IMPROVING 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 the 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 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 $\hat{f}$. We illustrate the results with a computational example and an empirical example dealing with age-height growth charts.

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

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

1. Introduction

A common problem in statistics is to approximate an unknown monotonic function on the basis of available samples. For example, biometric age-height charts should be monotonic in age; econometric demand functions should be monotonic in price; and quantile functions should be monotonic in the probability index. Suppose an original, possibly 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 conditional mean functions (Dette, Neumeyer, and Pilz 2006, Dette and Pilz 2006) and various conditional quantile and probability functions (Chernozhukov, Fernandez-Val, and Galichon (2006a, 2006b)). In this paper, it is shown that the rearrangement of the original estimate is useful not only for producing monotonicity, but also has the following important property: The rearrangement always improves over 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 paper is organized as follows. In Section 2.1, we motivate the monotonicity issue in regression problems, and discuss common estimates/approximations of regression functions that are not naturally monotonic. In Section 2.2, we analyze the improvements in estimation/approximation properties that the rearranged estimates deliver. In Section 2.3, we discuss the computation of the rearrangement, using sorting and simulation. In Section 2.4, we extend the analysis of Section 2.2 to multivariate functions. In Section 3, we provide proofs of the main results. In Section 4, we present an empirical application to biometric age-height charts. We show how the rearrangement monotonizes and improves the original estimates of the conditional mean function in this example, and quantify the improvement in a simulation example resembling the empirical one. In the same section, we also analyze estimation of conditional quantile processes for height given age that need to be monotonic in both age and the quantile index. We apply a multivariate rearrangement to doubly monotonize the estimates both in age and the quantile index. We show that the rearrangement monotonizes and improves the original estimates, and

quantify the improvement in a simulation example mimicking the empirical example. In Section 5 we offer a summary and a conclusion.

2. Improving Approximations of Monotonic Functions

2.1. **Common Estimates of Monotonic Functions.** A basic problem in many areas of analysis is to approximate an unknown function $f_0 : \mathbb{R}^d \to \mathbb{R}$ on the basis of some available information. In statistics, the common problem is to approximate an unknown regression function, such as the conditional mean or a conditional quantile, using an available sample. In numerical analysis, the common problem is to approximate an intractable target function by a more tractable function on the basis of the target function's values at a collection of points.

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). Likewise, in numerical analysis "data" often consist of values Y_i of a target function evaluated at a collection of mesh points $\{X_i, i = 1,, n\}$ and the mesh points themselves. The series estimates $x \mapsto \hat{f}(x) = P_{k_n}(x)'\hat{b}$ are widely used in data analysis due to their good approximation properties and computational tractability. However, these estimates need not be naturally monotone, unless explicit constraints are added into the optimization program (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 $x \mapsto \hat{f}(x)$ needs 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 re-arrangement 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 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, where K is a bounded subset of $\mathbb{R}$. 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* $p \in [1, \infty]$,

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

where $\eta_p = \inf\{|v - t'|^p + |v' - t|^p - |v - t|^p - |v' - t'|^p\}$ *and* $\eta_p > 0$ *for* $p \in (1, \infty)$, *with the infimum taken over all* v, v', t, t' *in the set* K *such that* $v' \geq v + \epsilon$ *and* $t' \geq t + \epsilon$.

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 $\hat{f}_0(x)$ 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 $F_{\hat{f}^*} = F_{\hat{f}}$.

This proposition establishes that the rearranged estimate $\hat{f}^*$ has a 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.

An indirect proof of the weak inequality (2.1) is a simple but important consequence of the following classical inequality due to Lorentz (1953): Let q and g be two functions mapping $\mathcal{X}$ to K, a bounded subset of $\mathbb{R}$. Let q^* and g^* denote their corresponding increasing rearrangements. Then,

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

for any submodular discrepancy function $L : \mathbb{R}^3 \mapsto \mathbb{R}$. 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. Moreover, $L(v, w, x) = |w - v|^p$ is submodular for $p \in [1, \infty)$. This proves the first part of the proposition above. For $p = \infty$, the first part follows by taking the limit as $p \to \infty$.

In Section 3 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.

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

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 net 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 the 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). As shown in Chernozhukov, Fernandez-Val, and Galichon (2006a), the density of $f(X)$, denoted $F'_f(t)$, exists if $f(x)$ is continuously differentiable and the number of elements in $\{x \in \mathcal{X} : f'(x) = 0\}$ is bounded; in particular,

$$F'_f(t) = \sum_{x \in \{r \in \mathcal{X} : f(r) = t\}} \frac{1}{|f'(x)|}. \tag{2.3}$$

Thus, the density $F'_f(t)$ is bounded away from zero if $f'(x)$ is bounded away from infinity. Interestingly, the density has infinite poles at $\{t \in \mathcal{X} : \text{there is an } x \text{ such that } f'(x) = 0 \text{ and } f(x) = t\}$.

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 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. 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:

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

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

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

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 $p \in [1, \infty]$,

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

where $\eta_p = \inf\{|v - t'|^p + |v' - t|^p - |v - t|^p - |v' - t'|^p\}$, and $\eta_p > 0$ for $p \in (1, \infty)$, 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 $p \in [1, \infty]$,

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

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

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

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. 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.

3. Proofs of Propositions

3.1. **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 ℓ be an integer in $1, ..., r$ such that $f_\ell > f_m$ for some $m > l$. If ℓ does not exist, set $S(f) = f$. If ℓ exists, set $S(f)$ to be a r-vector with the ℓ-th element equal to f_m, the m-th element equal to f_ℓ, and all other elements equal to the corresponding elements of f. For any submodular function $L : \mathbb{R}^2 \to \mathbb{R}_+$, by $f_\ell \geq f_m$, $f_{0m} \geq f_{0\ell}$ and the definition of the submodularity,

$$L(f_m, f_{0\ell}) + L(f_\ell, f_{0m}) \leq L(f_\ell, f_{0\ell}) + L(f_m, f_{0m}).$$

Therefore, we conclude that

$$\int_{\mathcal{X}} L(S(\hat{f})(x), f_0(x))dx \leq \int_{\mathcal{X}} L(\hat{f}(x), f_0(x))dx, \tag{3.1}$$

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 approximation error. Therefore,

$$\int_{\mathcal{X}} L(\hat{f}^*(x), f_0(x))dx \leq \int_{\mathcal{X}} L(\underbrace{S \circ ... \circ S}_{\text{finite times}}(\hat{f}), f_0(x))dx \leq \int_{\mathcal{X}} L(\hat{f}(x), f_0(x))dx. \tag{3.2}$$

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 (3.2) holds along the sequence, the dominated convergence theorem implies that (3.2) also holds for the general case. $\square$

Proof of Part 2. Let us first consider the case of simple functions, as defined in 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 $t' \geq t + \epsilon$.

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. $\quad\square$

3.2. **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

$$\bar{f}(X_j, x'_{-j}) \geq \bar{f}(X_j, x_{-j}) \text{ for } X_j \sim U(0, 1). \tag{3.3}$$

Therefore, the random variable on the left of (3.3) dominates the random variable on the right of (3.3) 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

$$\bar{f}_j^*(x_j, x'_{-j}) \geq \bar{f}_j^*(x_j, x_{-j}) \text{ for each } x_j \in \mathcal{X} = (0, 1). \tag{3.4}$$

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. $\quad\square$

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

$$\int_{\mathcal{X}} \left| \bar{f}_j^*(x_j, x_{-j}) - f_0(x_j, x_{-j}) \right|^p dx_j \leq \int_{\mathcal{X}} \left| \bar{f}(x_j, x_{-j}) - f_0(x_j, x_{-j}) \right|^p dx_j. \tag{3.5}$$

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 $p \to \infty$. $\quad\square$

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. $\quad\square$

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 (3.5), 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

$$\int_{\mathcal{X}} \left| \bar{f}_j^*(x_j, x_{-j}) - f_0(x_j, x_{-j}) \right|^p dx_j \leq \int_{\mathcal{X}} \left| \bar{f}(x_j, x_{-j}) - f_0(x_j, x_{-j}) \right|^p dx_j - \eta_p \delta, \quad (3.6)$$

where η_p is defined in the same way as in Proposition 1. Integrating the weak inequality (3.5) over $x_{-j} \in \mathcal{X}^{d-1} \setminus \mathcal{X}_{-j}$, of measure $1 - \nu$, and the strong inequality (3.6) over $\mathcal{X}_{-j}$, of measure ν, we obtain

$$\int_{\mathcal{X}^d} \left| \bar{f}_j^*(x) - f_0(x) \right|^p dx \leq \int_{\mathcal{X}^d} \left| \bar{f}(x) - f_0(x) \right|^p dx - \eta_p \delta \nu. \quad (3.7)$$

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 (3.7), 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

$$\left[\int_{\mathcal{X}^d} \left| \hat{f}^*(x) - f_0(x) \right|^p dx \right]^{1/p} = \left[\int_{\mathcal{X}^d} \left| \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \left(\hat{f}^*(x) - f_0(x) \right) \right|^p dx \right]^{1/p}$$
$$\leq \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \left[\int_{\mathcal{X}^d} \left| \hat{f}_\pi^*(x) - f_0(x) \right|^p dx \right]^{1/p}, \quad (3.8)$$

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

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 estimates, and then we quantify the improvement in a simulation example that mimics the empirical application.

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 asses 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 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%, median, 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) local linear, (c) spline, (d) global polynomial, (e) Fourier, and (f) flexible Fourier methods. For the kernel method, we provide a fit

on a cell-by-cell basis, with each cell corresponding to one month. For the local linear method, we choose a bandwidth of one year and a box kernel. For the spline 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 global polynomial method, we fit a quartic polynomial. For the Fourier method, we employ eight trigonometric terms, with four sines and four cosines. For the flexible Fourier method, we use a quadratic polynomial and four trigonometric terms, with two sines and two 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-F of Figure 1 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 sixteen and yield non-monotonic curves for the highest values of age. The Fourier series have a special difficulty approximating the aperiodic age-height relationship. 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. Moreover, the rearranged estimates necessarily improve upon the original estimates, since, by the theoretical results derived earlier, they are closer to the true functions than the original estimates. We quantify this improvement in the next subsection.

Figure 2 displays similar but more pronounced non-monotonicity patterns for the estimates of the conditional quantile functions. The rearrangement again performs well in delivering curves that improve upon the original estimates and that satisfy the natural monotonicity requirement.

Figures 3-7 illustrate 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, we do not show the multivariate age-quantile and quantile-age rearrangements separately, because

they are very similar to the average multivariate rearrangement.) We see from the contour plots that, for all of the estimation methods considered, the estimated CQP is non-monotone in age and non-monotone in the quantile index at extremal values of this index. The contour plots for the estimates based on the Fourier series best illustrate the non-monotonicity problem. We see that the average multivarite rearrangement fixes the non-monotonicity problem, and delivers 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.

4.2. **Monte-Carlo Illustration.** The following Monte Carlo experiment quantifies the improvement in the estimation/approximation properties of the rearranged estimates relative to the original estimates. The experiment closely matches the empirical application presented above.

Specifically, we consider the 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 CEF and CQP that is monotone in age over the range of 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 growth chart data set. In each replication, we estimate the CEF and CQP using the nonparametric methods described in the previous section.

In Table 1 we report the average L^p errors (for $p = 1, 2, 3, 4$ and ∞) for the original estimates and the rearranged estimates of the CEF. We also report the relative efficiency of the two estimates, measured as the ratio of the average error of the rearranged estimate to the average error of the original estimate. 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 rearrangement. For all of the methods considered, we find that the rearranged curves estimate the true CEF more accurately than the original curves, providing a 2% to 84% reduction in the average error, depending on the method and the norm (i.e. values of p).

In Table 2 we report the average L^p errors for the original estimates of the conditional quantile process and their multivariate rearrangement with respect to the age and quantile index arguments. We also report the ratio of the average error of the rearranged estimate to the average error of the original estimate. 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 multivariate rearrangement. 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.

In Table 3 we report the average L_p error for the univariate rearrangements of the conditional quantile function along either the age argument or the quantile index argument. We also report the ratio of the average error for these rearrangements to the average error of the original estimates. For all of the methods considered, we find that these rearranged curves estimate the true CQP more accurately than the original curves, providing noticeable reductions in the estimation error. Moreover, in this case the rearrangement along the age argument is more effective in reducing the estimation error than the rearrangement along the quantile index. Furthermore, by comparing Tables 2 and 3, we also see that the multivariate rearrangement provides an improvement over the individual univariate rearrangements, yielding estimates of the CQP that are often much closer to the true process.

5. Conclusion

Suppose that a target function is known to be weakly increasing, and we have an original estimate of this function, which is not weakly increasing. Common estimation methods provide estimates with such a property. We show that these estimates can always be improved using rearrangement techniques. The univariate and multivariate rearrangement methods transform the original estimate to a monotonic estimate. The resulting monotonic estimate is in fact closer to the target function in common metrics than the original estimate. We illustrate these theoretical results with a computational example and an empirical example, dealing with estimation of conditional mean and quantile functions of height given age. The rearrangement both monotonizes and improves the original non-monotone estimates. It would be interesting to determine whether this improved estimation/approximation property carries over to other methods of monotonization. We leave this extension for future research.

References

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

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): "Quantile and Probability Curves without Crossing," MIT Working Paper, www.arxiv.org and www.ssrn.com.

———— (2006b): "Rearranging Edgeworth-Cornish-Fisher Expansions," MIT Working Paper, coming to www.arxiv.org and www.ssrn.com.

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

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.

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

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.

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.

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., 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.

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

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.

RAMSAY, J. O., AND B. W. SILVERMAN (2005): *Functional data analysis*. Springer, New York, second edn.

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.

WEI, Y., A. PERE, R. KOENKER, AND X. HE (2006): "Quantile regression methods for reference growth charts," *Stat. Med.*, 25(8), 1369–1382.

TABLE 1. L^p Estimation/Approximation Error of Original and Rearranged Estimates of the Conditional Expectation Function, for $p = 1, 2, 3, 4$, and ∞. Univariate Rearrangement.

p	L_O^p	L_R^p	L_R^p/L_O^p	L_O^p	L_R^p	L_R^p/L_O^p
		A. Kernel			B. Local Polynomial	
1	3.69	1.33	0.36	0.79	0.76	0.96
2	4.80	1.84	0.38	1.00	0.96	0.96
3	5.81	2.46	0.42	1.17	1.13	0.96
4	6.72	3.12	0.46	1.33	1.28	0.96
∞	16.8	9.84	0.58	2.96	2.81	0.95
		C. Splines			D. Quartic	
1	0.87	0.81	0.93	1.33	1.19	0.89
2	1.10	1.02	0.93	1.64	1.46	0.89
3	1.31	1.22	0.93	1.89	1.68	0.89
4	1.52	1.39	0.92	2.10	1.87	0.89
∞	3.72	3.19	0.86	4.38	3.79	0.87
		E. Fourier			F. Flexible Fourier	
1	6.57	3.21	0.49	0.73	0.72	0.97
2	10.7	3.79	0.35	0.91	0.89	0.97
3	15.2	4.24	0.28	1.06	1.04	0.98
4	19.0	4.59	0.24	1.18	1.16	0.98
∞	48.9	7.79	0.16	2.44	2.40	0.98

Notes: The table is based on 10,000 replications.

L_O^p is the L^p error of the original estimate.

L_R^p is the L^p error of the rearranged estimate.

TABLE 2. L^p Estimation/Approximation Error of Original and Rearranged Estimates of the Conditional Quantile Process, for $p = 1, 2, 3, 4$, and ∞. Average Multivariate Rearrangement.

p	L_O^p	L_{RR}^p	L_{RR}^p/L_O^p	L_O^p	L_{RR}^p	L_{RR}^p/L_O^p
		A. Kernel			B. Local Polynomial	
1	5.35	3.13	0.58	1.21	1.09	0.91
2	6.97	4.37	0.63	1.61	1.46	0.91
3	8.40	5.49	0.65	2.03	1.84	0.91
4	9.72	6.49	0.67	2.48	2.24	0.91
∞	34.3	26.4	0.77	12.3	10.4	0.84
		C. Splines			D. Quartic	
1	1.33	1.20	0.90	1.49	1.35	0.90
2	1.78	1.60	0.90	1.87	1.69	0.90
3	2.30	2.03	0.88	2.23	1.99	0.89
4	2.92	2.50	0.86	2.62	2.29	0.87
∞	16.9	12.1	0.72	12.6	8.61	0.68
		E. Fourier			F. Flexible Fourier	
1	6.72	4.18	0.62	1.05	1.00	0.96
2	13.7	5.35	0.39	1.38	1.31	0.95
3	20.8	6.36	0.31	1.72	1.63	0.95
4	26.7	7.25	0.27	2.12	1.98	0.94
∞	84.9	21.9	0.26	10.9	9.13	0.84

Notes: The table is based on 1,000 replications.

L_O^p is the L^p error of the original estimate.

L_{RR}^p is the L^p error of the average multivariate rearranged estimate.

TABLE 3. L^p Estimation/Approximation Error of Rearranged Estimates of the Conditional Quantile Process, for $p = 1, 2, 3, 4, \infty$. Univariate Rearrangements.

p	$L^p_{R_u}$	$L^p_{R_x}$	$L^p_{R_u}/L^p_O$	$L^p_{R_x}/L^p_O$	$L^p_{R_u}$	$L^p_{R_x}$	$L^p_{R_u}/L^p_O$	$L^p_{R_x}/L^p_O$
	A. Kernel				B. Local Polynomial			
1	5.35	3.13	1.00	0.58	1.20	1.10	1.00	0.91
2	6.97	4.37	1.00	0.63	1.60	1.47	1.00	0.91
3	8.40	5.49	1.00	0.65	2.01	1.85	0.99	0.91
4	9.72	6.49	1.00	0.67	2.45	2.26	0.99	0.91
∞	34.3	26.4	1.00	0.77	11.8	10.8	0.96	0.88
	C. Splines				D. Quartic			
1	1.31	1.21	0.99	0.91	1.49	1.35	1.00	0.91
2	1.75	1.63	0.98	0.91	1.87	1.69	1.00	0.90
3	2.24	2.08	0.97	0.90	2.22	2.00	0.99	0.90
4	2.80	2.59	0.96	0.89	2.60	2.30	0.99	0.88
∞	14.4	13.9	0.85	0.82	11.9	9.11	0.95	0.72
	E. Fourier				F. Flexible Fourier			
1	6.71	4.19	1.00	0.62	1.04	1.01	0.99	0.96
2	13.7	5.36	1.00	0.39	1.36	1.32	0.99	0.96
3	20.8	6.37	1.00	0.31	1.70	1.65	0.99	0.96
4	26.7	7.26	1.00	0.27	2.08	2.02	0.98	0.95
∞	84.9	22.2	1.00	0.26	10.0	9.86	0.92	0.91

Notes. The table is based on 1,000 replications.

L^p_O is the L^p error of the original estimate.

$L^p_{R_x}$ is the L^p error of the estimate rearranged in age x.

$L^p_{R_u}$ is the L^p error of the estimate rearranged in the quantile index u.
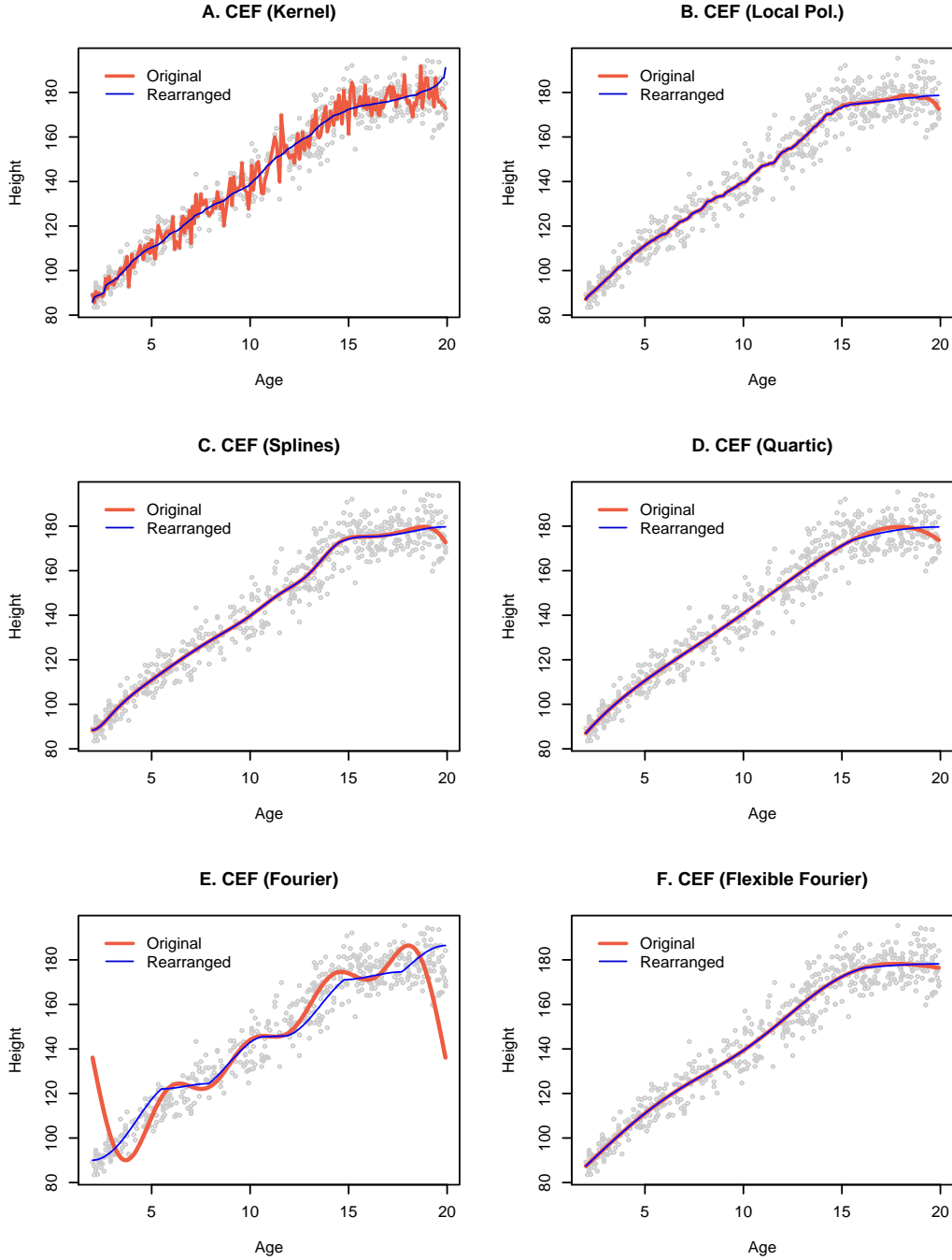
FIGURE 1. Nonparametric estimates of the Conditional Expectation Function (CEF) of height given age and their increasing rearrangements. Nonparametric estimates are obtained using kernel regression (A), locally linear regression (B), cubic B-splines series (C), a four degree polynomial (D), Fourier series (E), and flexible Fourier series (F).
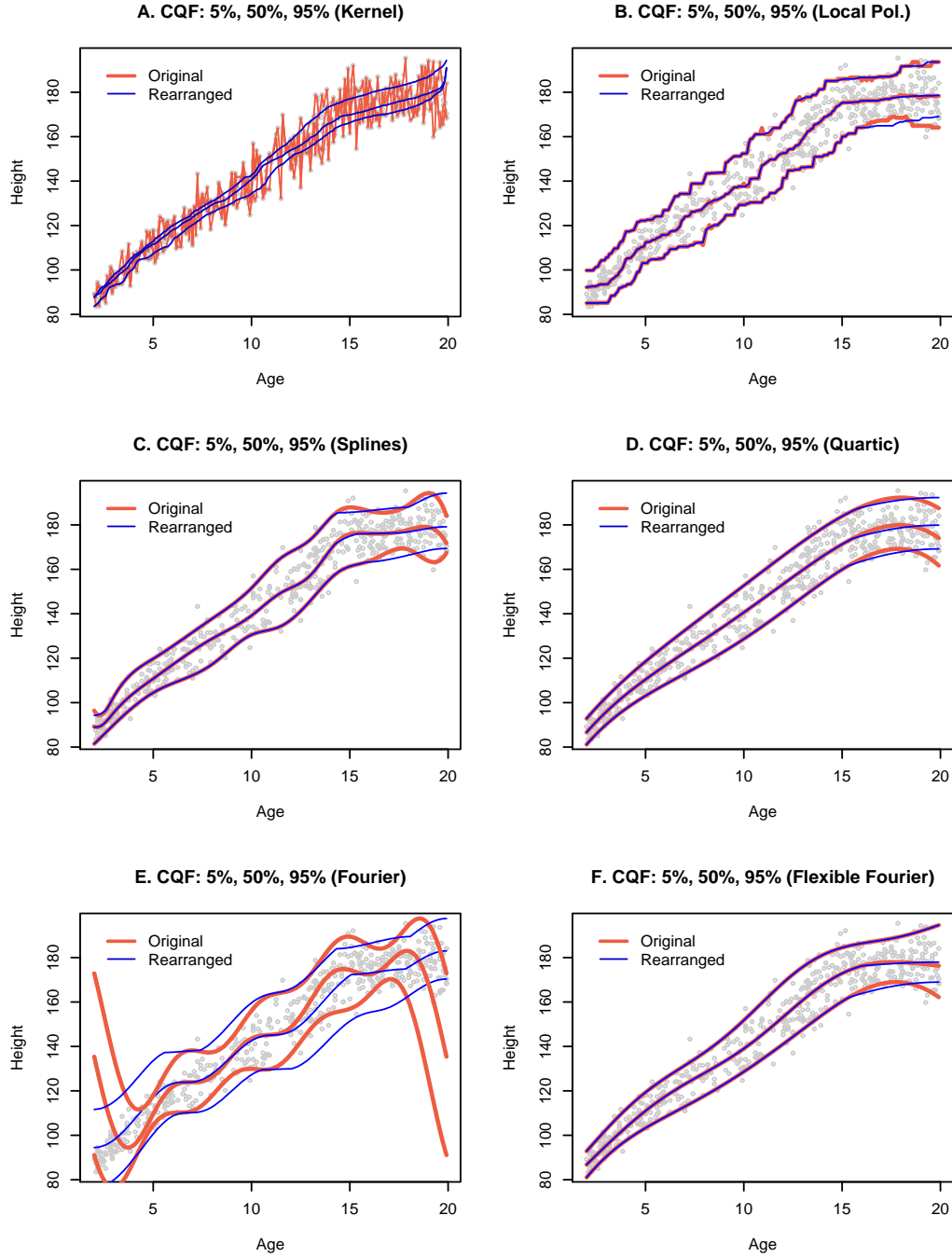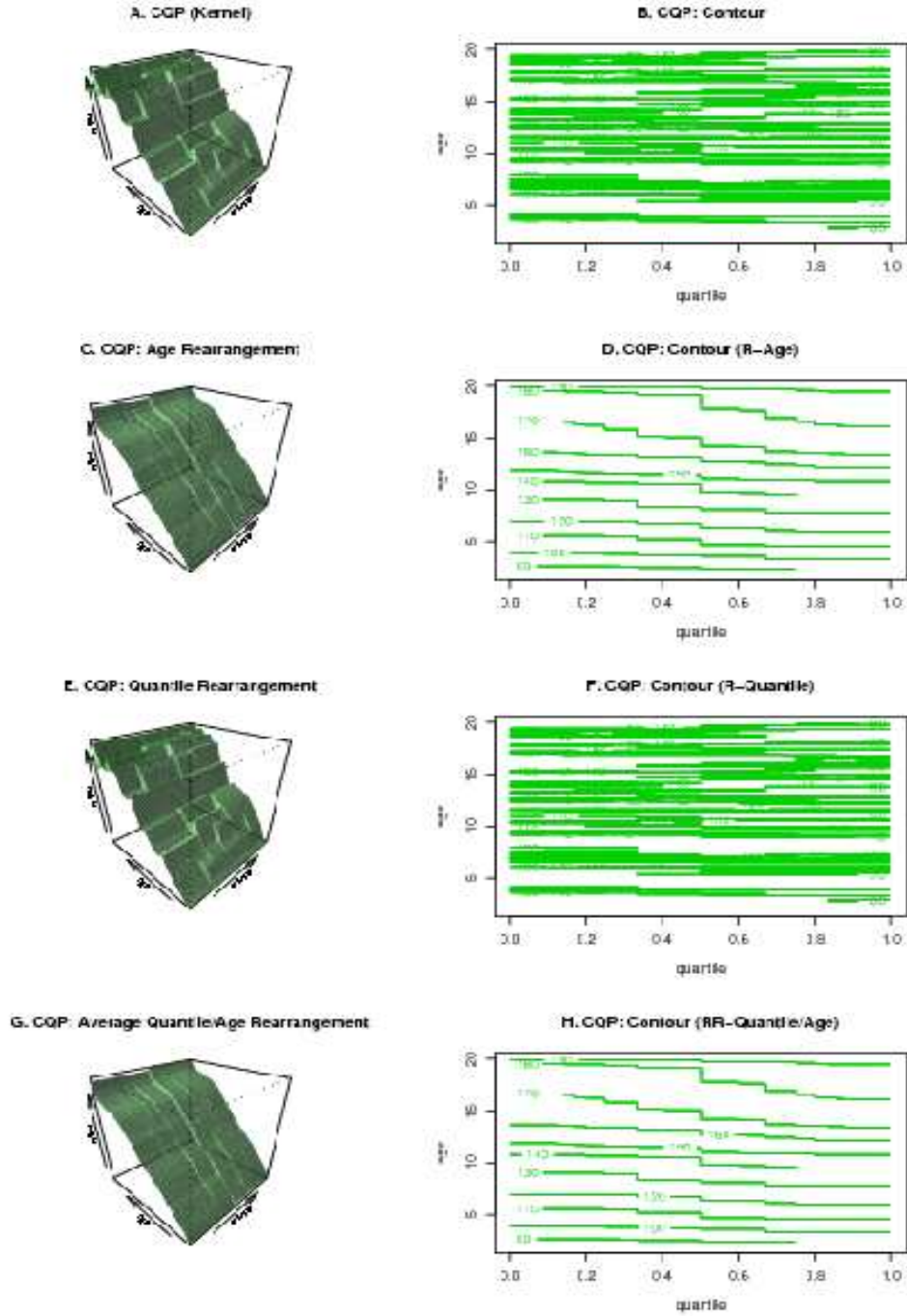
FIGURE 2. Nonparametric estimates of the 5%, 50%, and 95% Conditional Quantile Functions (CQF) of height given age and their increasing rearrangements. Nonparametric estimates are obtained using kernel regression (A), locally linear regression (B), cubic B-splines series (C), a four degree polynomial (D), Fourier series (E), and flexible Fourier series (F).

A. CQP (Kernel)

B. CQP: Contour

C. CQP: Age Rearrangement

D. CQP: Contour (R-Age)

E. CQP: Quantile Rearrangement

F. CQP: Contour (R-Quantile)

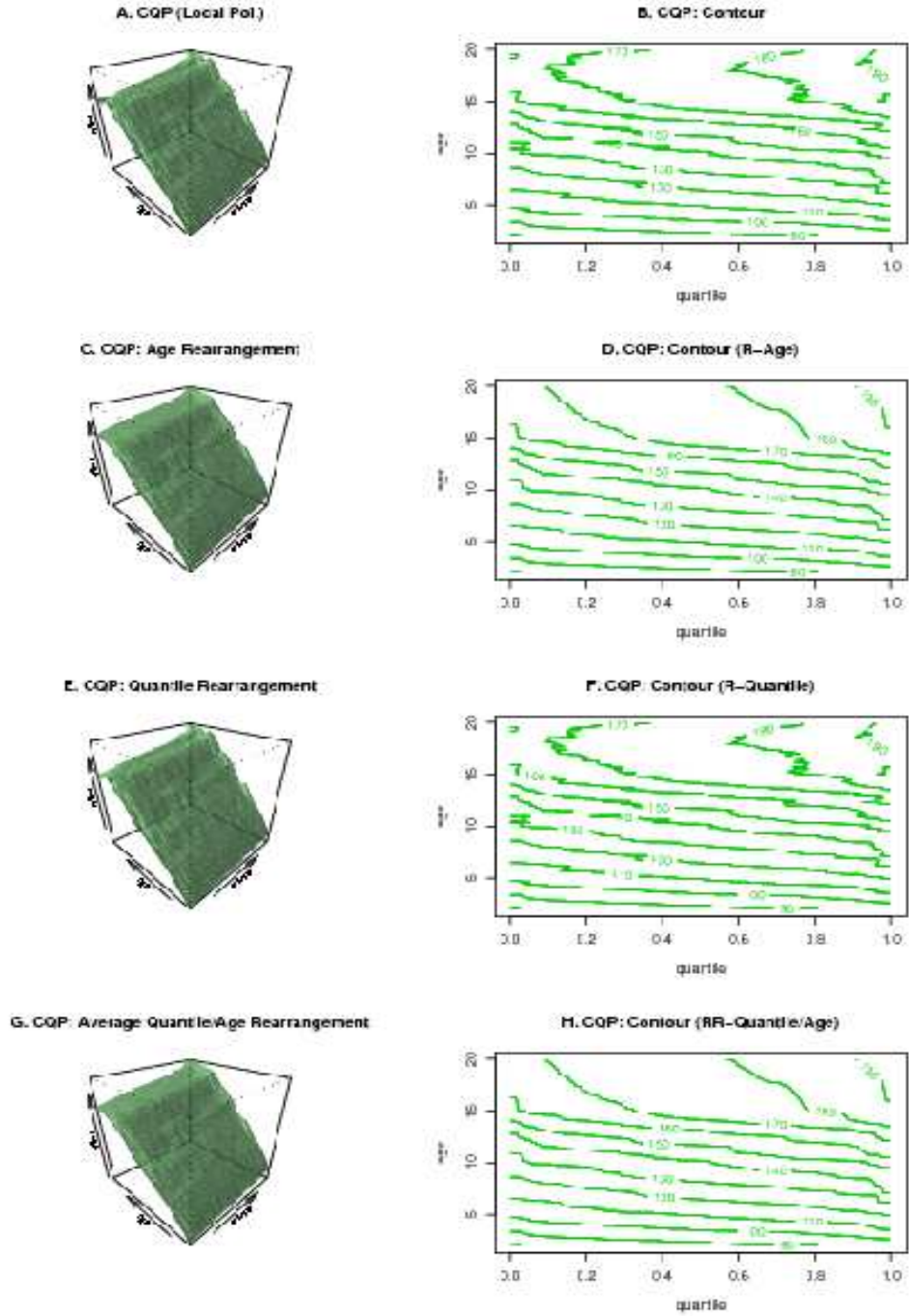G. CQP: Average Quantile/Age Rearrangement

H. CQP: Contour (RR-Quantile/Age)

FIGURE 3. Kernel 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.

A. CQP (Local Pol.)

B. CQP: Contour

C. CQP: Age Rearrangement

D. CQP: Contour (R-Age)

E. CQP: Quantile Rearrangement

F. CQP: Contour (R-Quantile)

G. CQP: Average Quantile/Age Rearrangement

H. CQP: Contour (RR-Quantile/Age)

FIGURE 4. Locally linear 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.
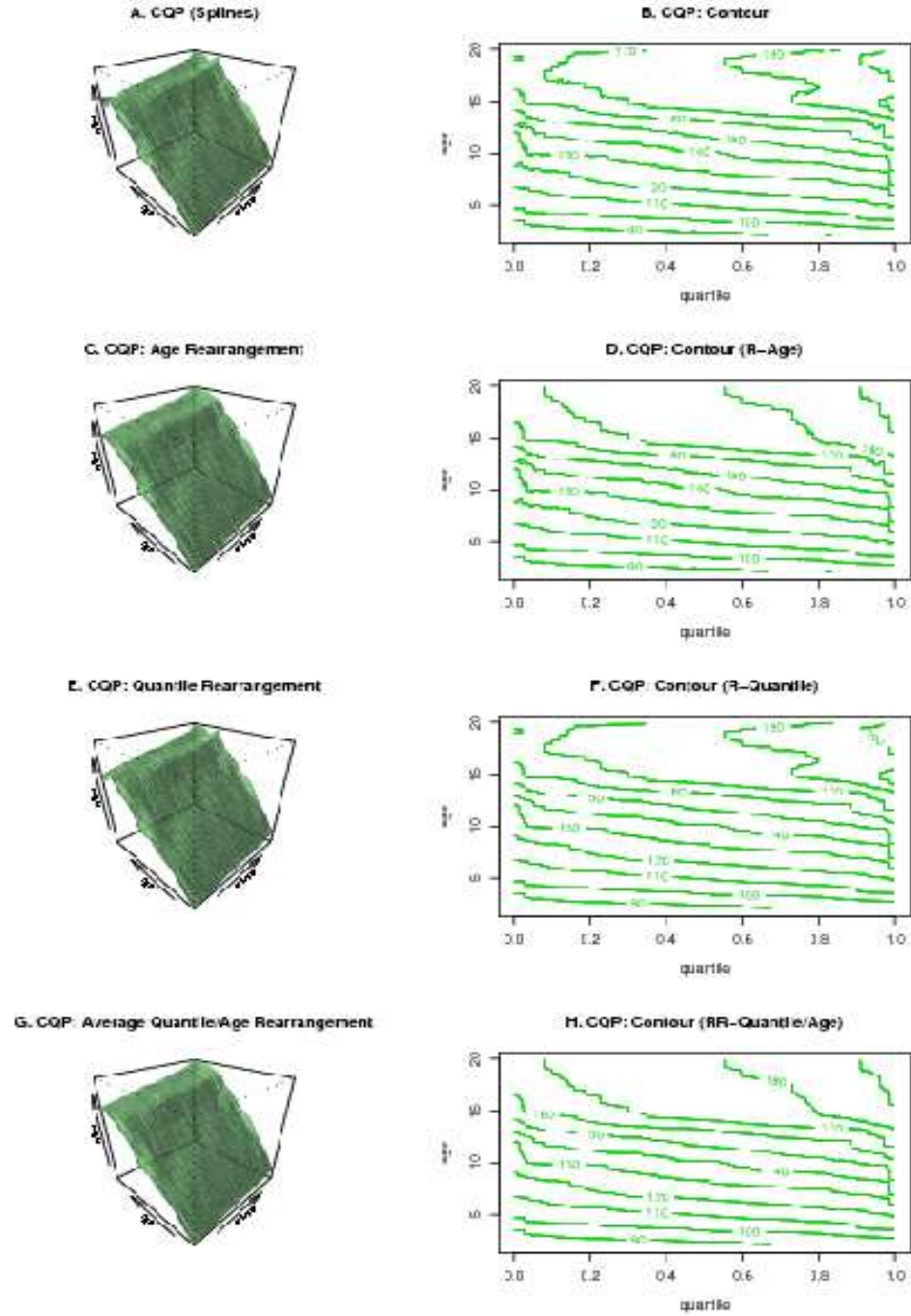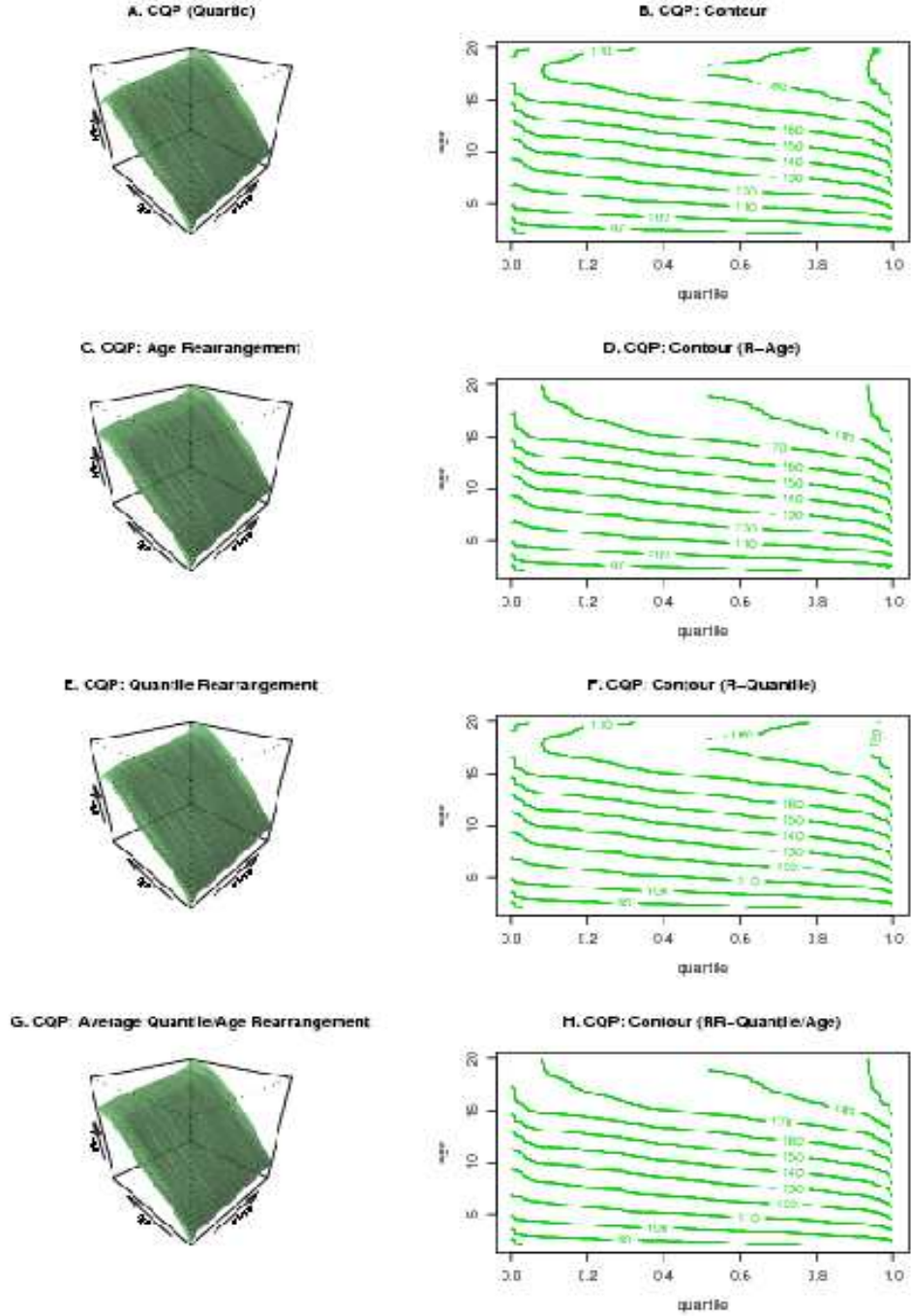
FIGURE 5. Cubic B-splines series estimates of the Conditional Quantile Process (CQP) of height given age and their increasing rearrangements. Panels C and E plot the one dimensional increasing rearrangement along the age and quantile dimension respectively; panel G shows the average multivariate rearrangement.

FIGURE 6. Quartic polynomial 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.
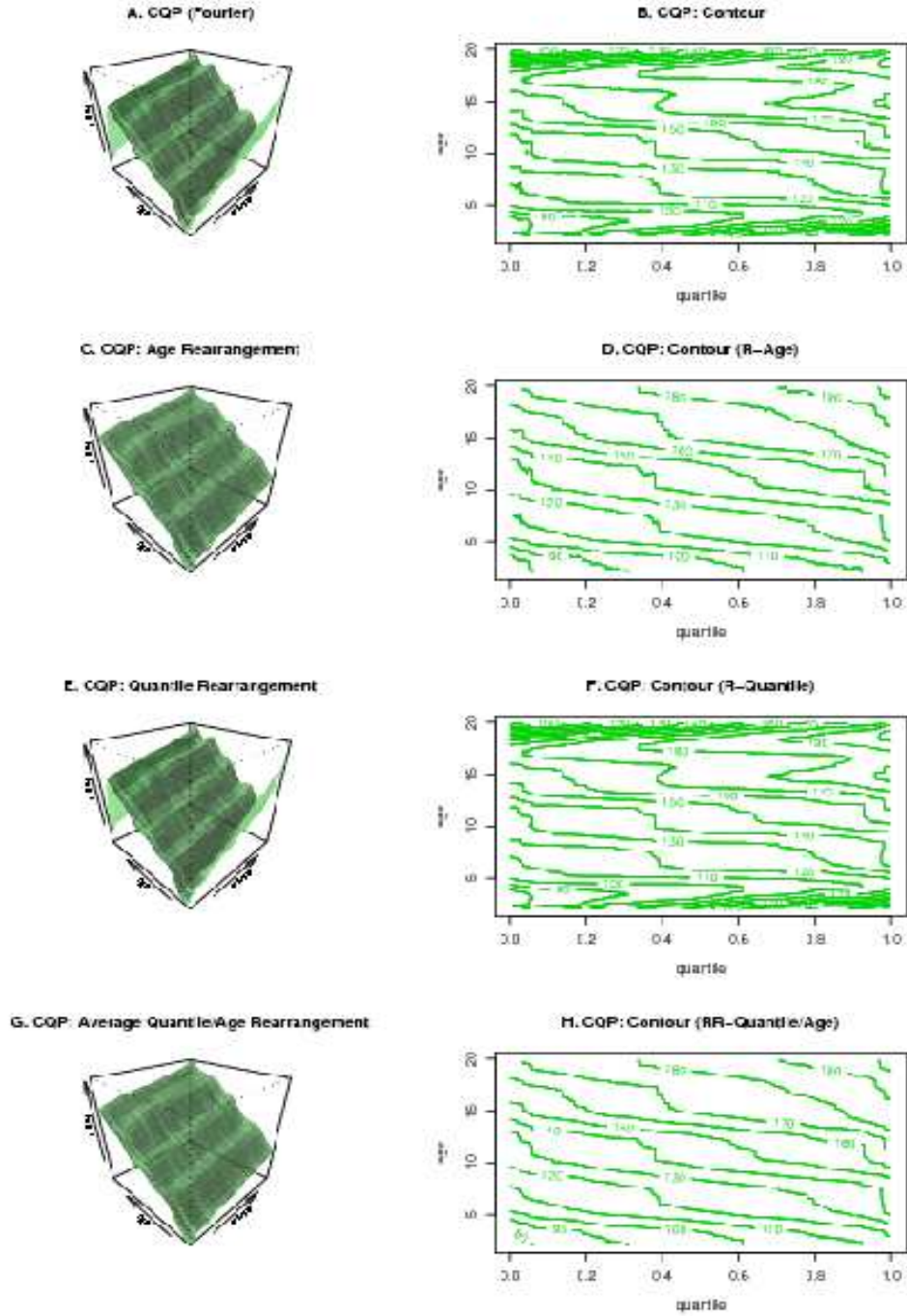
FIGURE 7. Fourier series estimates of the Conditional Quantile Process (CQP) of height given age and their increasing rearrangements. Panels C and E plot the one dimensional increasing rearrangement along the age and quantile dimension respectively; panel G shows the average multivariate rearrangement.
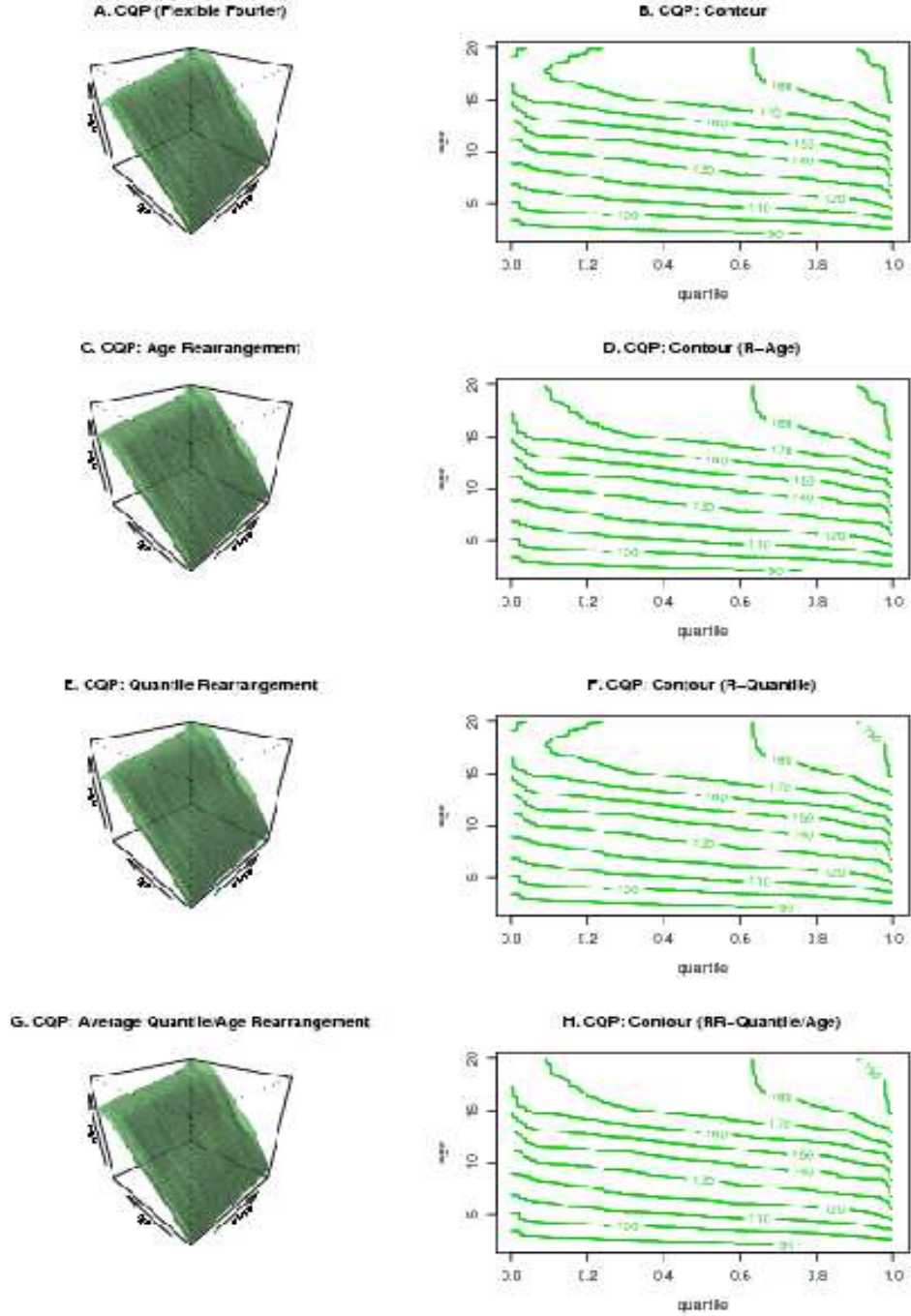
FIGURE 8. Flexible Fourier form series estimates of the Conditional Quantile Process (CQP) of height given age and their increasing rearrangements. Panels C and E plot the one dimensional increasing rearrangement along the age and quantile dimension respectively; panel G shows the average multivariate rearrangement.