namely weakly increasing, and an original estimate $\hat{f}$ of this target function is available, which is not weakly increasing. Many common estimation methods used in statistics produce such estimates $\hat{f}$. We show that these estimates can always be improved with no harm by using rearrangement techniques: The rearrangement methods, univariate and multivariate, transform the original estimate to a monotonic estimate $\hat{f}^*$, and the resulting estimate is *closer* to the true curve f_0 in common metrics than the original estimate $\hat{f}$. The improvement property of the rearrangement also extends to the construction of confidence bands for monotone functions. Let ℓ and u be the lower and upper endpoint functions of a simultaneous confidence interval $[\ell, u]$ that covers f_0 with probability $1 - \alpha$, then the rearranged confidence interval $[\ell^*, u^*]$, defined by the rearranged lower and upper end-point functions ℓ^* and u^*, is shorter in length in common norms than the original interval and covers f_0 with probability greater or equal to $1 - \alpha$. We illustrate the results with a computational example and an empirical example dealing with age-height growth charts.

KEY WORDS. Monotone function, improved estimation, improved inference, multivariate rearrangement, univariate rearrangement, Lorentz inequalities, growth chart, quantile regression, mean regression, series, locally linear, kernel methods

AMS SUBJECT CLASSIFICATION. Primary 62G08; Secondary 46F10, 62F35, 62P10

Date: First version is of December, 2006. Two ArXiv versions of paper are available, ArXiv: 0704.3686 (April, 2007) and ArXiv: 0806.4730 (June, 2008). This version is of July 12, 2008. The previous title of this paper was "Improving Estimates of Monotone Functions by Rearrangement". We would like to thank the editor, the associate editor, the anonymous referees of the journal, Richard Blundell, Eric Cator, Andrew Chesher, Moshe Cohen, Holger Dette, Emily Gallagher, Wesley Graybill, Piet Groeneboom, Raymond Guiteras, Xuming He, Roger Koenker, Charles Manski, Costas Meghir, Ilya Molchanov, Whitney Newey, Steve Portnoy, Alp Simsek, Jon Wellner and seminar participants at BU, CEMFI, CEMMAP Measurement Matters Conference, Columbia, Columbia Conference on Optimal Transportation, Conference Frontiers of Microeconometrics in Tokyo, Cornell, Cowless Foundation 75th Anniversary Conference, Duke-Triangle, Georgetown, MIT, MIT-Harvard, Northwestern, UBC, UCL, UIUC, University of Alicante, University of Gothenburg Conference "Nonsmooth Inference, Analysis, and Dependence," and Yale University for very useful comments that helped to considerably improve the paper.

[†] Massachusetts Institute of Technology, Department of Economics & Operations Research Center, and University College London, CEMMAP. E-mail: vchern@mit.edu. Research support from the Castle Krob Chair, National Science Foundation, the Sloan Foundation, and CEMMAP is gratefully acknowledged.

§ Boston University, Department of Economics. E-mail: ivanf@bu.edu. Research support from the National Science Foundation is gratefully acknowledged.

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

1. Introduction

A common problem in statistics is the approximation of an unknown monotonic function using an available sample. There are many examples of monotonically increasing functions, including biometric age-height charts, which should be monotonic in age; econometric demand functions, which should be monotonic in price; and quantile functions, which should be monotonic in the probability index. Suppose an original, potentially non-monotonic, estimate is available. Then, the rearrangement operation from variational analysis (Hardy, Littlewood, and Pólya 1952, Lorentz 1953, Villani 2003) can be used to monotonize the original estimate. The rearrangement has been shown to be useful in producing monotonized estimates of density functions (Fougeres 1997), conditional mean functions (Davydov and Zitikis 2005, Dette, Neumeyer, and Pilz 2006, Dette and Pilz 2006, Dette and Scheder 2006), and various conditional quantile and distribution functions (Chernozhukov, Fernandez-Val, and Galichon (2006b, 2006c)).

In this paper, we show, using Lorentz inequalities and their appropriate generalizations, that the rearrangement of the original estimate is not only useful for producing monotonicity, but also has the following important property: The rearrangement always improves upon the original estimate, whenever the latter is not monotonic. Namely, the rearranged curves are always closer (often considerably closer) to the target curve being estimated. Furthermore, this improvement property is generic, i.e., it does not depend on the underlying specifics of the original estimate and applies to both univariate and multivariate cases. The improvement property of the rearrangement also extends to the construction of confidence bands for monotone functions. We show that we can increase the coverage probabilities and reduce the lengths of the confidence bands for monotone functions by rearranging the upper and lower bounds of the confidence bands.

Monotonization procedures have a long history in the statistical literature, mostly in relation to isotone regression. While we will not provide an extensive literature review, we reference the methods other than rearrangement that are most related to the present paper. Mammen (1991) studies the effect of smoothing on monotonization by isotone regression. He considers two alternative two-step procedures that differ in the ordering of the smoothing and monotonization steps. The resulting estimators are asymptotically equivalent up to first order if an optimal bandwidth choice is used in the smoothing step. Mammen, Marron, Turlach, and Wand (2001) show that most smoothing problems, notably including smoothed isotone regression problems, can be recast as a projection problem with respect to a given norm. Another approach is the one-step procedure of Ramsay (1988), which projects on a class of monotone spline functions called I-splines. Later in the paper we will make both analytical and numerical comparisons of these procedures with the rearrangement.

2. Improving Point Estimates of Monotone Functions by Rearrangement

2.1. **Common Estimates of Monotonic Functions.** A basic problem in many areas of statistics is the estimation of an unknown function $f_0 : \mathbb{R}^d \to \mathbb{R}$ using the available information. Suppose we know that the target function f_0 is monotonic, namely weakly increasing. Suppose further that an original estimate $\hat{f}$ is available, which is not necessarily monotonic. Many common estimation methods do indeed produce such estimates. Can these estimates always be improved with no harm? The answer provided by this paper is yes: the rearrangement method transforms the original estimate to a monotonic estimate $\hat{f}^*$, and this estimate is in fact *closer* to the true curve f_0 than the original estimate $\hat{f}$ in common metrics. Furthermore, the rearrangement is computationally tractable, and thus preserves the computational appeal of the original estimates.

Estimation methods, specifically the ones used in regression analysis, can be grouped into global methods and local methods. An example of a global method is the series estimator of f_0 taking the form

$$\hat{f}(x) = P_{k_n}(x)'\hat{b},$$

where $P_{k_n}(x)$ is a k_n-vector of suitable transformations of the variable x, such as B-splines, polynomials, and trigonometric functions. Section 4 lists specific examples in the context of an empirical example. The estimate $\hat{b}$ is obtained by solving the regression problem

$$\hat{b} = \arg\min_{b \in \mathbb{R}^{k_n}} \sum_{i=1}^{n} \rho(Y_i - P_{k_n}(X_i)'b),$$

where $(Y_i, X_i), i = 1, ..., n$ denotes the data. In particular, using the square loss $\rho(u) = u^2$ produces estimates of the conditional mean of Y_i given X_i (Gallant 1981, Andrews 1991, Stone 1994, Newey 1997), while using the asymmetric absolute deviation loss $\rho(u) = (u - 1(u < 0))u$ produces estimates of the conditional u-quantile of Y_i given X_i (Koenker and Bassett 1978, Portnoy 1997, He and Shao 2000). The series estimates $x \mapsto \hat{f}(x) = P_{k_n}(x)'\hat{b}$ are widely used in data analysis due to their desirable approximation properties and computational tractability. However, these estimates need not be naturally monotone, unless explicit constraints are added into the optimization program (see, for example, Matzkin (1994), Silvapulle and Sen (2005), and Koenker and Ng (2005)).

Examples of local methods include kernel and locally polynomial estimators. A kernel estimator takes the form

$$\hat{f}(x) = \arg\min_{b \in \mathbb{R}} \sum_{i=1}^{n} w_i \rho(Y_i - b), \quad w_i = K\left(\frac{X_i - x}{h}\right),$$

where the loss function ρ plays the same role as above, $K(u)$ is a standard, possibly high-order, kernel function, and $h > 0$ is a vector of bandwidths (see, for example, Wand and Jones (1995)

and Ramsay and Silverman (2005)). The resulting estimate $x \mapsto \hat{f}(x)$ need not be naturally monotone. Dette, Neumeyer, and Pilz (2006) show that the rearrangement transforms the kernel estimate into a monotonic one. We further show here that the rearranged estimate necessarily improves upon the original estimate, whenever the latter is not monotonic. The locally polynomial regression is a related local method (Chaudhuri 1991, Fan and Gijbels 1996). In particular, the locally linear estimator takes the form

$$(\hat{f}(x), \hat{d}(x)) = \underset{b\in\mathbb{R}, d\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \rho(Y_i - b - d(X_i - x))^2, \quad w_i = K\left(\frac{X_i - x}{h}\right).$$

The resulting estimate $x \mapsto \hat{f}(x)$ may also be non-monotonic, unless explicit constrains are added to the optimization problem. Section 4 illustrates the non-monotonicity of the locally linear estimate in an empirical example.

In summary, there are many attractive estimation and approximation methods in statistics that do not necessarily produce monotonic estimates. These estimates do have other attractive features though, such as good approximation properties and computational tractability. Below we show that the rearrangement operation applied to these estimates produces (monotonic) estimates that improve the approximation properties of the original estimates by bringing them closer to the target curve. Furthermore, the rearrangement is computationally tractable, and thus preserves the computational appeal of the original estimates.

2.2. **The Rearrangement and its Approximation Property: The Univariate Case.** In what follows, let $\mathcal{X}$ be a compact interval. Without loss of generality, it is convenient to take this interval to be $\mathcal{X} = [0, 1]$. Let $f(x)$ be a measurable function mapping $\mathcal{X}$ to K, a bounded subset of $\mathbb{R}$. Let $F_f(y) = \int_{\mathcal{X}} 1\{f(u) \leq y\}du$ denote the distribution function of $f(X)$ when X follows the uniform distribution on $[0, 1]$. Let

$$f^*(x) := Q_f(x) := \inf\{y \in \mathbb{R} : F_f(y) \geq x\}$$

be the quantile function of $F_f(y)$. Thus,

$$f^*(x) := \inf\left\{y \in \mathbb{R} : \left[\int_{\mathcal{X}} 1\{f(u) \leq y\}du\right] \geq x\right\}.$$

This function f^* is called the increasing rearrangement of the function f.

Thus, the rearrangement operator simply transforms a function f to its quantile function f^*. That is, $x \mapsto f^*(x)$ is the quantile function of the random variable $f(X)$ when $X \sim U(0, 1)$. It is also convenient to think of the rearrangement as a sorting operation: given values of the function $f(x)$ evaluated at x in a fine enough net of equidistant points, we simply sort the values in an increasing order. The function created in this way is the rearrangement of f.

The first main point of this paper is the following:

Proposition 1. *Let $f_0 : \mathcal{X} \to K$ be a weakly increasing measurable function in x. This is the target function. Let $\hat{f} : \mathcal{X} \to K$ be another measurable function, an initial estimate of the target function f_0.*

1. For any $p \in [1, \infty]$, the rearrangement of $\hat{f}$, denoted $\hat{f}^*$, weakly reduces the estimation error:

$$\left[\int_{\mathcal{X}} \left|\hat{f}^*(x) - f_0(x)\right|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}} \left|\hat{f}(x) - f_0(x)\right|^p dx\right]^{1/p}. \tag{2.1}$$

2. Suppose that there exist regions $\mathcal{X}_0$ and $\mathcal{X}_0'$, each of measure greater than $\delta > 0$, such that for all $x \in \mathcal{X}_0$ and $x' \in \mathcal{X}_0'$ we have that (i) $x' > x$, (ii) $\hat{f}(x) > \hat{f}(x') + \epsilon$, and (iii) $f_0(x') > f_0(x) + \epsilon$, for some $\epsilon > 0$. Then the gain in the quality of approximation is strict for $p \in (1, \infty)$. Namely, for any $p \in (1, \infty)$,

$$\left[\int_{\mathcal{X}} \left|\hat{f}^*(x) - f_0(x)\right|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}} \left|\hat{f}(x) - f_0(x)\right|^p dx - \delta\eta_p\right]^{1/p}, \tag{2.2}$$

where $\eta_p = \inf\{|v - t'|^p + |v' - t|^p - |v - t|^p - |v' - t'|^p\} > 0$, with the infimum taken over all v, v', t, t' in the set K such that $v' \geq v + \epsilon$ and $t' \geq t + \epsilon$.

This proposition establishes that the rearranged estimate $\hat{f}^*$ has a smaller (and often strictly smaller) estimation error in the L_p norm than the original estimate whenever the latter is not monotone. This is a very useful and generally applicable property that is independent of the sample size and of the way the original estimate $\hat{f}$ is obtained. The first part of the proposition states the weak inequality (2.1), and the second part states the strict inequality (2.2). For example, the inequality is strict for $p \in (1, \infty)$ if the original estimate $\hat{f}(x)$ is decreasing on a subset of $\mathcal{X}$ having positive measure, while the target function $f_0(x)$ is increasing on $\mathcal{X}$ (by increasing, we mean strictly increasing throughout). Of course, if $f_0(x)$ is constant, then the inequality (2.1) becomes an equality, as the distribution of the rearranged function $\hat{f}^*$ is the same as the distribution of the original function $\hat{f}$, that is $F_{\hat{f}^*} = F_{\hat{f}}$.

The weak inequality (2.1) is a direct (yet important) consequence of the classical rearrangement inequality due to Lorentz (1953): Let q and g be two functions mapping $\mathcal{X}$ to K. Let q^* and g^* denote their corresponding increasing rearrangements. Then,

$$\int_{\mathcal{X}} L(q^*(x), g^*(x))dx \leq \int_{\mathcal{X}} L(q(x), g(x))dx,$$

for any submodular discrepancy function $L : \mathbb{R}^2 \mapsto \mathbb{R}_+$. Set $q(x) = \hat{f}(x)$, $q^*(x) = \hat{f}^*(x)$, $g(x) = f_0(x)$, and $g^*(x) = f_0^*(x)$. Now, note that in our case $f_0^*(x) = f_0(x)$ almost everywhere, that is, the target function is its own rearrangement. Let us recall that L is submodular if for each pair of vectors (v, v') and (t, t') in $\mathbb{R}^2$, we have that

$$L(v \wedge v', t \wedge t') + L(v \vee v', t \vee t') \leq L(v, t) + L(v', t'), \tag{2.3}$$

In other words, a function L measuring the discrepancy between vectors is submodular if co-monotonization of vectors reduces the discrepancy. When a function L is smooth, submodularity is equivalent to the condition $\partial_v \partial_t L(v,t) \leq 0$ holding for each (v,t) in $\mathbb{R}^2$. Thus, for example, power functions $L(v,t) = |v-t|^p$ for $p \in [1,\infty)$ and many other loss functions are submodular.

In the Appendix, we provide a proof of the strong inequality (2.2) as well as the direct proof of the weak inequality (2.1). The direct proof illustrates how reductions of the estimation error arise from even a partial sorting of the values of the estimate $\hat{f}$. Moreover, the direct proof characterizes the conditions for the strict reduction of the estimation error.

It is also worth emphasizing the following immediate asymptotic implication of the above finite-sample result: The rearranged estimate $\hat{f}^*$ inherits the L_p rates of convergence from the original estimates $\hat{f}$. For $p \in [1,\infty]$, if $\lambda_n = [\int_{\mathcal{X}} |\hat{f}(x) - f_0(x)|^p du]^{1/p} = O_P(a_n)$ for some sequence of constants a_n, then $[\int_{\mathcal{X}} |\hat{f}^*(x) - f_0(x)|^p du]^{1/p} \leq \lambda_n = O_P(a_n)$.

2.3. **Computation of the Rearranged Estimate.** One of the following methods can be used for computing the rearrangement. Let $\{X_j, j = 1, ..., B\}$ be either (1) a set of equidistant points in $[0,1]$ or (2) a sample of i.i.d. draws from the uniform distribution on $[0,1]$. Then the rearranged estimate $\hat{f}^*(u)$ at point $u \in \mathcal{X}$ can be approximately computed as the u-quantile of the sample $\{f(X_j), j = 1, ..., B\}$. The first method is deterministic, and the second is stochastic. Thus, for a given number of draws B, the complexity of computing the rearranged estimate $f^*(u)$ in this way is equivalent to the complexity of computing the sample u-quantile in a sample of size B. The number of evaluations B can depend on the problem. Suppose that the density function of the random variable $f(X)$, when $X \sim U(0,1)$, is bounded away from zero over a neighborhood of $f^*(x)$. Then $f^*(x)$ can be computed with the accuracy of $O_P(1/\sqrt{B})$, as $B \to \infty$, where the rate follows from the results of Knight (2002).

2.4. **The Rearrangement and Its Approximation Property: The Multivariate Case.** In this section we consider multivariate functions $f : \mathcal{X}^d \to K$, where $\mathcal{X}^d = [0,1]^d$ and K is a bounded subset of $\mathbb{R}$. The notion of monotonicity we seek to impose on f is the following: We say that the function f is weakly increasing in x if $f(x') \geq f(x)$ whenever $x' \geq x$. The notation $x' = (x'_1, ..., x'_d) \geq x = (x_1, ..., x_d)$ means that one vector is weakly larger than the other in each of the components, that is, $x'_j \geq x_j$ for each $j = 1, ..., d$. In what follows, we use the notation $f(x_j, x_{-j})$ to denote the dependence of f on its j-th argument, x_j, and all other arguments, x_{-j}, that exclude x_j. The notion of monotonicity above is equivalent to the requirement that for each j in $1, ..., d$ the mapping $x_j \mapsto f(x_j, x_{-j})$ is weakly increasing in x_j, for each x_{-j} in $\mathcal{X}^{d-1}$.

Define the rearranged operator R_j and the rearranged function $f_j^*(x)$ with respect to the j-th argument as follows:

$$f_j^*(x) := R_j \circ f(x) := \inf\left\{y : \left[\int_{\mathcal{X}} 1\{f(x_j', x_{-j}) \leq y\} dx_j'\right] \geq x_j\right\}.$$

This is the one-dimensional increasing rearrangement applied to the one-dimensional function $x_j \mapsto f(x_j, x_{-j})$, holding the other arguments x_{-j} fixed. The rearrangement is applied for every value of the other arguments x_{-j}.

Let $\pi = (\pi_1, ..., \pi_d)$ be an ordering, i.e., a permutation, of the integers $1, ..., d$. Let us define the π-rearrangement operator R_π and the π-rearranged function $f_\pi^*(x)$ as follows:

$$f_\pi^*(x) := R_\pi \circ f(x) := R_{\pi_1} \circ ... \circ R_{\pi_d} \circ f(x).$$

For any ordering π, the π-rearrangement operator rearranges the function with respect to all of its arguments. As shown below, the resulting function $f_\pi(x)$ is weakly increasing in x.

In general, two different orderings π and π' of $1, ..., d$ can yield different rearranged functions $f_\pi^*(x)$ and $f_{\pi'}^*(x)$. Therefore, to resolve the conflict among rearrangements done with different orderings, we may consider averaging among them: letting Π be any collection of distinct orderings π, we can define the average rearrangement as

$$f^*(x) := \frac{1}{|\Pi|} \sum_{\pi \in \Pi} f_\pi^*(x), \tag{2.4}$$

where $|\Pi|$ denotes the number of elements in the set of orderings Π. Dette and Scheder (2006) also proposed averaging all the possible orderings of the smoothed rearrangement in the context of monotone conditional mean estimation. As shown below, the approximation error of the average rearrangement is weakly smaller than the average of approximation errors of individual π-rearrangements.

The following proposition describes the properties of multivariate π-rearrangements:

Proposition 2. *Let the target function $f_0 : \mathcal{X}^d \to K$ be weakly increasing and measurable in x. Let $\hat{f} : \mathcal{X}^d \to K$ be a measurable function that is an initial estimate of the target function f_0. Let $\bar{f} : \mathcal{X}^d \to K$ be another estimate of f_0, which is measurable in x, including, for example, a rearranged $\hat{f}$ with respect to some of the arguments. Then,*

1. For each ordering π of $1, ..., d$, the π-rearranged estimate $\hat{f}_\pi^*(x)$ is weakly increasing in x. Moreover, $\hat{f}^*(x)$, an average of π-rearranged estimates, is weakly increasing in x.

2. (a) For any j in $1, ..., d$ and any p in $[1, \infty]$, the rearrangement of $\bar{f}$ with respect to the j-th argument produces a weak reduction in the approximation error:

$$\left[\int_{\mathcal{X}^d} |\bar{f}_j^*(x) - f_0(x)|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}^d} |\bar{f}(x) - f_0(x)|^p dx\right]^{1/p}. \tag{2.5}$$

(b) Consequently, a π-rearranged estimate $\hat{f}^*_\pi(x)$ of $\hat{f}(x)$ weakly reduces the approximation error of the original estimate:

$$\left[\int_{\mathcal{X}^d} |\hat{f}^*_\pi(x) - f_0(x)|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}^d} |\hat{f}(x) - f_0(x)|^p dx\right]^{1/p}. \tag{2.6}$$

3. Suppose that $\bar{f}(x)$ and $f_0(x)$ have the following properties: there exist subsets $\mathcal{X}_j \subset \mathcal{X}$ and $\mathcal{X}'_j \subset \mathcal{X}$, each of measure $\delta > 0$, and a subset $\mathcal{X}_{-j} \subseteq \mathcal{X}^{d-1}$, of measure $\nu > 0$, such that for all $x = (x_j, x_{-j})$ and $x' = (x'_j, x_{-j})$, with $x'_j \in \mathcal{X}'_j$, $x_j \in \mathcal{X}_j$, $x_{-j} \in \mathcal{X}_{-j}$, we have that (i) $x'_j > x_j$, (ii) $\bar{f}(x) > \bar{f}(x') + \epsilon$, and (iii) $f_0(x') > f_0(x) + \epsilon$, for some $\epsilon > 0$.

(a) Then, for any $p \in (1, \infty)$,

$$\left[\int_{\mathcal{X}^d} |\bar{f}^*_j(x) - f_0(x)|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}^d} |\bar{f}(x) - f_0(x)|^p dx - \eta_p \delta \nu\right]^{1/p}, \tag{2.7}$$

where $\eta_p = \inf\{|v - t'|^p + |v' - t|^p - |v - t|^p - |v' - t'|^p\} > 0$, with the infimum taken over all v, v', t, t' in the set K such that $v' \geq v + \epsilon$ and $t' \geq t + \epsilon$.

(b) Further, for an ordering $\pi = (\pi_1, ..., \pi_k, ..., \pi_d)$ with $\pi_k = j$, let $\bar{f}$ be a partially rearranged function, $\bar{f}(x) = R_{\pi_{k+1}} \circ ... \circ R_{\pi_d} \circ \hat{f}(x)$ (for $k = d$ we set $\bar{f}(x) = \hat{f}(x)$). If the function $\bar{f}(x)$ and the target function $f_0(x)$ satisfy the condition stated above, then, for any $p \in (1, \infty)$,

$$\left[\int_{\mathcal{X}^d} |\hat{f}^*_\pi(x) - f_0(x)|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}^d} |\hat{f}(x) - f_0(x)|^p dx - \eta_p \delta \nu\right]^{1/p}. \tag{2.8}$$

4. The approximation error of an average rearrangement is weakly smaller than the average approximation error of the individual π- rearrangements: For any $p \in [1, \infty]$,

$$\left[\int_{\mathcal{X}^d} |\hat{f}^*(x) - f_0(x)|^p dx\right]^{1/p} \leq \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \left[\int_{\mathcal{X}^d} |\hat{f}^*_\pi(x) - f_0(x)|^p dx\right]^{1/p}. \tag{2.9}$$

This proposition generalizes the results of Proposition 1 to the multivariate case, also demonstrating several features unique of the multivariate case. We see that the π-rearranged functions are monotonic in all of the arguments. Dette and Scheder (2006), using a different argument, showed that their smoothed rearrangement for conditional mean functions is monotonic in both arguments for the bivariate case in large samples. The rearrangement along any argument improves the approximation properties of the estimate. Moreover, the improvement is strict when the rearrangement with respect to a j-th argument is performed on an estimate that is decreasing in the j-th argument, while the target function is increasing in the same j-th argument, in the sense precisely defined in the proposition. Moreover, averaging different π-rearrangements is better (on average) than using a single π-rearrangement chosen at random. All other basic implications of the proposition are similar to those discussed for the univariate case.

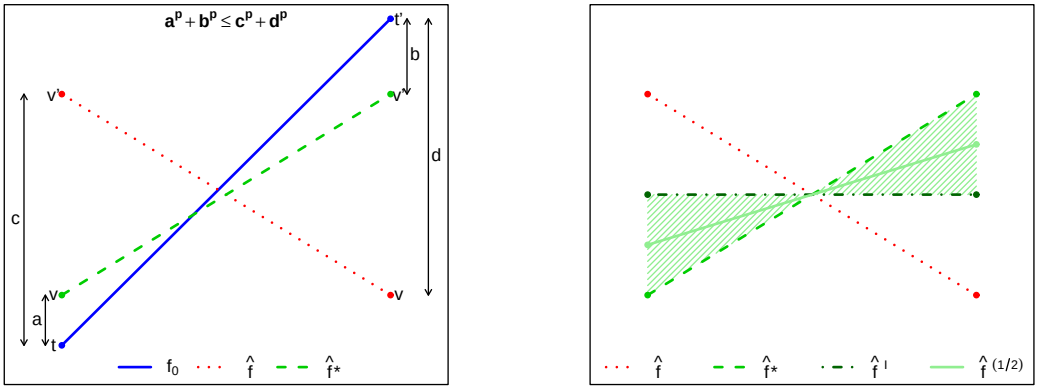

FIGURE 1. Graphical illustration for the proof of Proposition 1 (left panel) and comparison to isotonic regression (right panel). In the figure, f_0 represents the target function, $\hat{f}$ the original estimate, $\hat{f}^*$ the rearranged estimate, $\hat{f}^I$ the isotonized estimate, and $\hat{f}^{(1/2)}$ the average of the rearranged and isotonized estimates. In the left panel $L(v,t) = a^p$, $L(v',t) = c^p$, $L(v',t') = b^p$, and $L(v,t') = d^p$.

2.5. **Discussion and Comparisons.** In what follows we informally explain why rearrangement provides the improvement property and compare rearrangement to isotonization.

Let us begin by noting that the proof of the improvement property can be first reduced to the case of simple functions or, equivalently, functions with a finite domain, and then to the case of "very simple" functions with a two-point domain. The improvement property for these very simple functions then follows from the submodularity property (2.3). In the left panel of Figure 1 we illustrate this property geometrically by plotting the original estimate $\hat{f}$, the rearranged estimate $\hat{f}^*$, and the true function f_0. In this example, the original estimate is decreasing and hence violates the monotonicity requirement. We see that the two-point rearrangement co-monotonizes $\hat{f}^*$ with f_0 and thus brings $\hat{f}^*$ closer to f_0. Also, we can view the rearrangement as a projection on the set of weakly increasing functions that have the same distribution as the original estimate $\hat{f}$.

Next in the right panel of Figure 1 we plot both the rearranged and isotonized estimates. The isotonized estimate $\hat{f}^I$ is a projection of the original estimate $\hat{f}$ on the set of weakly increasing functions (that only preserves the mean of the original estimate). We can compute the two values of the isotonized estimate $\hat{f}^I$ by assigning both of them the average of the two

values of the original estimate $\hat{f}$, whenever the latter violate the monotonicity requirement, and leaving the original values unchanged, otherwise. We see from Figure 1 that in our example this produces a flat function $\hat{f}^I$. This computational procedure, known as "pool adjacent violators," naturally extends to domains with more than two points by simply applying the procedure iteratively to any pair of points at which the monotonicity requirement remains violated (Ayer, Brunk, Ewing, Reid, and Silverman 1955).

Using the computational definition of isotonization, one can show that, like rearrangement, isotonization also improves upon the original estimate, for any $p \in [1, \infty]$:

$$\left[\int_{\mathcal{X}} |\hat{f}^I(x) - f_0(x)|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}} |\hat{f}(x) - f_0(x)|^p dx\right]^{1/p}, \tag{2.10}$$

see, e.g., Barlow, Bartholomew, Bremner, and Brunk (1972). Therefore, it follows that any function $\hat{f}^\lambda$ in the convex hull of the rearranged and the isotonized estimate both monotonizes and improves upon the original estimate. The first property is obvious and the second follows from homogeneity and subadditivity of norms, that is for any $p \in [1, \infty]$:

$$\begin{aligned}\left[\int_{\mathcal{X}} |\hat{f}^\lambda(x) - f_0(x)|^p dx\right]^{1/p} &\leq \lambda \left[\int_{\mathcal{X}} |\hat{f}^I(x) - f_0(x)|^p dx\right]^{1/p} + (1-\lambda) \left[\int_{\mathcal{X}} |\hat{f}^*(x) - f_0(x)|^p dx\right]^{1/p} \\ &\leq \left[\int_{\mathcal{X}} |\hat{f}(x) - f_0(x)|^p dx\right]^{1/p}, \end{aligned} \tag{2.11}$$

where $\hat{f}^\lambda(x) = \lambda \hat{f}^*(x) + (1-\lambda)\hat{f}^I(x)$ for any $\lambda \in [0, 1]$. Before proceeding further, let us also note that, by an induction argument similar to that presented in the previous section, the improvement property listed above extends to the sequential multivariate isotonization and to its convex hull with the sequential multivariate rearrangement.

Thus, we see that a rather rich class of procedures (or operators) both monotonizes the original estimate and reduces the distance to the true target function. It is also important to note that there is no single best distance-reducing monotonizing procedure. Indeed, whether the rearranged estimate $\hat{f}^*$ approximates the target function better than the isotonized estimate $\hat{f}^I$ depends on how steep or flat the target function is. We illustrate this point via a simple example plotted in the right panel of Figure 1: Consider any increasing target function taking values in the shaded area between $\hat{f}^*$ and $\hat{f}^I$, and also the function $\hat{f}^{1/2}$, the average of the isotonized and the rearranged estimate, that passes through the middle of the shaded area. Suppose first that the target function is steeper than $\hat{f}^{1/2}$, then $\hat{f}^*$ has a smaller approximation error than $\hat{f}^I$. Now suppose instead that the target function is flatter than $\hat{f}^{1/2}$, then $\hat{f}^I$ has a smaller approximation error than $\hat{f}^*$. It is also clear that, if the target function is neither very steep nor very flat, $\hat{f}^{1/2}$ can outperform either $\hat{f}^*$ or $\hat{f}^I$. Thus, in practice we can choose rearrangement, isotonization, or, some combination of the two, depending on our beliefs about how steep or flat the target function is in a particular application.

3. Improving Interval Estimates of Monotone Functions by Rearrangement

In this section we propose to directly apply the rearrangement, univariate and multivariate, to simultaneous confidence intervals for functions. We show that our proposal will necessarily improve the original intervals by decreasing their length and increasing their coverage.

Suppose that we are given an initial simultaneous confidence interval

$$[\ell, u] = ([\ell(x), u(x)], x \in \mathcal{X}^d), \tag{3.1}$$

where $\ell(x)$ and $u(x)$ are the lower and upper end-point functions such that $\ell(x) \leq u(x)$ for all $x \in \mathcal{X}^d$.

We further suppose that the confidence interval $[\ell, u]$ has either the exact or the asymptotic *confidence property* for the estimand function f, namely, for a given $\alpha \in (0, 1)$,

$$\text{Prob}_P\{f \in [\ell, u]\} = \text{Prob}_P\{\ell(x) \leq f(x) \leq u(x), \text{ for all } x \in \mathcal{X}^d\} \geq 1 - \alpha, \tag{3.2}$$

for all probability measures P in some set $\mathcal{P}_n$ containing the true probability measure P_0. We assume that property (3.2) holds either in the finite sample sense, that is, for the given sample size n, or in the asymptotic sense, that is, for all but finitely many sample sizes n (Lehmann and Romano 2005).

A common type of a confidence interval for functions is one where

$$\ell(x) = \hat{f}(x) - s(x)\hat{c} \text{ and } u(x) = \hat{f}(x) + s(x)\hat{c}, \tag{3.3}$$

where $\hat{f}(x)$ is a point estimate, $s(x)$ is the standard error of the point estimate, and $\hat{c}$ is the critical value chosen so that the confidence interval $[\ell, u]$ in (3.1) covers the function f with the specified probability, as stated in (3.2). There are many well-established methods for the construction of the critical value, ranging from analytical tube methods to the bootstrap, both for parametric and non-parametric estimators (see, e.g., Johansen and Johnstone (1990), and Hall (1993)). The Wasserman (2006) book provides an excellent overview of the existing methods for inference on functions. The problem with such confidence intervals, similar to the point estimates themselves, is that these intervals need not be monotonic. Indeed, typical inferential procedures do not guarantee that the end-point functions $\hat{f}(x) \pm s(x)\hat{c}$ of the confidence interval are monotonic. This means that such a confidence interval contains non-monotone functions that can be excluded from it.

In some cases the confidence intervals mentioned above may not contain any monotone functions at all, for example, due to a small sample size or misspecification. We define the case of misspecification or incorrect centering of the confidence interval $[\ell, u]$ as any case where the estimand f being covered by $[\ell, u]$ is not equal to the weakly increasing target function f_0, so that f may not be monotone. Misspecification is a rather common occurrence both

in parametric and non-parametric estimation. Indeed, correct centering of confidence intervals in parametric estimation requires perfect specification of functional forms and is generally hard to achieve. On the other hand, correct centering of confidence intervals in nonparametric estimation requires the so-called undersmoothing, a delicate requirement, which amounts to using a relatively large number of terms in series estimation and a relatively small bandwidth in kernel-based estimation. In real applications with many regressors, researchers tend to use oversmoothing rather than undersmoothing. In a recent development, Genovese and Wasserman (2008) provide, in our interpretation, some formal justification for oversmoothing: targeting inference on functions f, that represent various smoothed versions of f_0 and thus summarize features of f_0, may be desirable to make inference more robust, or, equivalently, to enlarge the class of data-generating processes $\mathcal{P}_n$ for which the confidence interval property (3.2) holds. In any case, regardless of the reasons for why confidence intervals may target f instead of f_0, our procedures will work for inference on the monotonized version f^* of f.

Our proposal for improved interval estimates is to rearrange the entire simultaneous confidence intervals into a monotonic interval given by

$$[\ell^*, u^*] = ([\ell^*(x), u^*(x)], x \in \mathcal{X}^d), \tag{3.4}$$

where the lower and upper end-point functions ℓ^* and u^* are the increasing rearrangements of the original end-point functions ℓ and u. In the multivariate case, we use the symbols ℓ^* and u^* to denote either π-multivariate rearrangements ℓ^*_π and u^*_π or average multivariate rearrangements ℓ^* and u^*, whenever we do not need to specifically emphasize the dependence on π.

The following proposition describes the formal property of the rearranged confidence intervals.

Proposition 3. *Let $[\ell, u]$ in (3.1) be the original confidence interval that has the confidence interval property (3.2) for the estimand function $f : \mathcal{X}^d \mapsto K$ and let the rearranged confidence interval $[\ell^*, u^*]$ be defined as in (3.4).*

1. The rearranged confidence interval $[\ell^*, u^*]$ is weakly increasing and non-empty, in the sense that the end-point functions ℓ^* and u^* are weakly increasing on $\mathcal{X}^d$ and satisfy $\ell^* \leq u^*$ on $\mathcal{X}^d$. Moreover, the event that $[\ell, u]$ contains the estimand f implies the event that $[\ell^*, u^*]$ contains the rearranged, hence monotonized, version f^* of the estimand f:

$$f \in [\ell, u] \text{ implies } f^* \in [\ell^*, u^*]. \tag{3.5}$$

In particular, under the correct specification, when f equals a weakly increasing target function f_0, we have that $f = f^* = f_0$, so that

$$f_0 \in [\ell, u] \text{ implies } f_0 \in [\ell^*, u^*]. \tag{3.6}$$

Therefore, $[\ell^*, u^*]$ covers f^*, which is equal to f_0 under the correct specification, with a probability that is greater or equal to the probability that $[\ell, u]$ covers f.

2. The rearranged confidence interval $[\ell^*, u^*]$ is weakly shorter than the initial confidence interval $[\ell, u]$ in the average L^p length: for each $p \in [1, \infty]$:

$$\left[\int_{\mathcal{X}^d} \left|\ell^*(x) - u^*(x)\right|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}^d} \left|\ell(x) - u(x)\right|^p dx\right]^{1/p}. \tag{3.7}$$

3. In the univariate case, suppose that $\ell(x)$ and $u(x)$ have the following properties: there exist subsets $\mathcal{X}_0 \subset \mathcal{X}$ and $\mathcal{X}_0' \subset \mathcal{X}$, each of measure greater than $\delta > 0$ such that for all $x' \in \mathcal{X}_0'$ and $x \in \mathcal{X}_0$, we have that $x' > x$, and either (i) $\ell(x) > \ell(x') + \epsilon$, and $u(x') > u(x) + \epsilon$, for some $\epsilon > 0$ or (ii) $\ell(x') > \ell(x) + \epsilon$ and $u(x) > u(x') + \epsilon$, for some $\epsilon > 0$. Then, for any $p \in (1, \infty)$

$$\left[\int_{\mathcal{X}} \left|\ell^*(x) - u^*(x)\right|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}} \left|\ell(x) - u(x)\right|^p - \eta_p \delta\right]^{1/p}, \tag{3.8}$$

where $\eta_p = \inf\{|v - t'|^p + |v' - t|^p - |v - t|^p - |v' - t'|^p\} > 0$, where the infimum is taken over all v, v', t, t' in the set K such that $v' \geq v + \epsilon$ and $t' \geq t + \epsilon$ or such that $v \geq v' + \epsilon$ and $t \geq t' + \epsilon$.

In the multivariate case with $d \geq 2$, for an ordering $\pi = (\pi_1, ..., \pi_k, ..., \pi_d)$ of integers $\{1, ..., d\}$ with $\pi_k = j$, let $\bar{g}$ denote the partially rearranged function, $\bar{g}(x) = R_{\pi_{k+1}} \circ ... \circ R_{\pi_d} \circ \hat{g}(x)$, where for $k = d$ we set $\bar{g}(x) = g(x)$. Suppose that $\bar{\ell}(x)$ and $\bar{u}(x)$ have the following properties: there exist subsets $\mathcal{X}_j \subset \mathcal{X}$ and $\mathcal{X}_j' \subset \mathcal{X}$, each of measure greater than $\delta > 0$, and a subset $\mathcal{X}_{-j} \subseteq \mathcal{X}^{d-1}$, of measure $\nu > 0$, such that for all $x = (x_j, x_{-j})$ and $x' = (x_j', x_{-j})$, with $x_j' \in \mathcal{X}_j'$, $x_j \in \mathcal{X}_j$, $x_{-j} \in \mathcal{X}_{-j}$, we have that (i) $x_j' > x_j$, and either (ii) $\bar{\ell}(x) > \bar{\ell}(x') + \epsilon$, and $\bar{u}(x') > \bar{u}(x) + \epsilon$, for some $\epsilon > 0$ or (iii) $\bar{\ell}(x') > \bar{\ell}(x) + \epsilon$ and $\bar{u}(x) > \bar{u}(x') + \epsilon$, for some $\epsilon > 0$. Then, for any $p \in (1, \infty)$

$$\left[\int_{\mathcal{X}^d} \left|\ell_\pi^*(x) - u_\pi^*(x)\right|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}^d} \left|\ell(x) - u(x)\right|^p - \eta_p \delta \nu\right]^{1/p}, \tag{3.9}$$

where $\eta_p > 0$ is defined as above.

The proposition shows that the rearranged confidence intervals are weakly shorter than the original confidence intervals, and also qualifies when the rearranged confidence intervals are strictly shorter. In particular, the inequality (3.7) is necessarily strict for $p \in (1, \infty)$ in the univariate case, if there is a region of positive measure in $\mathcal{X}$ over which the end-point functions $x \mapsto \ell(x)$ and $x \mapsto u(x)$ are not comonotonic. This weak shortening result follows for univariate cases directly from the rearrangement inequality of Lorentz (1953), and the strong shortening follows from a simple strengthening of the Lorentz inequality, as argued in the proof. The shortening results for the multivariate case follow by an induction argument. Moreover, the order-preservation property of the univariate and multivariate rearrangements, demonstrated in the proof, implies that the rearranged confidence interval $[\ell^*, u^*]$ has a weakly higher coverage

than the original confidence interval $[\ell, u]$. We do not qualify strict improvements in coverage, but we demonstrate them through the examples in the next section.

Our idea of directly monotonizing the interval estimates also applies to other monotonization procedures. Indeed, the proof of Proposition 3 reveals that part 1 of Proposition 3 applies to any order-preserving monotonization operator T, such that

$$g(x) \leq m(x) \text{ for all } x \in \mathcal{X}^d \text{ implies } T \circ g(x) \leq T \circ m(x) \text{ for all } x \in \mathcal{X}^d. \tag{3.10}$$

Furthermore, part 2 of Proposition 3 on the weak shortening of the confidence intervals applies to any distance-reducing operator T such that

$$\left[\int_{\mathcal{X}^d} |T \circ \ell(x) - T \circ u(x)|^p dx\right]^{1/p} \leq \left[\int_{\mathcal{X}^d} |\ell(x) - u(x)|^p dx\right]^{1/p}. \tag{3.11}$$

Rearrangements, univariate and multivariate, are one instance of order-preserving and distance-reducing operators. Isotonization, univariate and multivariate, is another important instance (Robertson, Wright, and Dykstra 1988). Moreover, convex combinations of order-preserving and distance-reducing operators, such as the average of rearrangement and isotonization, are also order-preserving and distance-reducing. We demonstrate the inferential implications of these properties further in the computational experiments reported in Section 4.

4. Illustrations

In this section we provide an empirical application to biometric age-height charts. We show how the rearrangement monotonizes and improves various nonparametric point and interval estimates for functions, and then we quantify the improvement in a simulation example that mimics the empirical application. We carried out all the computations using the software `R` (R Development Core Team 2008), the quantile regression package `quantreg` (Koenker 2008), and the functional data analysis package `fda` (Ramsay, Wickham, and Graves 2007).

4.1. **An Empirical Illustration with Age-Height Reference Charts.**

Since their introduction by Quetelet in the 19th century, reference growth charts have become common tools to assess an individual's health status. These charts describe the evolution of individual anthropometric measures, such as height, weight, and body mass index, across different ages. See Cole (1988) for a classical work on the subject, and Wei, Pere, Koenker, and He (2006) for a recent analysis from a quantile regression perspective, and additional references.

To illustrate the properties of the rearrangement method we consider the estimation of growth charts for height. It is clear that height should naturally follow an increasing relationship with age. Our data consist of repeated cross sectional measurements of height and age from the 2003-2004 National Health and Nutrition Survey collected by the National Center for Health Statistics. Height is measured as standing height in centimeters, and age is recorded in

months and expressed in years. To avoid confounding factors that might affect the relationship between age and height, we restrict the sample to US-born white males of age two through twenty. Our final sample consists of 533 subjects almost evenly distributed across these ages.

Let Y and X denote height and age, respectively. Let $E[Y|X = x]$ denote the conditional expectation of Y given $X = x$, and $Q_Y(u|X = x)$ denote the u-th quantile of Y given $X = x$, where u is the quantile index. The population functions of interests are (1) the conditional expectation function (CEF), (2) the conditional quantile functions (CQF) for several quantile indices (5%, 50%, and 95%), and (3) the entire conditional quantile process (CQP) for height given age. In the first case, the target function $x \mapsto f_0(x)$ is $x \mapsto E[Y|X = x]$; in the second case, the target function $x \mapsto f_0(x)$ is $x \mapsto Q_Y[u|X = x]$, for $u = 5\%$, 50%, and 95%; and, in the third case, the target function $(u, x) \mapsto f_0(u, x)$ is $(u, x) \mapsto Q_Y[u|X = x]$. The natural monotonicity requirements for the target functions are the following: The CEF $x \mapsto E[Y|X = x]$ and the CQF $x \mapsto Q_Y(u|X = x)$ should be increasing in age x, and the CQP $(u, x) \mapsto Q_Y[u|X = x]$ should be increasing in both age x and the quantile index u.

We estimate the target functions using non-parametric ordinary least squares or quantile regression techniques and then rearrange the estimates to satisfy the monotonicity requirements. We consider (a) kernel, (b) locally linear, (c) regression splines, and (d) Fourier series methods. For the kernel and locally linear methods, we choose a bandwidth of one year and a box kernel. For the regression splines method, we use cubic B-splines with a knot sequence $\{3, 5, 8, 10, 11.5, 13, 14.5, 16, 18\}$, following Wei, Pere, Koenker, and He (2006). For the Fourier method, we employ eight trigonometric terms, with four sines and four cosines. Finally, for the estimation of the conditional quantile process, we use a net of two hundred quantile indices $\{0.005, 0.010, ..., 0.995\}$. In the choice of the parameters for the different methods, we select values that either have been used in the previous empirical work or give rise to specifications with similar complexities for the different methods.

The panels A-D of Figure 2 show the original and rearranged estimates of the conditional expectation function for the different methods. All the estimated curves have trouble capturing the slowdown in the growth of height after age fifteen and yield non-monotonic curves for the highest values of age. The Fourier series performs particularly poorly in approximating the aperiodic age-height relationship and has many non-monotonicities. The rearranged estimates correct the non-monotonicity of the original estimates, providing weakly increasing curves that coincide with the original estimates in the parts where the latter are monotonic. Figure 3 displays similar but more pronounced non-monotonicity patterns for the estimates of the conditional quantile functions. In all cases, the rearrangement again performs well in delivering curves that improve upon the original estimates and that satisfy the natural monotonicity requirement. We quantify this improvement in the next subsection.

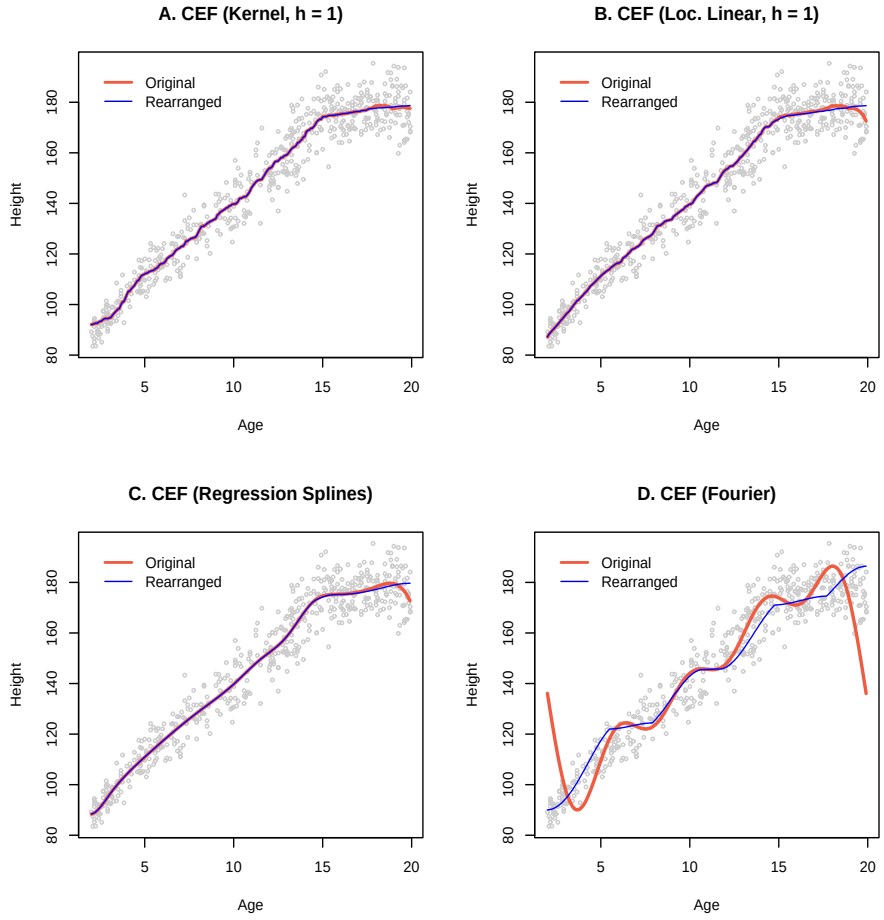

FIGURE 2. Nonparametric estimates of the Conditional Expectation Function (CEF) of height given age and their increasing rearrangements. Nonparametric estimates are obtained using kernel regression (A), locally linear regression (B), cubic regression B-splines series (C), and Fourier series (D).

Figure 4 illustrates the multivariate rearrangement of the conditional quantile process (CQP) along both the age and the quantile index arguments. We plot, in three dimensions, the original estimate, its age rearrangement, its quantile rearrangement, and its average multivariate rearrangement (the average of the age-quantile and quantile-age rearrangements). We also plot the corresponding contour surfaces. Here, for brevity, we focus on the Fourier series estimates, which have the most severe non-monotonicity problems. (Analogous figures for the other estimation methods considered can be found in the working paper version Chernozhukov, Fernandez-Val, and Galichon (2006a)). Moreover, we do not show the multivariate age-quantile and quantile-age rearrangements separately, because they are very similar to the

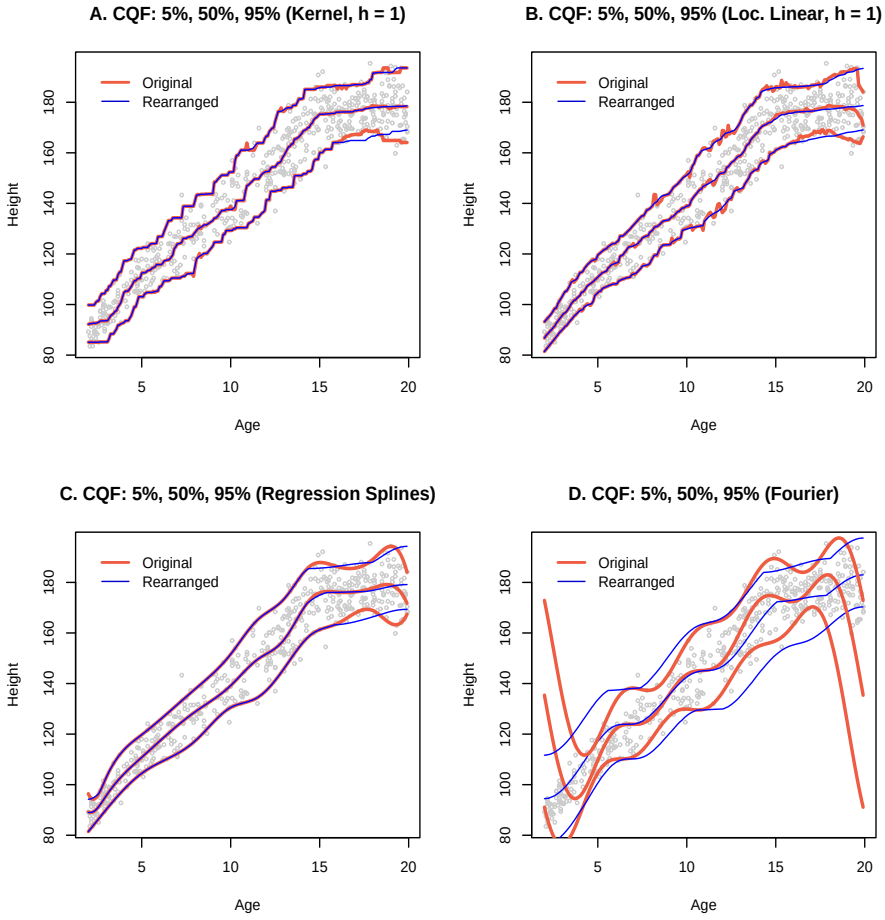

FIGURE 3. Nonparametric estimates of the 5%, 50%, and 95% Conditional Quantile Functions (CQF) of height given age and their increasing rearrangements. Nonparametric estimates are obtained using kernel regression (A), locally linear regression (B), cubic regression B-splines series (C), and Fourier series (D).

average multivariate rearrangement. We see from the contour plots that the estimated CQP is non-monotone in age and non-monotone in the quantile index at extremal values of this index. The average multivariate rearrangement fixes the non-monotonicity problem delivering an estimate of the CQP that is monotone in both the age and the quantile index arguments. Furthermore, by the theoretical results of the paper, the multivariate rearranged estimates necessarily improve upon the original estimates.

In Figures 5 and 6, we illustrate the inference properties of the rearranged confidence intervals. Figure 5 shows 90% uniform confidence intervals for the conditional expectation function and three conditional quantile functions for the 5%, 50%, and 95% quantiles based on Fourier

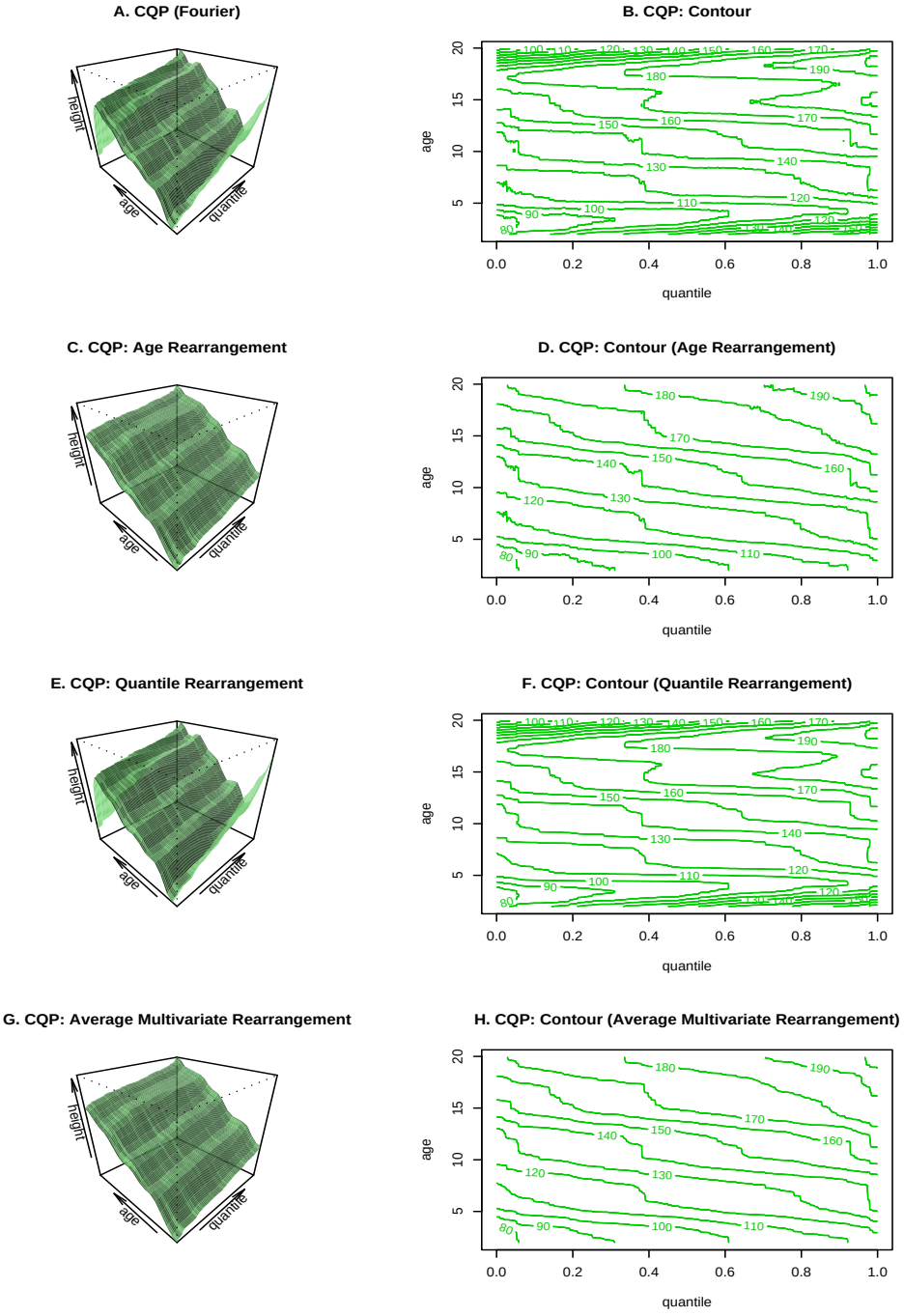

FIGURE 4. Fourier series estimates of the Conditional Quantile Process (CQP) of height given age and their increasing rearrangements. Panels C and E plot the one dimensional increasing rearrangement along the age and quantile dimension, respectively; panel G shows the average multivariate rearrangement.

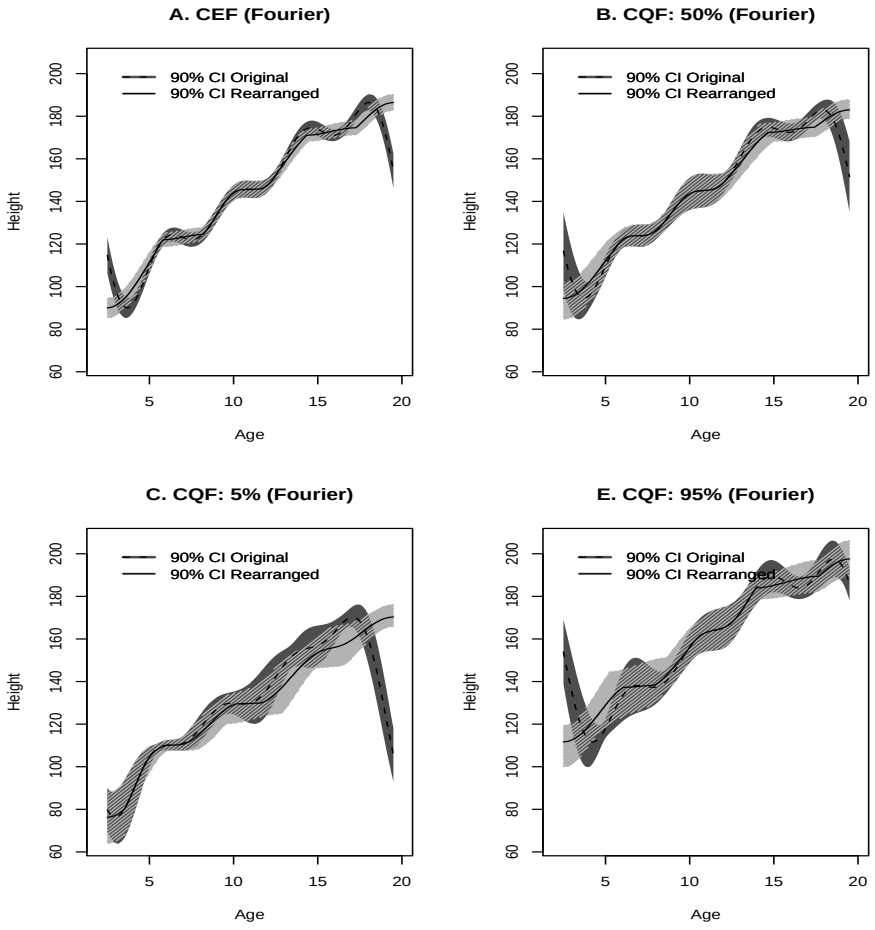

FIGURE 5. 90% confidence intervals for the Conditional Expectation Function (CEF), and 5%, 50% and 95% Conditional Quantile Functions (CQF) of height given age and their increasing rearrangements. Nonparametric estimates are based on Fourier series and confidence bands are obtained by bootstrap with 200 repetitions.

series estimates. We obtain the initial confidence intervals of the form (3.3) using the bootstrap with 200 repetitions to estimate the critical values (Hall 1993). We then obtain the rearranged confidence intervals by rearranging the lower and upper end-point functions of the initial confidence intervals, following the procedure defined in Section 3. In Figure 6, we illustrate the construction of the confidence intervals in the multidimensional case by plotting the initial and rearranged 90% uniform confidence bands for the entire conditional quantile process based on the Fourier series estimates. We see from the figures that the rearranged confidence intervals correct the non-monotonicity of the original confidence intervals and reduce their average length, as we shall verify numerically in the next section.

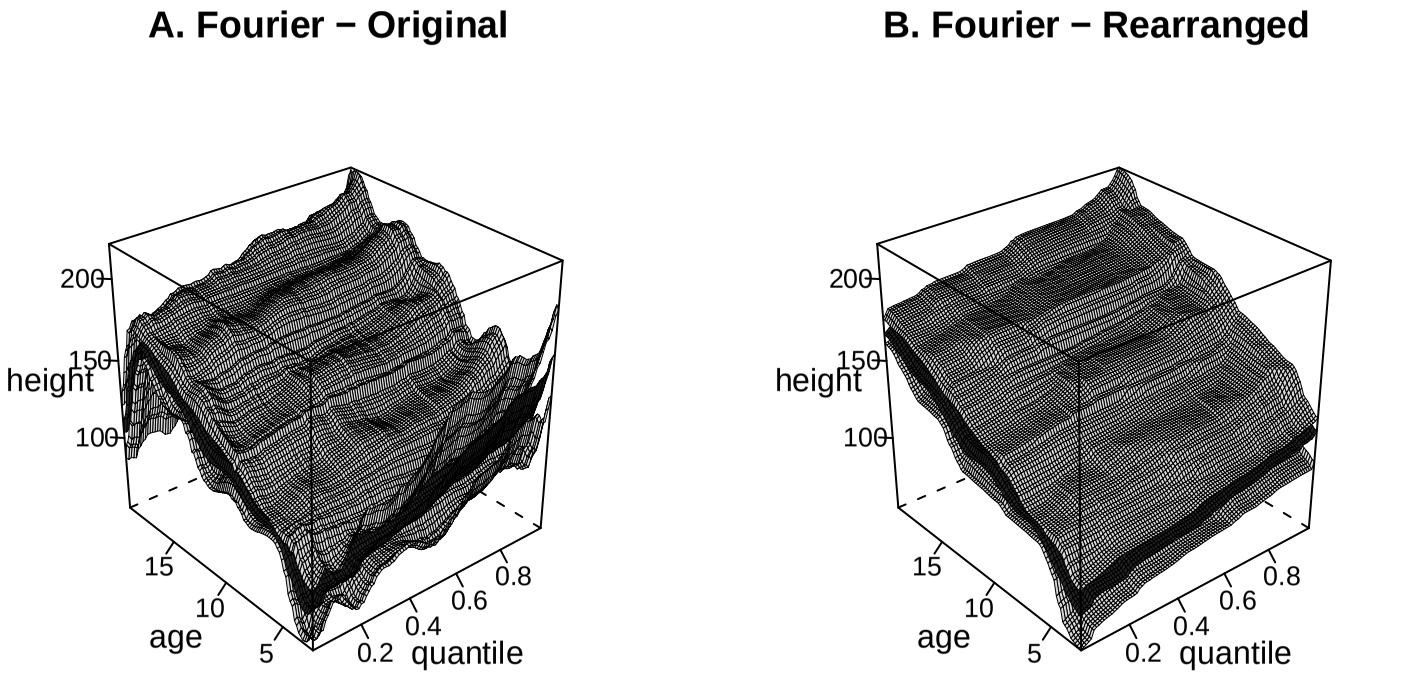

FIGURE 6. Fourier series 90% confidence interval for the Conditional Quantile Process (CQP) of height given age and average multivariate rearrangement. Confidence bands are obtained by bootstrap with 200 repetitions.

4.2. **Monte-Carlo Illustration.** The following Monte Carlo experiment quantifies the improvement in the point and interval estimation that rearrangement provides relative to the original estimates. We also compare rearrangement to isotonization and to convex combinations of rearrangement and isotonization.

Our experiment closely matches the empirical application presented above. Specifically, we consider a design where the outcome variable Y equals a location function plus a disturbance ϵ, $Y = Z(X)'\beta + \epsilon$, and the disturbance is independent of the regressor X. The vector $Z(X)$ includes a constant and a piecewise linear transformation of the regressor X with three changes of slope, namely $Z(X) = (1, X, 1\{X > 5\}\cdot(X-5), 1\{X > 10\}\cdot(X-10), 1\{X > 15\}\cdot(X-15))$. This design implies the conditional expectation function

$$E[Y|X] = Z(X)'\beta, \tag{4.1}$$

and the conditional quantile function

$$Q_Y(u|X) = Z(X)'\beta + Q_\epsilon(u). \tag{4.2}$$

We select the parameters of the design to match the empirical example of growth charts in the previous subsection. Thus, we set the parameter β equal to the ordinary least squares estimate obtained in the growth chart data, namely (71.25, 8.13, −2.72, 1.78, −6.43). This parameter value and the location specification (4.2) imply a model for the CEF and CQP that is monotone in age over the range of ages 2-20. To generate the values of the dependent variable, we draw disturbances from a normal distribution with the mean and variance equal to the mean and variance of the estimated residuals, $\epsilon = Y - Z(X)'\beta$, in the growth chart data. We fix the regressor X in all of the replications to be the observed values of age in the data set. In each replication, we estimate the CEF and CQP using the nonparametric methods described in the previous section, along with a global polynomial and a flexible Fourier methods. For the global polynomial method, we fit a quartic polynomial. For the flexible Fourier method, we use a quadratic polynomial and four trigonometric terms, with two sines and two cosines.

In Table 1 we report the average L^p errors (for $p = 1, 2$, and ∞) for the original estimates of the CEF. We also report the relative efficiency of the rearranged estimates, measured as the ratio of the average error of the rearranged estimate to the average error of the original estimate; together with relative efficiencies for an alternative approach based on isotonization of the original estimates, and an approach consisting of averaging the rearranged and isotonized estimates. The two-step approach based on isotonization corresponds to the SI estimator in Mammen (1991), where the isotonization step is carried out using the pool-adjacent violator algorithm (PAVA). For regression splines, we also consider the one-step monotone regression splines of Ramsay (1998).

TABLE 1. L^p Estimation Errors of Original, Rearranged, Isotonized, Average Rearranged-Isotonized, and Monotone Estimates of the Conditional Expectation Function, for $p = 1, 2$, and ∞. Univariate Case.

p	L^p_O	L^p_R/L^p_O	L^p_I/L^p_O	$L^p_{(R+I)/2}/L^p_O$	L^p_M/L^p_O	L^p_O	L^p_R/L^p_O	L^p_I/L^p_O	$L^p_{(R+I)/2}/L^p_O$
	A. Kernel					B. Locally Linear			
1	1.00	0.97	0.98	0.98		0.79	0.96	0.97	0.96
2	1.30	0.98	0.99	0.98		0.99	0.96	0.97	0.97
∞	4.54	0.99	1.00	1.00		2.93	0.95	0.95	0.95
	C. Regression Splines					D. Quartic			
1	0.87	0.93	0.95	0.94	0.99	1.33	0.89	0.87	0.87
2	1.09	0.93	0.95	0.94	0.99	1.64	0.89	0.88	0.87
∞	3.68	0.85	0.88	0.86	0.84	4.38	0.86	0.86	0.86
	E. Fourier					F. Flexible Fourier			
1	6.57	0.49	0.59	0.40		0.73	0.97	0.99	0.98
2	10.8	0.35	0.45	0.30		0.91	0.98	0.99	0.98
∞	48.9	0.16	0.34	0.20		2.40	0.98	0.98	0.98

Notes: The table is based on 1,000 replications. The algorithm for the monotone regression splines stopped with an error message in 6 cases; these cases were discarded for all the estimators. L^p_O is the L^p error of the error of the original estimate; L^p_R is the L^p error of the rearranged estimate; L^p_I is the L^p error of the isotonized estimate; $L^p_{(R+I)/2}$ is the L^p error of the average of the rearranged and isotonized estimates; L^p_M is the L^p error of the monotone regression splines estimates.

We calculate the average L^p error as the Monte Carlo average of

$$L^p := \left[\int_{\mathcal{X}} |\bar{f}(x) - f_0(x)|^p dx \right]^{1/p},$$

where the target function $f_0(x)$ is the CEF $E[Y|X = x]$, and the estimate $\bar{f}(x)$ denotes either the original nonparametric estimate of the CEF or its increasing transformation. For all of the methods considered, we find that the rearranged curves estimate the true CEF more accurately than the original curves, providing a 1% to 84% reduction in the average error, depending on the method and the norm (i.e., values of p). In this example, there is no uniform winner between rearrangement and isotonic regression. The rearrangement works better than isotonization for Kernel, Local Polynomials, Splines, Fourier, and Flexible Fourier estimates, but it works worse than isotonization for global Quartic polynomials for some norms. Averaging the two procedures seems to be a good compromise for all the estimation methods considered. For regression splines, the performance of the rearrangement is comparable to the computationally more intensive one-step monotone splines procedure.

In Table 2 we report the average L^p errors for the original estimates of the conditional quantile process. We also report the ratio of the average error of the multivariate rearranged

estimate, with respect to the age and quantile index arguments, to the average error of the original estimate; together with the same ratios for isotonized estimates and average rearranged-isotonized estimated. The isotonized estimates are obtained by sequentially applying the PAVA to the two arguments, and then averaging for the two possible orderings age-quantile and quantile-age.

TABLE 2. L^p Estimation Errors of Original, Rearranged, Isotonized, and Average Rearranged-Isotonized Estimates of the Conditional Quantile Process, for $p = 1, 2$, and ∞. Multivariate Case.

p	L^p_O	L^p_R/L^p_O	L^p_I/L^p_O	$L^p_{(R+I)/2}/L^p_O$	L^p_O	L^p_R/L^p_O	L^p_I/L^p_O	$L^p_{(R+I)/2}/L^p_O$
		A. Kernel				B. Locally Linear		
1	1.49	0.95	0.97	0.96	1.21	0.91	0.93	0.92
2	1.99	0.96	0.98	0.97	1.61	0.91	0.93	0.92
∞	13.7	0.92	0.97	0.94	12.3	0.84	0.87	0.85
		C. Splines				D. Quartic		
1	1.33	0.90	0.93	0.91	1.49	0.90	0.89	0.89
2	1.78	0.90	0.92	0.90	1.87	0.90	0.89	0.89
∞	16.9	0.72	0.76	0.73	12.6	0.68	0.69	0.68
		E. Fourier				F. Flexible Fourier		
1	6.72	0.62	0.77	0.64	1.05	0.96	0.97	0.96
2	13.7	0.39	0.58	0.44	1.38	0.95	0.97	0.96
∞	84.9	0.26	0.47	0.36	10.9	0.84	0.86	0.85

Notes: The table is based on 1,000 replications. L^p_O is the L^p error of the original estimate; L^p_R is the L^p error of the average multivariate rearranged estimate; L^p_I is the L^p error of the average multivariate isotonized estimate; $L^p_{(R+I)/2}$ is the L^p error of the mean of the average multivariate rearranged and isotonized estimates.

The average L_p error is the Monte Carlo average of

$$L^p := \left[\int_{\mathcal{U}} \int_{\mathcal{X}} |\bar{f}(u,x) - f_0(u,x)|^p dx du \right]^{1/p},$$

where the target function $f_0(u,x)$ is the conditional quantile process $Q_Y(u|X=x)$, and the estimate $\bar{f}(u,x)$ denotes either the original nonparametric estimate of the conditional quantile process or its monotone transformation. We present the results for the average multivariate rearrangement only. The age-quantile and quantile-age multivariate rearrangements give errors that are very similar to their average multivariate rearrangement, and we therefore do not report them separately. For all the methods considered, we find that the multivariate rearranged curves estimate the true CQP more accurately than the original curves, providing a 4% to 74% reduction in the approximation error, depending on the method and the norm. As in

the univariate case, there is no uniform winner between rearrangement and isotonic regression and their average estimate gives a good balance.

Table 3 reports Monte Carlo coverage frequencies and integrated lengths for the original and monotonized 90% confidence bands for the CEF. For a measure of length, we used the integrated L^p length, as defined in Proposition 3, with $p = 1, 2$, and ∞. We constructed the original confidence intervals of the form specified in equations (3.3) by obtaining the pointwise standard errors of the original estimates using the bootstrap with 200 repetitions, and we calibrated the critical value so that the original confidence bands cover the entire true function with the exact frequency of 90%. We constructed monotonized confidence intervals by applying rearrangement, isotonization, and a rearrangement-isotonization average to the end-point functions of the original confidence intervals, as suggested in Section 3. Here, we find that in all cases the rearrangement and other monotonization methods increase the coverage of the confidence intervals while reducing their length. In particular, we see that monotonization increases coverage especially for the local estimation methods, whereas it reduces length most noticeably for the global estimation methods. For the most problematic Fourier estimates, there are both important increases in coverage and reductions in length.

Appendix A. Proofs of Propositions

A.1. **Proof of Proposition 1.** The first part establishes the weak inequality, following in part the strategy in Lorentz's (1953) proof. The proof focuses directly on obtaining the result stated in the proposition. The second part establishes the strong inequality.

Proof of Part 1. We assume at first that the functions $\hat{f}(\cdot)$ and $f_0(\cdot)$ are simple functions, constant on intervals $((s-1)/r, s/r]$, $s = 1, ..., r$. For any simple $f(\cdot)$ with r steps, let f denote the r-vector with the s-th element, denoted f_s, equal to the value of $f(\cdot)$ on the s-th interval. Let us define the sorting operator $S(f)$ as follows: Let k be an integer in $1, ..., r$ such that $f_k > f_m$ for some $m > k$. If k does not exist, set $S(f) = f$. If k exists, set $S(f)$ to be a r-vector with the k-th element equal to f_m, the m-th element equal to f_k, and all other elements equal to the corresponding elements of f. For any submodular function $L : \mathbb{R}^2 \to \mathbb{R}_+$, by $f_k \geq f_m$, $f_{0m} \geq f_{0k}$ and the definition of the submodularity, $L(f_m, f_{0k}) + L(f_k, f_{0m}) \leq L(f_k, f_{0k}) + L(f_m, f_{0m})$. Therefore, we conclude that $\int_{\mathcal{X}} L(S(\hat{f})(x), f_0(x))dx \leq \int_{\mathcal{X}} L(\hat{f}(x), f_0(x))dx$, using that we integrate simple functions.

Applying the sorting operation a sufficient finite number of times to $\hat{f}$, we obtain a completely sorted, that is, rearranged, vector $\hat{f}^*$. Thus, we can express $\hat{f}^*$ as a finite composition $\hat{f}^* = S \circ ... \circ S(\hat{f})$. By repeating the argument above, each composition weakly reduces the

TABLE 3. Coverage Probabilities and Integrated Lengths of Original, Rearranged, Isotonized, and Average Rearranged-Isotonized 90% Confidence Intervals for the Conditional Expectation Function.

Interval	Cover	Length				Cover	Length			
		L^1	L^1/L^1_O	L^2/L^2_O	L^∞/L^∞_O		L^1	L^1/L^1_O	L^2/L^2_O	L^∞/L^∞_O
		A. Kernel					B. Locally Linear			
O	.90	8.80				.90	8.63			
R	.96	8.79	1	1	.99	.96	8.63	1	1	.97
I	.94	8.80	1	1	.99	.94	8.63	1	1	.98
$(R+I)/2$	.95	8.80	1	1	.99	.95	8.63	1	1	.97
		C. Splines					D. Quartic			
O	.90	6.32				.90	10.43			
R	.91	6.32	1	1	1	.90	10.41	1	.99	.93
I	.91	6.32	1	1	1	.90	10.43	1	1	.93
$(R+I)/2$	.91	6.32	1	1	1	.90	10.42	1	1	.93
		E. Fourier					F. Flexible Fourier			
O	.90	24.91				.90	6.45			
R	1	24.52	.98	.94	0.63	.90	6.45	1	1	.97
I	1	24.91	1	.97	0.69	.90	6.45	1	1	.97
$(R+I)/2$	1	24.71	.99	.95	0.65	.90	6.45	1	1	.97

Notes: The table is based on 1,000 replications. O, R, I, and $(R+I)/2$ refer to original, rearranged, isotonized, and average rearranged-isotonized confidence intervals. Coverage probabilities (Cover) are for the entire function. Original confidence intervals calibrated to have 90% coverage probabilities.

approximation error. Therefore,

$$\int_{\mathcal{X}} L(\hat{f}^*(x), f_0(x))dx \leq \int_{\mathcal{X}} L(\underbrace{S \circ ... \circ S}_{\text{finite times}}(\hat{f}), f_0(x))dx \leq \int_{\mathcal{X}} L(\hat{f}(x), f_0(x))dx. \tag{A.1}$$

Furthermore, this inequality is extended to general measurable functions $\hat{f}(\cdot)$ and $f_0(\cdot)$ mapping $\mathcal{X}$ to K by taking a sequence of bounded simple functions $\hat{f}^{(r)}(\cdot)$ and $f_0^{(r)}(\cdot)$ converging to $\hat{f}(\cdot)$ and $f_0(\cdot)$ almost everywhere as $r \to \infty$. The almost everywhere convergence of $\hat{f}^{(r)}(\cdot)$ to $\hat{f}(\cdot)$ implies the almost everywhere convergence of its quantile function $\hat{f}^{*(r)}(\cdot)$ to the quantile function of the limit, $\hat{f}^*(\cdot)$. Since inequality (A.1) holds along the sequence, the dominated convergence theorem implies that (A.1) also holds for the general case. □

Proof of Part 2. Let us first consider the case of simple functions, as defined in the proof of Part 1. We take the functions to satisfy the following hypotheses: there exist regions $\mathcal{X}_0$ and $\mathcal{X}_0'$, each of measure greater than $\delta > 0$, such that for all $x \in \mathcal{X}_0$ and $x' \in \mathcal{X}_0'$, we have that (i) $x' > x$, (ii) $\hat{f}(x) > \hat{f}(x') + \epsilon$, and (iii) $f_0(x') > f_0(x) + \epsilon$, for $\epsilon > 0$ specified in the proposition. For any strictly submodular function $L : \mathbb{R}^2 \to \mathbb{R}_+$ we have that $\eta =$

$\inf\{L(v',t)+L(v,t')-L(v,t)-L(v',t')\}>0$, where the infimum is taken over all v,v',t,t' in the set K such that $v'\geq v+\epsilon$ and $t'\geq t+\epsilon$. A simple graphical illustration for this property is given in Figure 1.

We can begin sorting by exchanging an element $\hat{f}(x)$, $x\in\mathcal{X}_0$, of r-vector $\hat{f}$ with an element $\hat{f}(x')$, $x'\in\mathcal{X}_0'$, of r-vector $\hat{f}$. This induces a sorting gain of at least η times $1/r$. The total mass of points that can be sorted in this way is at least δ. We then proceed to sort all of these points in this way, and then continue with the sorting of other points. After the sorting is completed, the total gain from sorting is at least $\delta\eta$. That is, $\int_{\mathcal{X}} L(\hat{f}^*(x),f_0(x))dx\leq\int_{\mathcal{X}} L(\hat{f}(x),f_0(x))dx-\delta\eta$.

We then extend this inequality to the general measurable functions exactly as in the proof of part one. □

A.2. **Proof of Proposition 2.** The proof consists of the following four parts.

Proof of Part 1. We prove the claim by induction. The claim is true for $d=1$ by $\hat{f}^*(x)$ being a quantile function. We then consider any $d\geq 2$. Suppose the claim is true in $d-1$ dimensions. If so, then the estimate $\bar{f}(x_j,x_{-j})$, obtained from the original estimate $\hat{f}(x)$ after applying the rearrangement to all arguments x_{-j} of x, except for the argument x_j, must be weakly increasing in x_{-j} for each x_j. Thus, for any $x'_{-j}\geq x_{-j}$, we have that

$$\bar{f}(X_j,x'_{-j})\geq\bar{f}(X_j,x_{-j}) \text{ for } X_j\sim U[0,1]. \tag{A.2}$$

Therefore, the random variable on the left of (A.2) dominates the random variable on the right of (A.2) in the stochastic sense. Therefore, the quantile function of the random variable on the left dominates the quantile function of the random variable on the right, namely

$$\bar{f}_j^*(x_j,x'_{-j})\geq\bar{f}_j^*(x_j,x_{-j}) \text{ for each } x_j\in\mathcal{X}=[0,1]. \tag{A.3}$$

Moreover, for each x_{-j}, the function $x_j\mapsto\bar{f}_j^*(x_j,x_{-j})$ is weakly increasing by virtue of being a quantile function. We conclude therefore that $x\mapsto\bar{f}_j^*(x)$ is weakly increasing in all of its arguments at all points $x\in\mathcal{X}^d$. The claim of Part 1 of the Proposition now follows by induction. □

Proof of Part 2 (a). By Proposition 1, we have that for each x_{-j},

$$\int_{\mathcal{X}}\left|\bar{f}_j^*(x_j,x_{-j})-f_0(x_j,x_{-j})\right|^p dx_j\leq\int_{\mathcal{X}}\left|\bar{f}(x_j,x_{-j})-f_0(x_j,x_{-j})\right|^p dx_j. \tag{A.4}$$

Now, the claim follows by integrating with respect to x_{-j} and taking the p-th root of both sides. For $p=\infty$, the claim follows by taking the limit as $p\to\infty$. □

Proof of Part 2 (b). We first apply the inequality of Part 2(a) to $\bar{f}(x)=\hat{f}(x)$, then to $\bar{f}(x)=R_{\pi_d}\circ\hat{f}(x)$, then to $\bar{f}(x)=R_{\pi_{d-1}}\circ R_{\pi_d}\circ\hat{f}(x)$, and so on. In doing so, we recursively generate a sequence of weak inequalities that imply the inequality (2.6) stated in the Proposition. □

Proof of Part 3 (a). For each $x_{-j} \in \mathcal{X}^{d-1} \setminus \mathcal{X}_{-j}$, by Part 2(a), we have the weak inequality (A.4), and for each $x_{-j} \in \mathcal{X}_{-j}$, by the inequality for the univariate case stated in Proposition 1 Part 2, we have the strong inequality

$$\int_{\mathcal{X}} \left|\bar{f}_j^*(x_j, x_{-j}) - f_0(x_j, x_{-j})\right|^p dx_j \leq \int_{\mathcal{X}} \left|\bar{f}(x_j, x_{-j}) - f_0(x_j, x_{-j})\right|^p dx_j - \eta_p \delta, \tag{A.5}$$

where η_p is defined in the same way as in Proposition 1. Integrating the weak inequality (A.4) over $x_{-j} \in \mathcal{X}^{d-1} \setminus \mathcal{X}_{-j}$, of measure $1-\nu$, and the strong inequality (A.5) over $\mathcal{X}_{-j}$, of measure ν, we obtain

$$\int_{\mathcal{X}^d} \left|\bar{f}_j^*(x) - f_0(x)\right|^p dx \leq \int_{\mathcal{X}^d} \left|\bar{f}(x) - f_0(x)\right|^p dx - \eta_p \delta \nu. \tag{A.6}$$

The claim now follows. □

Proof of Part 3 (b). As in Part 2(a), we can recursively obtain a sequence of weak inequalities describing the improvements in approximation error from rearranging sequentially with respect to the individual arguments. Moreover, at least one of the inequalities can be strengthened to be of the form stated in (A.6), from the assumption of the claim. The resulting system of inequalities yields the inequality (2.8), stated in the proposition. □

Proof of Part 4. We can write

$$\begin{aligned} \left[\int_{\mathcal{X}^d} \left|\hat{f}^*(x) - f_0(x)\right|^p dx\right]^{1/p} &= \left[\int_{\mathcal{X}^d} \left|\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\left(\hat{f}^*(x) - f_0(x)\right)\right|^p dx\right]^{1/p} \\ &\leq \frac{1}{|\Pi|}\sum_{\pi\in\Pi}\left[\int_{\mathcal{X}^d} \left|\hat{f}_\pi^*(x) - f_0(x)\right|^p dx\right]^{1/p}, \end{aligned} \tag{A.7}$$

where the last inequality follows by pulling out $1/|\Pi|$ and then applying the triangle inequality for the L_p norm. □

A.3. **Proof of Proposition 3.** Proof of Part 1. The monotonicity follows from Proposition 2. The rest of the proof relies on establishing the order-preserving property of the π-rearrangement operator: for any measurable functions $g, m : \mathcal{X}^d \to \mathbb{R}$, we have that

$$g(x) \leq m(x) \text{ for all } x \in \mathcal{X}^d \text{ implies } g^*(x) \leq m^*(x) \text{ for all } x \in \mathcal{X}^d. \tag{A.8}$$

Given the property we have that

$$\ell(x) \leq f(x) \leq u(x) \text{ for all } x \in \mathcal{X}^d \text{ implies } \ell^*(x) \leq f^*(x) \leq u^*(x) \text{ for all } x \in \mathcal{X}^d,$$

which verifies the claim of the first part. The claim also extends to the average multivariate rearrangement, since averaging preserves the order-preserving property.

It remains to establish the order-preserving property for $\pi-$rearrangement, which we do by induction. We first note that in the univariate case, when $d = 1$, order preservation is obvious from the rearrangement being a quantile function: the random variable $m(X)$, where

$X \sim$ Uniform$[\mathcal{X}]$, dominates the random variable $g(X)$ in the stochastic sense, hence the quantile function $m^*(x)$ of $m(X)$ must be greater than the quantile function $g^*(x)$ of $g(X)$ for each $x \in \mathcal{X}$. We then extend this to the multivariate case by induction: Suppose the order-preserving property is true for any $d-1$ with $d \geq 2$. If so, then for each $x_j \in \mathcal{X}$ and $x_{-j} \in \mathcal{X}^{d-1}$

$$g(x_j, x_{-j}) \leq m(x_j, x_{-j}) \text{ implies } \bar{g}(x_j, x_{-j}) \leq \bar{m}(x_j, x_{-j}),$$

where $\bar{g}$ and $\bar{m}$ are multivariate rearrangements of $x_{-j} \mapsto g(x_j, x_{-j})$ and $x_{-j} \mapsto m(x_j, x_{-j})$ with respect to x_{-j}, holding x_j fixed. Now apply the order-preserving property of the univariate rearrangement to the univariate functions $x_j \mapsto \bar{g}(x_j, x_{-j})$ and $x_j \mapsto \bar{m}(x_j, x_{-j})$, holding x_{-j} fixed, for each x_{-j}, to conclude that (A.8) holds. □

Proof of Part 2. As stated in the text, the weak inequality is due to Lorentz (1953). For completeness we only briefly note that the proof follows similarly to the proof of Proposition 1. Indeed, we can start with simple functions $\ell(\cdot)$ and $u(\cdot)$ and work with their equivalent vector representations ℓ and u. Then we apply the sorting operator S to the pair of r-vectors (ℓ, u) defined as $S(\ell, u) = (S(\ell), S'(u))$, where $S(\ell)$ is the sorting operator on the vector ℓ defined in the proof of Proposition 1, and $S'(u)$ is a subordinated sorting operator on the vector u defined by two conditions: (1) if $S(\ell)$ exchanges the k-th and m-th elements of ℓ, where $m > k$, then $S'(u)$ exchanges the k-th and m-th elements of u if these elements violate the monotonicity requirement, i.e., if $u_k > u_m$, and otherwise S' leaves all elements of u unchanged, i.e. $S'(u) = u$; and (2) if $S(\ell)$ exchanges no elements of ℓ, i.e. $S(\ell) = \ell$, then $S'(u)$ is simply the unrestricted $S(u)$ operator as defined in the proof of Proposition 1, i.e., $S'(u) = S(u)$. By the definition of submodularity (2.3), each application of S weakly reduces submodular discrepancies between vectors, so that pairs of vectors in the sequence $\{(\ell, u), S(\ell, u), ..., S \circ ... \circ S(\ell, u), (\ell^*, u^*)\}$ become progressively weakly closer to each other, and the sequence can be taken to be finite, where the last pair is the rearrangement (ℓ^*, u^*) of vectors (ℓ, u). The inequality extends to general bounded measurable functions by passing to the limit and using the dominated convergence theorem. To extend the proof to the multivariate case, we apply exactly the same induction strategy as in the proof of Proposition 2. □

Proof of Part 3. Finally, the proof of strict inequality in the univariate case is similar to the proof of Proposition 2, using the fact that for strictly submodular functions $L : \mathbb{R}^2 \mapsto \mathbb{R}_+$ we have that $\eta = \inf\{L(v', t) + L(v, t') - L(v, t) - L(v', t')\} > 0$, where the infimum is taken over all v, v', t, t' in the set K such that $v' \geq v + \epsilon$ and $t' \geq t + \epsilon$ or such that $v \geq v' + \epsilon$ and $t \geq t' + \epsilon$. The extension of the strict inequality to the multivariate case follows exactly as in the proof of Proposition 2. □

References

Andrews, D. W. K. (1991): "Asymptotic normality of series estimators for nonparametric and semiparametric regression models," *Econometrica*, 59(2), 307–345.

Ayer, M., H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman (1955): "An empirical distribution function for sampling with incomplete information," *Ann. Math. Statist.*, 26, 641–647.

Barlow, R. E., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk (1972): *Statistical inference under order restrictions. The theory and application of isotonic regression.* John Wiley & Sons, London-New York-Sydney, Wiley Series in Probability and Mathematical Statistics.

Chaudhuri, P. (1991): "Nonparametric estimates of regression quantiles and their local Bahadur representation," *Ann. Statist.*, 19, 760–777.

Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2006a): "Improving Approximations of Monotone Functions by Rearrangement," MIT and UCL CEMMAP Working Paper, available at www.arxiv.org, cemmap.ifs.org.uk, www.ssrn.com.

——— (2006b): "Quantile and Probability Curves without Crossing," MIT and UCL CEMMAP Working Paper, available at www.arxiv.org, cemmap.ifs.org.uk, www.ssrn.com.

——— (2006c): "Rearranging Edgeworth-Cornish-Fisher Expansions," MIT and UCL CEMMAP Working Paper, available at www.arxiv.org, cemmap.ifs.org.uk, www.ssrn.com.

Cole, T. J. (1988): "Fitting smoothed centile curves to reference data," *J Royal Stat Soc*, 151, 385–418.

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

Dette, H., N. Neumeyer, and K. F. Pilz (2006): "A simple nonparametric estimator of a strictly monotone regression function," *Bernoulli*, 12(3), 469–490.

Dette, H., and K. F. Pilz (2006): "A comparative study of monotone nonparametric kernel estimates," *J. Stat. Comput. Simul.*, 76(1), 41–56.

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.

Fan, J., and I. Gijbels (1996): *Local polynomial modelling and its applications*, vol. 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Fougeres, A.-L. (1997): "Estimation de densites unimodales," *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 25(3), 375–387.

Gallant, A. R. (1981): "On the bias in flexible functional forms and an essentially unbiased form: the Fourier flexible form," *J. Econometrics*, 15(2), 211–245.

Genovese, C., and L. Wasserman (2008): "Adaptive confidence bands," *Ann. Statist.*, 36(2), 875–905.

Hall, P. (1993): "On Edgeworth Expansion and Bootstrap Confidence Bands in Nonparametric Curve Estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(1), 291–304.

Hardy, G. H., J. E. Littlewood, and G. Pólya (1952): *Inequalities*. Cambridge University Press, 2d ed.

He, X., and Q.-M. Shao (2000): "On parameters of increasing dimensions," *J. Multivariate Anal.*, 73(1), 120–135.

Johansen, S., and I. M. Johnstone (1990): "Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis," *Ann. Statist.*, 18(2), 652–684.

Knight, K. (2002): "What are the limiting distributions of quantile estimators?," in *Statistical data analysis based on the L_1-norm and related methods (Neuchâtel, 2002)*, Stat. Ind. Technol., pp. 47–65. Birkhäuser, Basel.

Koenker, R. (2008): *quantreg: Quantile Regression*R package version 4.17.

Koenker, R., and G. S. Bassett (1978): "Regression quantiles," *Econometrica*, 46, 33–50.

Koenker, R., and P. Ng (2005): "Inequality constrained quantile regression," *Sankhyā*, 67(2), 418–440.

Lehmann, E. L., and J. P. Romano (2005): *Testing statistical hypotheses*, Springer Texts in Statistics. Springer, New York, third edn.

Lorentz, G. G. (1953): "An inequality for rearrangements," *Amer. Math. Monthly*, 60, 176–179.

Mammen, E. (1991): "Nonparametric Regression Under Qualitative Smoothness Assumptions," *The Annals of Statistics*, 19(2), 741–759.

Mammen, E., J. S. Marron, B. A. Turlach, and M. P. Wand (2001): "A general projection framework for constrained smoothing," *Statist. Sci.*, 16(3), 232–248.

Matzkin, R. L. (1994): "Restrictions of economic theory in nonparametric methods," in *Handbook of econometrics, Vol. IV*, vol. 2 of *Handbooks in Econom.*, pp. 2523–2558. North-Holland, Amsterdam.

Newey, W. K. (1997): "Convergence rates and asymptotic normality for series estimators," *J. Econometrics*, 79(1), 147–168.

Portnoy, S. (1997): "Local asymptotics for quantile smoothing splines," *Ann. Statist.*, 25(1), 414–434.

R Development Core Team (2008): *R: A Language and Environment for Statistical Computing*R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Ramsay, J. O. (1988): "Monotone Regression Splines in Action," *Statistical Science*, 3(4), 425–441.

Ramsay, J. O., and B. W. Silverman (2005): *Functional data analysis*. Springer, New York, second edn.

Ramsay, J. O., H. Wickham, and S. Graves (2007): *fda: Functional Data Analysis*R package version 1.2.3.

Robertson, T., F. T. Wright, and R. L. Dykstra (1988): *Order restricted statistical inference*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester.

Silvapulle, M. J., and P. K. Sen (2005): *Constrained statistical inference*, Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.

Stone, C. J. (1994): "The use of polynomial splines and their tensor products in multivariate function estimation," *Ann. Statist.*, 22(1), 118–184, With discussion by Andreas Buja and Trevor Hastie and a rejoinder by the author.

Villani, C. (2003): *Topics in optimal transportation*, vol. 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.

Wand, M. P., and M. C. Jones (1995): *Kernel smoothing*, vol. 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.

Wasserman, L. (2006): *All of nonparametric statistics*, Springer Texts in Statistics. Springer, New York.

Wei, Y., A. Pere, R. Koenker, and X. He (2006): "Quantile regression methods for reference growth charts," *Stat. Med.*, 25(8), 1369–1382.