



Published in final edited form as:

J Am Stat Assoc. 2017 ; 112(518): 613–622. doi:10.1080/01621459.2016.1149070.

Restoration of Monotonicity Respecting in Dynamic Regression

Yijian Huang

Professor, Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322

Abstract

Dynamic regression models, including the quantile regression model and Aalen’s additive hazards model, are widely adopted to investigate evolving covariate effects. Yet lack of monotonicity respecting with standard estimation procedures remains an outstanding issue. Advances have recently been made, but none provides a complete resolution. In this article, we propose a novel adaptive interpolation method to restore monotonicity respecting, by successively identifying and then interpolating nearest monotonicity-respecting points of an original estimator. Under mild regularity conditions, the resulting regression coefficient estimator is shown to be asymptotically equivalent to the original. Our numerical studies have demonstrated that the proposed estimator is much more smooth and may have better finite-sample efficiency than the original as well as, when available as only in special cases, other competing monotonicity-respecting estimators. Illustration with a clinical study is provided.

Keywords

Adaptive interpolation; Additive complementary log-log survival model; Additive hazards model; Censored quantile regression; Monotone function; Quantile regression

1. Introduction

Quantile regression (Koenker and Bassett 1978) models and estimates potentially dynamic covariate effects on quantiles. As popular as it has become, lack of monotonicity respecting remains an outstanding issue (e.g., He 1997). For example, an estimated 90th percentile may actually exceed its 95th counterpart. Although it may not have received as much attention elsewhere, this issue is general with linear dynamic regression models that relate a response variable Y with covariate vector \mathbf{X} through

$$D_{\mathbf{x}}(t) = \mathbf{x}^{\top} \boldsymbol{\beta}(t) \quad (t \in \mathbb{T}, \mathbf{x} \in \mathbb{X}). \quad (1)$$

Above, $D_{\mathbf{x}}(\cdot)$ is a functional that has a one-to-one mapping with the distribution function of Y given $\mathbf{X} = \mathbf{x}$, $\boldsymbol{\beta}(\cdot)$ is a regression coefficient that quantifies the dynamic covariate effects, \mathbb{T} is an interval index set on the real line, and \mathbb{X} is a covariate space. In the case of quantile

regression, $D_{\mathbf{x}}(\cdot)$ is the quantile function, possibly upon a transformation. One sub-class, due to Peng and Huang (2007), takes a transformation of the survival function as $D_{\mathbf{x}}(\cdot)$; the additive hazards model (Aalen 1980) and the additive complementary log-log survival model (Peng and Huang 2007) are special cases. With all these models, $D_{\mathbf{x}}(t)$ is a monotone function of $t \in \mathbb{T}$ for all $\mathbf{x} \in \mathbb{X}$. However, a standard regression procedure, with coefficient estimator $\hat{\beta}(\cdot)$, may not respect the monotonicity in the sense that $\hat{D}_{\mathbf{x}}(t) = \mathbf{x}^{\top} \hat{\beta}(t)$ is not monotone over $t \in \mathbb{T}$ at least for some $\mathbf{x} \in \mathbb{X}$. Without loss of generality, monotonicity is taken to be monotonically increasing throughout.

To elaborate, consider the typical circumstance that $\inf_{\mathbf{x} \in \mathbb{X}} \mathbf{x}^{\top} \{\beta(t) - \beta(s)\} > 0$ for any given $s < t$ and $\hat{\beta}(\cdot)$ is uniformly consistent for $\beta(\cdot)$. Then, with respect to discrete points in \mathbb{T} and in large sample, lack of monotonicity respecting can be lesser of an issue. Between a fixed point and any other one outside a fixed neighborhood, $\hat{\beta}(\cdot)$ is monotonicity-respecting with respect to them with probability tending to 1 as sample size n increases. So is for any finite set of fixed points. Nevertheless, $\mathbf{x}^{\top} \hat{\beta}(t)$ may not be monotone over all $t \in \mathbb{T}$ for most $\mathbf{x} \in \mathbb{X}$ even in large sample. With standard quantile regression of Koenker and Bassett (1978), $\hat{D}_{\mathbf{x}}(\cdot)$ is monotone when \mathbf{x} is the covariate sample average and by continuity in a neighborhood thereof. Neocleous and Portnoy (2008) argued that the neighborhood, however, is quite small and generally tends to zero in size as n increases. This might also explain the roughness of $\hat{\beta}(\cdot)$ as typically observed.

There are several notable recent developments to tackle this issue with quantile regression. Neocleous and Portnoy (2008) suggested to linearly interpolate the points of an imposed grid, and showed that the resulting estimator is asymptotically monotonicity-respecting and equivalent to the original Koenker–Bassett estimator if the mesh approaches 0 at an appropriate rate. However, monotonicity respecting is not guaranteed in finite sample. Wu and Liu (2009) and Bondell et al. (2010) investigated a somewhat different problem to target the estimation at discrete points only. Moreover, their methods may not adapt easily for other dynamic regression procedures.

As a related problem, curve monotonization has a much longer history in the literature. Many well-known methods, including isotonic regression (e.g., Barlow et al. 1972; Mammen 1991) and monotone regression splines (e.g., Ramsay 1988), were mostly developed for nonparametric regression. Even though they might be adapted, their statistical properties are largely unknown in our context (e.g., Chernozhukov et al. 2010), where the curve of concern $\hat{D}_{\mathbf{x}}(t)$ is dependent across $t \in \mathbb{T}$. Only a few are more relevant. For a baseline cumulative hazard function, Lin and Ying (1994) suggested to monotonize their original estimator by maximizing over the range below, and argued that the asymptotic properties are preserved. Recently, Chernozhukov et al. (2009, 2010) proposed and investigated the rearrangement method to monotonize an estimated curve. The rearranged curve is closer to the estimand in common metrics, and is asymptotically equivalent to the original under regularity conditions. Clearly, a curve monotonization method may be used to monotonize a fitted curve in a dynamic regression, i.e., $\hat{D}_{\mathbf{x}}(t)$ for a given \mathbf{x} , or curves.

However, these monotonized curves typically no longer respect the linearity in model (1). Thus, they may not correspond to a monotonicity-respecting regression coefficient estimator, which is of primary interest as a direct and meaningful measure of covariate effect. If the covariate space is of a special type, we later devise a procedure for monotonicity-respecting regression using a generic curve monotonization method. Nevertheless, this is not a general resolution.

In this article, we propose a novel adaptive interpolation method for monotonicity-respecting estimation of $\beta(\cdot)$ with a general covariate space \mathbb{X} . Starting from a standard estimator, the approach restores monotonicity respecting by construction. This general method applies across various dynamic regressions. As a byproduct, it also reduces to a new curve monotonization method. Section 2 presents the proposed method along with an asymptotic study. Simulations are reported in Section 3, and an illustration given in Section 4. Final remarks are provided in Section 5. Technical details are deferred to the Appendix.

2. The adaptive interpolation method

Standard dynamic regression procedures typically yield estimators that are cadlag step functions, such as cadlag version of the Koenker–Bassett estimator (Huang 2010, Section 3.4) for quantile regression, the estimator of Huang (2010) of censored quantile regression, the generalized Nelson–Aalen estimator under the additive hazards model, and the estimator of Peng and Huang (2007) under their dynamic survival regression models. These estimators are natural, being invariant to a monotonically increasing transformation of the index scale.

We shall thus focus on such an original estimator $\hat{\beta}(\cdot)$; see related discussion later in Remark 3.

For generality, the monotonicity respecting will be restored with respect to an arbitrary covariate sub-space, $\mathbb{X}_s \subseteq \mathbb{X}$. As such, the proposal also accommodates the usual practical situation with unknown \mathbb{X} , where a natural replacement is the convex hull of observed covariate values, or the empirical covariate space $\hat{\mathbb{X}}$. Moreover, the proposal reduces to a curve monotonization method when \mathbb{X}_s is singleton. Thus, \mathbb{X}_s can be random, and the linear space spanned by \mathbb{X}_s may have a lower dimension than that by \mathbb{X} .

2.1 The proposed estimator

Without loss of generality, suppose that the interval set \mathbb{T} is closed; a boundary point can always be included whereby a cadlag function at the boundary point may be set as an appropriate limit. Denote the interior and boundary of a set by circle superscript and ∂ , respectively. Write

$$\mathbb{D} = \{t \in \mathbb{T}^\circ : \hat{\beta}(t-) \neq \hat{\beta}(t)\} \cup \partial\mathbb{T},$$

which contains all the breakpoints of $\hat{\beta}(\cdot)$ along with the boundary points of \mathbb{T} . From a starting point $\tau \in \mathbb{D}$,

$$\max \left[t \in \mathbb{D}: t < \tau, \sup_{\mathbf{x} \in \mathbb{X}_S} \{ \mathbf{x}^\top \hat{\beta}(t) - \mathbf{x}^\top \hat{\beta}(\tau) \} \leq 0 \right], \quad (2)$$

is the left nearest monotonicity-respecting neighbor, in the sense that $\hat{\beta}(\cdot)$ respects the monotonicity at these two points. Recursively each identified point then has its own left nearest monotonicity-respecting neighbor determined, until such a neighbor no longer exists. In the other direction, right neighbors can be similarly obtained; the right nearest monotonicity-respecting neighbor to τ is

$$\min \left[t \in \mathbb{D}: t > \tau, \inf_{\mathbf{x} \in \mathbb{X}_S} \{ \mathbf{x}^\top \hat{\beta}(t) - \mathbf{x}^\top \hat{\beta}(\tau) \} \geq 0 \right]. \quad (3)$$

Denote the collection of all these points, including the starting one τ , by \mathbb{M} . Note that \mathbb{X}_S may be replaced with the vertex set of its convex hull in these neighbor definitions. Also, such nearest monotonicity-respecting neighbors are not necessarily mutual: That point A is the left one to point B does not imply that B is the right one to A .

Interpolating $\hat{\beta}(\cdot)$ linearly between adjacent points in \mathbb{M} then yields a monotonicity-respecting estimator $\tilde{\beta}(\cdot)$. for any t within two adjacent points in \mathbb{M} , say $\tau_l < \tau_r$,

$$\tilde{\beta}(t) = \frac{\tau_r - t}{\tau_r - \tau_l} \hat{\beta}(\tau_l) + \frac{t - \tau_l}{\tau_r - \tau_l} \hat{\beta}(\tau_r). \quad (4)$$

Moreover, set $\tilde{\beta}(t) = \hat{\beta}(\min \mathbb{M})$ for $t < \min \mathbb{M}$ and $\tilde{\beta}(t) = \hat{\beta}(\max \mathbb{M})$ for $t > \max \mathbb{M}$. Unlike $\hat{\beta}(\cdot)$, $\tilde{\beta}(\cdot)$ is a piecewise-linear continuous function.

The proposed adaptive interpolation method is invariant to linear transformation of the covariates in \mathbb{X}_S . In the case that \mathbb{X}_S is a singleton, say $\{\mathbf{x}\}$, the dynamic regression model (1) therefore becomes irrelevant and the proposal reduces to a curve monotonicization procedure, for $\hat{D}_x(t)$, just like Lin and Ying (1994) and Chernozhukov et al. (2009, 2010). Figure 1 illustrates the procedure in this special case. As an appealing characteristic, the monotonicity respecting is induced locally. For the large part, the proposed estimator is thus insensitive to potential tail instability of the original estimator so long as the starting point τ is positioned away from such a tail or tails. In contrast, the Lin–Ying and rearranged estimators are subject to influence by one tail and both tails, respectively. The proposed estimator has no more knots than breakpoints of the original estimator, and is always more smooth. However, unlike the rearranged estimator, the adaptive interpolation estimator is not necessarily closer to the estimand than the original. Nevertheless, it might be so on average as suggested by our numerical studies shown later.

This proposal requires a choice of the starting point, τ . Our asymptotic analysis and numerical studies will later show that a wide range of choices would make little difference in the resulting estimator. Nevertheless, a starting point at which the original estimator is most reliable seems natural; such a point depends on the specific regression model under consideration.

2.2 An asymptotic analysis

Mild regularity conditions are imposed for this purpose.

Condition 1—(Original estimator) $\hat{\beta}(\cdot)$ is consistent for $\beta(\cdot)$ uniformly on some interval $[a, b] \subseteq \mathbb{T}$. Furthermore, $n^{1/2}\{\hat{\beta}(\cdot) - \beta(\cdot)\}$ on $[a, b]$ converges weakly to a tight mean-zero Gaussian process with a continuous covariance function.

Condition 2—(Covariates) \mathbb{X} is bounded.

Condition 3—(Estimand) The derivative $\beta'(t)$ exists for all $t \in [a, b]$, and $\inf_{\mathbf{x} \in \mathbb{X}, t \in [a, b]} \mathbf{x}^\top \beta'(t) > 0$ strictly.

Condition 4—(Starting point) τ is in $[a, b]$ with probability tending to 1.

Uniform consistency and asymptotic normality as in Condition 1 hold in general for standard dynamic regression estimators. See Koenker (2005) for quantile regression, Peng and Huang (2008) and Huang (2010) for censored quantile regression, McKeague (1988) for the additive hazards model, and Peng and Huang (2007) for their dynamic survival regression models, among others.

We first characterize the distance between adjacent points in \mathbb{D} and \mathbb{M} , within interval $[a, b]$. Write the points in $(\mathbb{D} \cap [a, b]) \cup \{a, b\}$ as $a \equiv \nu_1 < \dots < \nu_J \equiv b$ and those in $(\mathbb{M} \cap [a, b]) \cup \{a, b\}$ as $a \equiv \tau_1 < \dots < \tau_K \equiv b$.

Lemma 1: Under Conditions 1, 2, and 3,

$$\max_{1 < j \leq J} (\nu_j - \nu_{j-1}) = o_p(n^{-1/2}). \tag{5}$$

With Condition 4 in addition,

$$\max_{1 < k \leq K} (\tau_k - \tau_{k-1}) = O_p(n^{-1/2}). \tag{6}$$

This result paves the way for establishing the statistical properties of $\tilde{\beta}(\cdot)$. Denote the L^p norm, $p \geq 1$, by $\|\cdot\|_p$.

Theorem 1: Suppose that Conditions 1, 2, 3, and 4 hold. Then,

$$\sup_{t \in [a+\iota, b-\iota]} \|\tilde{\beta}(t) - \hat{\beta}(t)\|_2 = o_p(n^{-1/2}) \quad (7)$$

for any small $\iota > 0$. If $\tau = a$ or $\tau = b$, the asymptotic equivalence of $\tilde{\beta}(\cdot)$ and $\hat{\beta}(\cdot)$ as above holds over an extended interval $[a, b - \iota]$ or $[a + \iota, b]$, respectively.

Since ι can be made arbitrarily small, $\tilde{\beta}(\cdot)$ is asymptotically equivalent to $\hat{\beta}(\cdot)$ over essentially the same interval where the latter is consistent and asymptotically normal.

Accordingly, for inference with the monotonicity-respecting $\tilde{\beta}(\cdot)$, both interval estimation and hypothesis testing using the original $\hat{\beta}(\cdot)$, based on either asymptotic distribution theory or re-sampling methods, can be easily adapted. Specifically, the same standard error may be used for $\tilde{\beta}(\cdot)$, and a pointwise confidence interval or simultaneous confidence band with $\tilde{\beta}(\cdot)$ can be obtained by re-centering the interval or band with $\hat{\beta}(\cdot)$ around $\tilde{\beta}(\cdot)$. Furthermore, a test statistic with $\tilde{\beta}(\cdot)$ in place of $\hat{\beta}(\cdot)$ would typically remain asymptotically equivalent. This inference strategy was used in the numerical studies presented later.

As a result of selecting monotonicity-respecting points, the maximum distance between adjacent points increases from $o_p(n^{-1/2})$ in \mathbb{D} to $O_p(n^{-1/2})$ in \mathbb{M} . If a step function instead is used to connect the points in \mathbb{M} , the resulting monotonicity-respecting estimator then has a bias of $O_p(n^{-1/2})$, which is not sufficient for an asymptotic equivalence result like (7). The linear interpolation in the proposal ensures a negligible bias of $o_p(n^{-1/2})$.

Remark 1: A restricted parameter space is typically associated with improved estimation efficiency. It is not so for the monotonicity-respecting estimation in terms of first-order asymptotic efficiency, although finite-sample improvement might still be possible.

Remark 2: Although $\beta(\cdot)$ is introduced through the dynamic regression model (1), the model is nonessential whereby $\beta(\cdot)$ can be interpreted merely as the limit of $\hat{\beta}(\cdot)$. However, the strict monotonicity of $\mathbf{x}^\top \beta(t)$ over t for all \mathbf{x} , given by Condition 3, is critical, so as to bound the distance between adjacent points in \mathbb{M} and in turn give rise to the asymptotic equivalence of $\tilde{\beta}(\cdot)$ and $\hat{\beta}(\cdot)$.

Remark 3: Monotonicity respecting may be restored for a general original estimator without much additional difficulty. Approximation by a step function is one approach. Alternatively, the proposal may be adapted to include all points in \mathbb{T} as candidates in the nearest monotonicity-respecting neighbor identification. When the original is actually a cadlag step function, the adaptation is not the same as the proposal, however, and our proposed estimator tends to be more smooth.

Remark 4: The proposal shares one ingredient, linear interpolation, with the approach of Neocleous and Portnoy (2008). However, an important distinction is that ours is adaptive so as to guarantee monotonicity respecting in finite sample. Moreover, their results are specific

to quantile regression and implicitly require existence of the second derivative of $\beta(\cdot)$. Finally, our gap between knots, as given by (6), is tighter than their grid mesh in order.

3. Simulations

Simulations were conducted to investigate the performance of our proposal under practical sample size with various dynamic regressions. Overall precision was assessed using maximum absolute error and root mean integrated squared error; for example, in the case of estimator $\tilde{\beta}(\cdot)$ over interval $[c_1, c_2]$, these measures are given by

$$\max_{t \in [c_1, c_2]} \|\tilde{\beta}(t) - \beta(t)\|_\infty, \left\{ \int_{c_1}^{c_2} \|\tilde{\beta}(t) - \beta(t)\|_2^2 dt \right\}^{1/2},$$

respectively.

3-1 Quantile regression models

The first model contained an intercept and two non-constant covariates. The response Y followed an extreme value distribution conditionally. The two covariates were independent and identically distributed as uniform between 0 and 1. Under formulation (1), $D_x(t)$ was the conditional t -th quantile of Y and $\beta(\cdot)$ consisted of an intercept and two slopes:

$$\beta(t) = [\log\{-\log(1-t)\}, t, t^2]^\top.$$

Cadlag version of the Koenker–Bassett estimator was taken as the original, and its standard error was obtained from the multiplier bootstrap of size 200 as in Huang (2010). With the proposed adaptive interpolation method, we considered the covariate space \mathbb{X} or its empirical counterpart $\hat{\mathbb{X}}$. So far as τ is concerned, it might not take 0 or 1 since $\beta(\cdot)$ was not even bounded at the extremes. Instead, any value in the middle would be viable. A natural choice was the left breakpoint of the original estimator closest to 0.5, denoted by $\tau \doteq 0.5$, and we also considered $\tau \doteq 0.8$. Meanwhile, two additional estimators were also studied, each involving a single component of the two in the adaptive interpolation method. One was the linearly interpolated original estimator, and the other was the stairwise monotonicity-respecting estimator that jumps only at the points in \mathbb{M} as obtained with $\hat{\mathbb{X}}$ and $\tau \doteq 0.5$. Sample sizes of 200 and 400 were studied.

Table 1 reports the results on the regression coefficient estimators from 1000 replications. Not surprisingly, the interpolated original estimator behaved similarly to the original one; asymptotically the interpolation is negligible since the adjacent points in \mathbb{D} are sufficiently close to each other, as given by (5). However, the stairwise monotonicity-respecting estimator showed notable bias, highlighting the role of interpolation in the proposed adaptive interpolation method. Among the adaptive interpolation estimators, empirical $\hat{\mathbb{X}}$ versus theoretical \mathbb{X} and $\tau \doteq 0.5$ versus $\tau \doteq 0.8$ made little difference. They all had considerably smaller maximum absolute error on average than the original estimator, although the magnitude was more modest in terms of root mean integrated squared error; these improvements were larger with smaller sample size. Meanwhile, they had roughly 10% as

many knots on average as the breakpoints of the original estimator; the average numbers of breakpoints were 257 and 517 for the original estimator for sample sizes 200 and 400, respectively. At given probability indices, the proposed adaptive interpolation estimators had smaller standard deviation than the original estimator, especially in the circumstance of smaller sample size, but behaved similarly in bias and coverage of Wald confidence interval.

We also investigated estimated conditional quantile functions. The proposed adaptive interpolation method with $\mathbb{X}_s = \mathbb{X}$ yields a monotone estimated conditional quantile function for a given covariate \mathbf{x} . Meanwhile, a curve monotonicization method may also be used to monotonicize $\mathbf{x}^T \hat{\beta}(\cdot)$. Such methods include the adaptive interpolation with $\mathbb{X}_s = \{\mathbf{x}\}$ and the rearrangement of Chernozhukov et al. (2009, 2010). From the same simulations as for Table 1, Table 2 show the estimated conditional quantile functions corresponding to non-constant covariate values of $(1, 0)^T$, $(0, 1)^T$, and $(1, 1)^T$, three out of the four vertices of \mathbb{X} ; the other vertex takes $(0, 0)^T$ and the conditional quantile function is the intercept. Note that the lack of monotonicity with the original estimator $\mathbf{x}^T \hat{\beta}(\cdot)$ might be more serious when \mathbf{x} is farther away from the covariate sample average; see Section 1. Overall, all these different estimated conditional quantile functions had comparable performance at given probability indices. This is expected since they are all first-order asymptotically equivalent. Nevertheless, the adaptive interpolation estimator with $\mathbb{X}_s = \mathbb{X}$ typically had better average maximum absolute error and average root mean integrated squared error, as well as far fewer knots on average. This suggests that the stronger constraint as imposed helps improve finite-sample estimation efficiency.

To evaluate the sensitivity to covariate dimension, seven additional covariates were incorporated to the previous model. All these non-constant covariates were independent and identically distributed as uniform between 0 and 1, and the seven additional coefficients were set to 0 for model comparability. Table 3 shows the simulation results of the original Koenker–Bassett estimator versus the proposed adaptive interpolation estimator, using empirical $\hat{\mathbb{X}}$ and $\tau \doteq 0.5$, with sample size of 400 and from 1000 replications. In comparison with Table 1, the most notable change in the performance of the proposed estimator was the reduced average number of knots relative to that of breakpoints of the original estimator; the average numbers of breakpoints for the proposed and original were 8.4 and 831, respectively. This reduction was partly attributable to the stronger monotonicity-respecting constraint associated with the increase of covariate dimension. Other summary statistics remained largely comparable, and the performance of the proposed method was satisfactory.

3-2 Additive hazards model with simplex covariate space

A curve monotonicization method may be used to attain monotonicity-respecting regression when the covariate space has a simplex convex hull with the same number of affinely independent vertices as the coefficient dimension. In this circumstance, the curves of $D_{\mathbf{x}}(\cdot)$ corresponding to those vertices have a one-to-one linear mapping with the regression coefficient $\beta(\cdot)$. Accordingly, once these estimated curves are monotonicized, they uniquely determine a monotonicity-respecting estimator of $\beta(\cdot)$. This simplex condition is automatically satisfied in a model with an intercept and a single non-constant covariate. We simulated such an additive hazards model and compared so adapted monotonicity-respecting

regression methods with our proposal. Note that the simplex condition, of course, is restrictive in general; for example, it is not satisfied by the covariate spaces of the foregoing quantile regression models.

For the additive hazards model, $D_{\mathbf{x}}(t)$ in (1) was the conditional cumulative hazard function and

$$\beta(t) = (0.1t^2, 0.5t)^\top.$$

The covariate followed the uniform distribution between 0 and 1. Response Y was subject to right censoring, with the censoring time being uniformly distributed between 0 and 8 and independent of Y and the covariate. The response and censoring time were not directly observed but only through their minimum and the censoring indicator. The censoring rate was approximately 24.5%. The generalized Nelson–Aalen estimator was used as the original. Four methods were investigated to restore monotonicity respecting with respect to the empirical covariate space $\hat{\mathbb{X}}$. The first was the adaptive interpolation method taking $\mathbb{X}_S = \hat{\mathbb{X}}$, and the others were adapted from three curve monotonicization procedures: the adaptive interpolation method taking $\mathbb{X}_S = \{\mathbf{x}\}$, i.e., the singleton of a given covariate, the Lin–Ying method, and the rearrangement method. For the adaptive interpolation in both methods, the starting point $\tau = 0$ was a natural choice for this model and it also guaranteed respecting for nonnegative cumulative hazard. The adapted Lin–Ying method ensured this additional property as well, but the adapted rearrangement might not. The martingale-based standard error was adopted for the original estimator, and also used for inference with all the monotonicity-respecting estimators.

Table 4 reports the simulation results with sample size 400 from 1000 replications. On average, the proposed adaptive interpolation estimator with $\mathbb{X}_S = \hat{\mathbb{X}}$ had the smallest maximum absolute error and root mean integrated squared error, and far fewer knots than breakpoints of other estimators. Its performance was otherwise similar to that of the original, the adaptive interpolation with $\mathbb{X}_S = \{\mathbf{x}\}$, and adapted Lin–Ying estimators. The adapted rearrangement estimator, however, performed poorly, with serious bias, larger standard deviation, and confidence interval under-coverage particularly at small t . Apparently, right tail instability in the original estimator had a substantial impact on the rearrangement method; see the discussion in Section 2. This was verified by the improved performance of a limited rearrangement where the original estimator beyond $t = 5$ was curtailed. As a note, $t = 1, 2, 3, 4, 5$ correspond to the 29, 58, 79, 91, 97-th percentiles, respectively, of the marginal Y distribution.

3-3 Remark on computation

All these numerical studies and others exhibited high computational efficiency of the proposed adaptive interpolation method, despite that our implementation was seemingly primitive. Particularly in the case that the empirical covariate space $\hat{\mathbb{X}}$ was adopted, each and every observation in a sample was assessed for the purpose of identifying nearest monotonicity-respecting neighbors as in (2) and (3). Nevertheless, the computation time was

only a small fraction of that for the original estimator. Taking quantile regression as an example, the adaptive interpolation method took only about 3% of the computation time for point estimation of the Koenker–Bassett estimator over a wide range of sample size and covariate dimension; the efficient algorithm of Huang (2010) was used for the latter. This provided little impetus for an obvious and more sophisticated implementation that takes the vertex set of the empirical $\hat{\mathbf{X}}$ instead in the nearest neighbor identification. The vertex set size could be much smaller than the sample size when the covariate dimension is small, say, 2 or 3. However, the difference reduces fast with a higher covariate dimension. Meanwhile, the cost of computing the vertex set escalates tremendously. Therefore, such an implementation may not always fare well and we have not pursued it further beyond the initial exploration.

4. Illustration with a clinical study

Huang (2010) analyzed the Mayo primary biliary cirrhosis study (Fleming and Harrington 1991, Appendix D) using his censored quantile regression procedure. The data consisted of 416 primary biliary cirrhosis patients, with a median follow-up time of 4.74 years and a censoring rate of 61.5%. The survival time, on the logarithmic scale, was regressed over five covariates: age, edema, log(bilirubin), log(albumin), and log(prothrombin time). We applied our proposal to obtain monotonicity-respecting regression.

With censored quantile regression, choice of the starting point τ is not obvious. Typically, $\beta(t)$ becomes non-identifiable as t approaches 1 but the identifiability upper bound is unknown. Huang (2013) suggested an estimate that is less than the identifiability bound with probability tending to 1, yet with its limit reasonably close to the bound nevertheless. For this data set, we obtained such an identifiability bound surrogate, 0.828, corresponding to Huang (2013, Equation 12) with $\kappa = 0.05^{\dim(\mathbf{X})}$; $\dim(\mathbf{X})$ is the dimension of covariate vector \mathbf{X} including the unity element. Our starting point τ was taken to be the left breakpoint of the original estimator closest to half of the identifiability bound surrogate. The monotonicity respecting was restored with respect to the empirical covariate space $\hat{\mathbf{X}}$.

Figure 2 displays both the original estimate of Huang (2010) and our proposed adaptive interpolation estimate for the regression coefficients. The former is fairly rugged, with 267 breakpoints, a large part of which results presumably from random noise. In contrast, the latter is much more smooth, and has only 9 knots with $\min \mathbb{M}$ and $\max \mathbb{M}$ being 0.0397 and 0.856, respectively. Nevertheless, the latter tracks the former quite well, except for the tails as expected. Thus, the monotonicity-respecting estimation might help prevent an over-interpretation of the dynamic effects. Figure 3 shows that estimated conditional quantile functions have their monotonicity restored by the proposed method.

5. Discussion

The proposed method can be applied more generally. Additional pertinent models include those with a mixture of constant and varying effects (e.g., Qian and Peng 2010) and for recurrent events (Fine et al. 2004; Huang and Peng 2009). The proposal may also accommodate time-dependent covariates in, e.g., the additive hazards model, provided that

the covariate space across $t \in \mathbb{T}$ remains the same. In fact, it suffices to consider time-independent covariates only in the monotonicity respecting induction, since monotonicity of the estimated $D_{\mathbf{x}}(\cdot)$ corresponding to any time-dependent covariate process then would also be guaranteed.

Alternative definition for the nearest monotonicity-respecting neighbor was explored. A more stringent left neighbor to starting point $\tau \in \mathbb{D}$ is given by

$$\max \left[t \in \mathbb{D}: t < \tau, \sup_{\mathbf{x} \in \mathbb{X}, s \in \mathbb{D}: s \leq t} \{ \mathbf{x}^\top \hat{\beta}(s) - \mathbf{x}^\top \hat{\beta}(\tau) \} \leq 0 \right],$$

and the right counterpart can be accordingly defined. As such, the monotonicity respecting is also required with respect to the current point and every point beyond the neighbor. Fewer knots would thus be involved in the interpolation. Despite that the asymptotic results remain largely the same, finite-sample performance can be inferior since the estimator might become sensitive to tail behavior of the original estimator; see related discussion in Section 2.

As a small price to pay, our monotonicity-respecting estimator is not invariant to a monotonically increasing transformation of the index scale in the dynamic regression model (1). By the same token, the interpolation of our proposal may be generalized since the linear interpolation on a nonlinearly transformed index scale is no longer linear on the original scale. To be slightly more general, the interpolated value given by (4) at t between adjacent points τ_l and τ_r can be extended to

$$\frac{h_{\tau_l, \tau_r}(\tau_r) - h_{\tau_l, \tau_r}(t)}{h_{\tau_l, \tau_r}(\tau_r) - h_{\tau_l, \tau_r}(\tau_l)} \hat{\beta}(\tau_l) + \frac{h_{\tau_l, \tau_r}(t) - h_{\tau_l, \tau_r}(\tau_l)}{h_{\tau_l, \tau_r}(\tau_r) - h_{\tau_l, \tau_r}(\tau_l)} \hat{\beta}(\tau_r),$$

for a given monotonically increasing function $h_{\tau_l, \tau_r}(\cdot)$ specific to the interval $[\tau_l, \tau_r]$. If $h_{\tau_l, \tau_r}(\cdot)$ is differentiable and its derivative is bounded away from both 0 and ∞ uniformly for all τ_l and τ_r , then the resulting estimator has the same asymptotic equivalence result as $\tilde{\beta}(\cdot)$ given by Theorem 1. This generalization offers the possibility to further subject a monotonicity-respecting estimator to, e.g., certain smoothness, which merits further investigation.

Throughout, we have focused on the linear dynamic regression model (1). Its restoration of monotonicity respecting is distinct from that of a nonparametric dynamic regression model, which imposes little structure on the covariate effects. For the latter, it is a matter of curve monotonicization only, for estimated $D_{\mathbf{x}}(\cdot)$ with given \mathbf{x} . Dette and Volgushev (2008) and Qu and Yoon (2015), among others, recently developed monotone estimation for nonparametric quantile regression. These methods share similar ideas with Chernozhukov et al. (2009, 2010) and Neocleous and Portnoy (2008) to exploit rearrangement and interpolation for monotonicization. Our proposed adaptive interpolation as a curve monotonicization procedure might also be applicable, and in-depth analytical and numerical studies are warranted.

Acknowledgments

The author thanks the reviewers for helpful comments and suggestions that have led to an improvement in the presentation. This research was partially supported by the National Science Foundation grant DMS-1208874 and the National Institutes of Health grants HL113451 and AI050409.

Appendix: Proofs of Lemma 1 and Theorem 1

We first establish a result on $\hat{\beta}(\cdot)$ that will be used repeatedly in the proofs. Write $\phi(\cdot)$ as the limit of $n^{1/2}\{\hat{\beta}(\cdot) - \beta(\cdot)\}$, on $[a, b]$. Denote the first elements of $\phi(\cdot)$, $\hat{\beta}(\cdot)$, and $\beta(\cdot)$ by $\phi^{(1)}(\cdot)$, $\hat{\beta}^{(1)}(\cdot)$, and $\beta^{(1)}(\cdot)$, respectively. Since $n^{1/2}\{\hat{\beta}^{(1)}(\cdot) - \beta^{(1)}(\cdot)\}$ converges weakly to the tight Gaussian process $\phi^{(1)}(\cdot)$ by Condition 1, it is implied that, for every $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \Pr^* \left\{ \sup_{s, t \in [a, b]: \rho(s, t) < \delta} n^{1/2} \left| \{\hat{\beta}^{(1)}(s) - \beta^{(1)}(s)\} - \{\hat{\beta}^{(1)}(t) - \beta^{(1)}(t)\} \right| > \varepsilon \right\} = 0$$

where \Pr^* denotes outer probability and $\rho(s, t) = [\text{var}\{\phi^{(1)}(s) - \phi^{(1)}(t)\}]^{1/2}$ (Kosorok 2008, Theorem 2.1). Meanwhile, the covariance function of $\phi^{(1)}(\cdot)$ is continuous by Condition 1, and hence uniformly continuous by the Heine–Cantor theorem. Therefore, for any $\delta > 0$, there exists $\eta > 0$ such that $|s - t| < \eta$ implies $\rho(s, t) < \delta$ for all $s, t \in [a, b]$. Consequently, $\rho(s, t)$ in the above equation can be replaced with $|s - t|$. Furthermore, the same result applies to each and every other element of $\hat{\beta}(\cdot) - \beta(\cdot)$. Thus, for every $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \Pr^* \left\{ \sup_{s, t \in [a, b]: |s - t| < \delta} n^{1/2} \left\| \{\hat{\beta}(s) - \beta(s)\} - \{\hat{\beta}(t) - \beta(t)\} \right\|_2 > \varepsilon \right\} = 0. \tag{A.1}$$

In words, $n^{1/2}\{\hat{\beta}(\cdot) - \beta(\cdot)\}$ is asymptotically uniformly equicontinuous in probability.

Proof of Lemma 1

Write $c = \inf_{\mathbf{x} \in \mathbb{X}_S, t \in [a, b]} \mathbf{x}^\top \beta'(t)$, where $c > 0$ by Condition 3. Then, for any $\mathbf{x} \in \mathbb{X}_S$ and $1 < j < J$,

$$c(\nu_j - \nu_{j-1}) \leq \mathbf{x}^\top \{\beta(\nu_j) - \beta(\nu_{j-1})\} \leq \|\mathbf{x}\|_2 \|\beta(\nu_j) - \beta(\nu_{j-1})\|_2 = O(\|\beta(\nu_j) - \beta(\nu_{j-1})\|_2)$$

by Cauchy–Schwarz inequality and Condition 2. Meanwhile,

$$\max_{1 < j \leq J} \|\beta(\nu_j) - \beta(\nu_{j-1})\|_2 = \max_{1 < j \leq J} \sup_{s, t \in [\nu_{j-1}, \nu_j]} \|\{\hat{\beta}(s) - \beta(s)\} - \{\hat{\beta}(t) - \beta(t)\}\|_2 \leq 2 \sup_{t \in [a, b]} \|\hat{\beta}(t) - \beta(t)\|_2 = O_p(n^{-1/2})$$

$$\tag{A.2}$$

by Condition 1. Therefore,

$$\max_{1 < j \leq J} (\nu_j - \nu_{j-1}) = O_p(n^{-1/2}).$$

In light of (A.1), the above result implies that the right-hand side of (A.2) is actually $o_p(n^{-1/2})$. Accordingly, the refinement (5) is obtained.

For any $\mathbf{x} \in \mathbb{X}_S$ and any $s, t \in [a, b]$ such that $t - s = n^{-1/2}$,

$$\mathbf{x}^\top \{\hat{\boldsymbol{\beta}}(t) - \hat{\boldsymbol{\beta}}(s)\} = \mathbf{x}^\top \{\boldsymbol{\beta}(t) - \boldsymbol{\beta}(s)\} + \mathbf{x}^\top [\{\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\} - \{\hat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}(s)\}] \geq cn^{-1/2} - \|\mathbf{x}\|_2 \|\{\hat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}(s)\} - \{\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\}\|_2.$$

From (A.1),

$$\sup_{s, t \in [a, b]: t - s = n^{-1/2}} \|\{\hat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}(s)\} - \{\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\}\|_2 = o_p(n^{-1/2}).$$

Therefore, with probability tending to 1,

$$\min_{\mathbf{x} \in \mathbb{X}_S, s, t \in [a, b]: t - s = n^{-1/2}} \mathbf{x}^\top \{\hat{\boldsymbol{\beta}}(t) - \hat{\boldsymbol{\beta}}(s)\} > 0$$

and subsequently

$$\max_{1 < k \leq K} (\tau_k - \tau_{k-1}) \leq n^{-1/2} + \max_{1 < j \leq J} (\nu_j - \nu_{j-1})$$

under Condition 4. Thus, equation (6) is obtained.

Proof of Theorem 1

By definition (4), for $t \in [\tau_j, \tau_{j+1}]$,

$$\begin{aligned}
& \tilde{\beta}(t) - \hat{\beta}(t) \\
&= \frac{\tau_r - t}{\tau_r - \tau_l} [\{\hat{\beta}(\tau_l) \\
&\quad - \beta(\tau_l)\} \\
&\quad - \{\hat{\beta}(t) \\
&\quad - \beta(t)\}] + \frac{t - \tau_l}{\tau_r - \tau_l} [\{\hat{\beta}(\tau_r) \\
&\quad - \beta(\tau_r)\} \\
&\quad - \{\hat{\beta}(t) \\
&\quad - \beta(t)\}] \\
&\quad + \frac{\tau_r - t}{\tau_r - \tau_l} \beta(\tau_l) \\
&\quad + \frac{t - \tau_l}{\tau_r - \tau_l} \beta(\tau_r) - \beta(t).
\end{aligned}$$

Lemma 1 implies that, with probability tending to 1, each and every point in $[a + \iota, b - \iota]$ is within two adjacent points in $\mathbb{M} \cap [a, b]$. Given bound (6) on the gap, result (A.1) implies that the supremum over $t \in [a + \iota, b - \iota]$ of the first line on the right-hand side above is $o_p(n^{-1/2})$. Moreover, by Taylor expansion, the second line above is $o(\tau_r - \tau_l)$. Combining them yields (7).

If the starting point τ is equal to a or b , each and every point in $[a, b - \iota]$ or $[a + \iota, b]$, respectively, then lies within adjacent points in $\mathbb{M} \cap [a, b]$ with probability tending to 1.

Thus, the asymptotic equivalence of $\tilde{\beta}(\cdot)$ and $\hat{\beta}(\cdot)$ holds on the extended interval.

References

- Aalen, OO. Lecture Notes in Statistics. Klonecki, W.Kozek, A., Rosi ski, J., editors. Vol. 2. New York: Springer-Verlag; 1980. p. 1-25.
- Barlow, RE., Bartholomew, DJ., Bremner, JM., Brunk, HD. Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression. New York: Wiley; 1972.
- Bondell HD, Reich BJ, Wang H. Noncrossing Quantile Regression Curve Estimation. *Biometrika*. 2010; 97:825–838. [PubMed: 22822254]
- Chernozhukov V, Fernandez-val I, Galichon A. Improving Point and Interval Estimators of Monotone Functions by Rearrangement. *Biometrika*. 2009; 96:559–575.
- Chernozhukov V, Fernandez-val I, Galichon A. Quantile and Probability Curves Without Crossing. *Econometrica*. 2010; 78:1093–1125.
- Dette H, Volgushev S. Non-crossing Non-parametric Estimates of Quantile Curves. *Journal of the Royal Statistical Society*. 2008; 70:609–627. Ser B
- Fine JP, Yan J, Kosorok MR. Temporal Process Regression. *Biometrika*. 2004; 91:683–703.
- Fleming, TR., Harrington, DP. Counting Processes and Survival Analysis. New York: Wiley; 1991.
- He X. Quantile Curves without Crossing. *American Statistician*. 1997; 51:186–192.

- Huang Y. Quantile Calculus and Censored Regression. *The Annals of Statistics*. 2010; 38:1607–1637. [PubMed: 20592942]
- Huang Y. Fast Censored Linear Regression. *Scandinavian Journal of Statistics*. 2013; 40:789–806.
- Huang Y, Peng L. Accelerated Recurrence Time Models. *Scandinavian Journal of Statistics*. 2009; 36:636–648.
- Koenker, R. *Quantile Regression*. Cambridge: Cambridge University Press; 2005.
- Koenker R, Bassett G. Regression Quantiles. *Econometrica*. 1978; 46:33–50.
- Kosorok, MR. *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer; 2008.
- Lin DY, Ying Z. Semiparametric Analysis of the Additive Risk Model. *Biometrika*. 1994; 81:61–71.
- Mammen E. Estimating a Smooth Monotone Regression Function. *The Annals of Statistics*. 1991; 19:724–740.
- McKeague IW. Asymptotic Theory for Weighted Least Squares Estimators in Aalen’s Additive Risk Model. *Contemporary Mathematics*. 1988; 80:139–152.
- Neocleous T, Portnoy S. On Monotonicity of Regression Quantile Functions. *Statistics and Probability Letters*. 2008; 78:1226–1229.
- Peng L, Huang Y. Survival Analysis with Temporal Covariate Effects. *Biometrika*. 2007; 94:719–733.
- Peng L, Huang Y. Survival Analysis with Quantile Regression Models. *Journal of the American Statistical Association*. 2008; 103:637–649.
- Qian J, Peng L. Quantile Regression with Partially Functional Effects. *Biometrika*. 2010; 97:839–850.
- Qu Z, Yoon J. Nonparametric estimation and Inference on Conditional Quantile Processes. *Journal of Econometrics*. 2015; 185:1–19.
- Ramsay JO. Monotone Regression Splines in Action. *Statistical Science*. 1988; 3:425–441.
- Wu Y, Liu Y. Stepwise Multiple Quantile Regression Estimation Using Non-crossing Constraints. *Statistics and Interface*. 2009; 2:299–310.

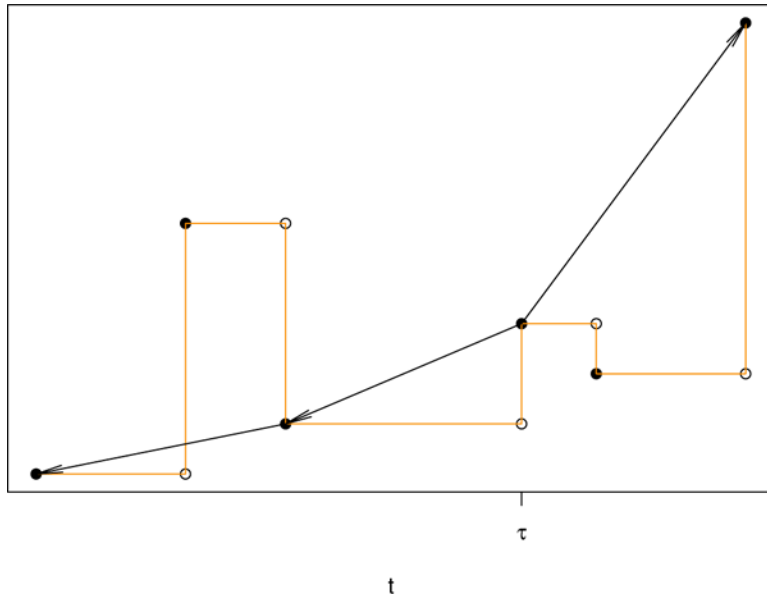


Figure 1. An illustration of the proposed adaptive interpolation method in the case that \mathbb{X}_S is a singleton, $\{\mathbf{x}\}$. The light line is the original $\hat{D}_{\mathbf{x}}(t)$. The dark line is the adaptive interpolation estimate, where each arrow points to a nearest monotonicity-respecting neighbor.

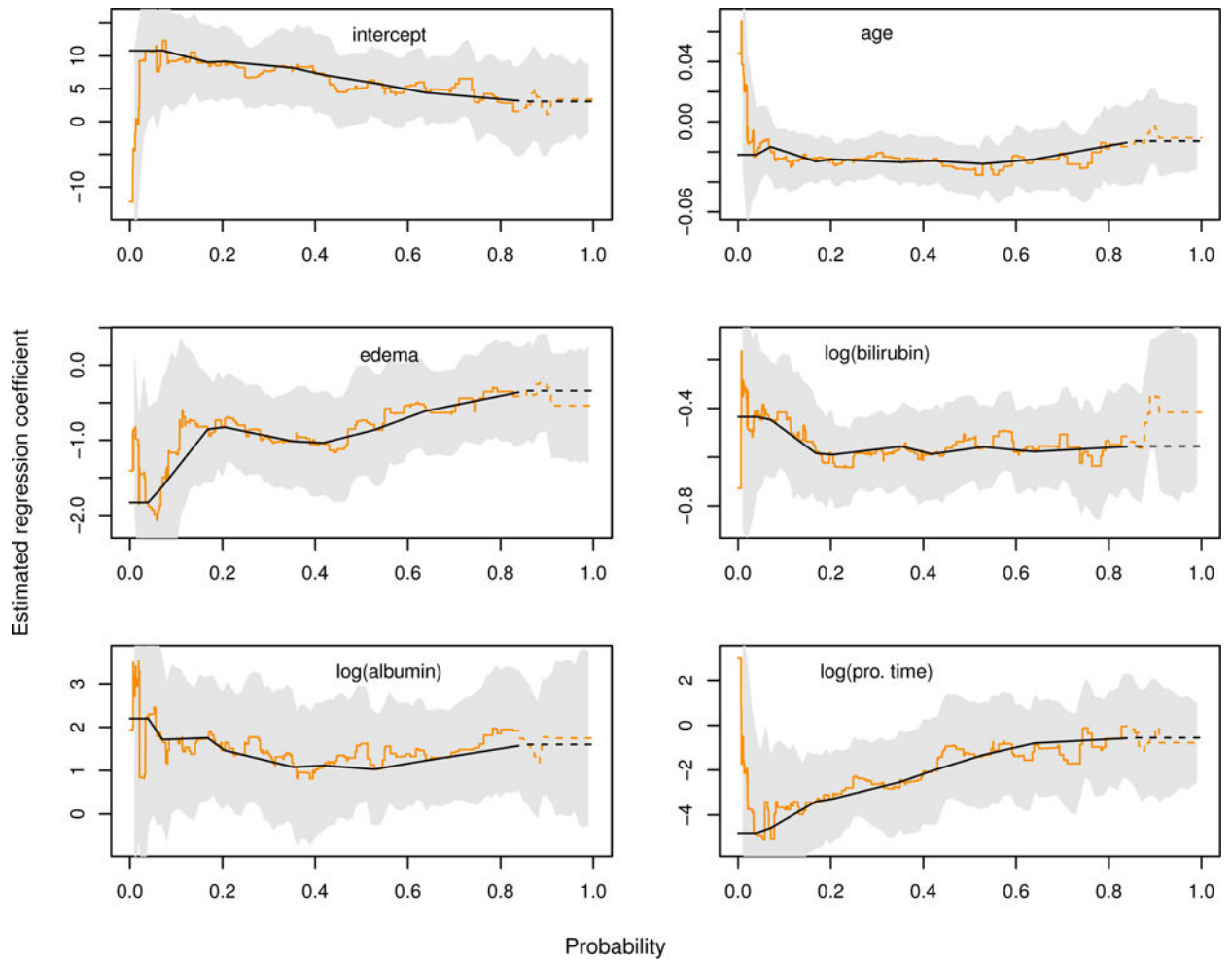


Figure 2. Estimated regression coefficients from the Mayo primary biliary cirrhosis study. The light lines are the original regression coefficient estimate, $\hat{\beta}(\cdot)$, with their pointwise Wald 95% confidence intervals in shaded areas, from censored quantile regression (Huang 2010). The dark lines correspond to the proposed adaptive interpolation estimate $\tilde{\beta}(\cdot)$. Portions of these estimates beyond the identifiability bound surrogate, 0.828, are denoted with dashed lines.

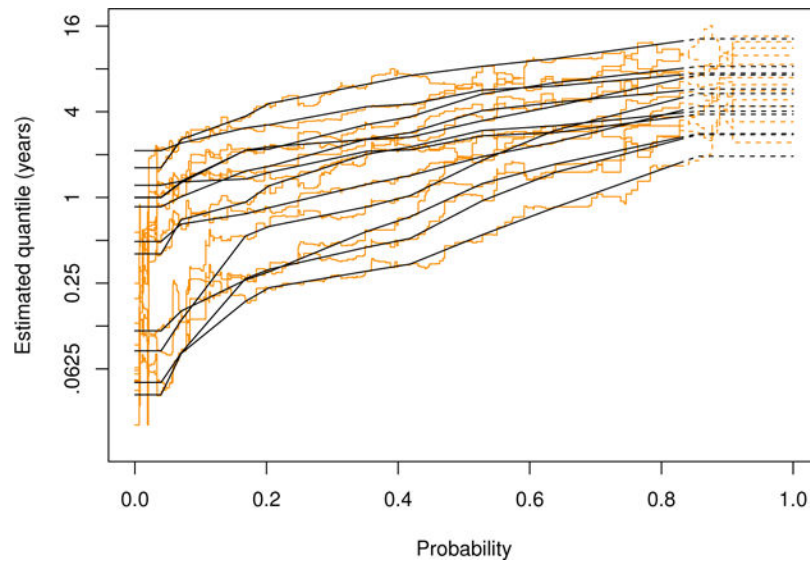


Figure 3. Estimated conditional quantile functions from the Mayo primary biliary cirrhosis study. The original estimates, in light, and those from the adaptive interpolation method, in dark, are shown for 12 covariate values as chosen from the 102 vertices of $\hat{\mathcal{X}}$. Portions of these estimators beyond the identifiability bound surrogate, 0.828, are denoted with dashed lines.

Table 1

Regression coefficient estimation under a quantile regression model

		$\tau = 0.2$			$\tau = 0.5$			$\tau = 0.8$		
sample size: 200										
Orig	B	-40	47	17	-27	28	9	-17	13	-3
	SD	466	631	611	343	470	480	308	440	435
	SE	485	659	657	347	491	493	314	448	451
	CI	93.4	95.1	95.9	94.3	94.7	94.8	94.3	94.6	95.2
IO	MAE	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RMISE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SW	MAE	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
	RMISE	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
AI- \hat{X}	MAE	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87
	RMISE	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
AI- \bar{X}	MAE	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86
	RMISE	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
AI- \hat{X}	MAE	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87
	RMISE	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
sample size: 400										
Orig	B	2	8	-5	-4	8	7	-1	-1	-2
	SD	330	429	446	241	338	320	215	314	304
	SE	338	456	457	242	345	343	219	314	313
	CI	94.2	95.0	94.2	92.9	94.5	96.1	94.2	93.9	94.1
IO	MAE	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RMISE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SW	MAE	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RMISE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

			$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$							
AI- \hat{X}	RMISE	1.03	SD 346	437	458	246	344	320	217	314	310	
	MAE	0.90	B	1	12	0	-3	9	7	11	-6	-12
	RMISE	0.98	SD	324	420	434	240	339	320	210	307	299
AI- \bar{X}	BRK	0.10	CI	94.8	95.1	94.4	92.8	94.5	96.0	94.3	94.7	95.0
	MAE	0.90	B	1	14	-1	-3	9	7	12	-7	-13
	RMISE	0.97	SD	322	419	432	240	339	319	210	307	298
AI- \hat{X}	BRK	0.10	CI	94.9	95.2	94.5	92.8	94.5	96.1	94.3	95.1	95.2
	MAE	0.91	B	1	12	0	-11	19	18	0	1	-1
	RMISE	0.97	SD	324	420	434	239	335	316	213	313	301
AI- \bar{X}	BRK	0.10	CI	94.8	95.1	94.4	93.3	94.7	96.0	94.2	93.6	94.6

Orig: the original estimator; IO: the interpolated original estimator; SW: the stairwise monotonicity-respecting estimator; AI- \hat{X} : the proposed adaptive interpolation estimator with \hat{X}_S and, unless otherwise specified, $\tau \neq 0.5$.

MAE: maximum absolute error over [0.1, 0.9]; RMISE: root mean integrated squared error over [0.1, 0.9]; BRK: number of knots or breakpoints. All are average measures, reported as relative to the original estimator. B: Empirical bias ($\times 1000$); SD: Empirical standard deviation ($\times 1000$); SE: Average standard error ($\times 1000$); CI: Empirical coverage (%) of 95% Wald confidence interval.

Three columns under each τ value correspond to the estimated intercept and two slopes.

Table 2

Estimation of conditional quantile functions

		$t = 0.2$			$t = 0.5$			$t = 0.8$			
		sample size: 200									
Orig		B	7	-23	24	1	-18	10	-4	-20	-7
		SD	472	466	488	364	359	388	333	344	356
		SE	508	496	510	378	374	396	337	347	368
		CI	93.9	94.0	94.9	93.6	94.2	93.9	93.8	93.7	94.7
AI- \bar{X}	MAE	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
	RMISE	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	BRK	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
AI-{x}	MAE	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
	RMISE	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	BRK	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
RA	MAE	0.91	0.91	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
	RMISE	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	BRK	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		sample size: 400									
Orig		B	10	-3	5	4	3	12	-2	-3	-4
		SD	335	323	342	252	250	268	236	240	255
		SE	348	342	359	263	261	279	235	244	256
		CI	94.5	95.1	95.3	95.0	94.4	94.6	92.8	93.7	93.4
AI- \bar{X}	MAE	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
	RMISE	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	BRK	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
AI-{x}	MAE	0.93	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	RMISE	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	BRK	0.40	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
RA	MAE	0.94	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	RMISE	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

	$\tau = 0.2$			$\tau = 0.5$			$\tau = 0.8$						
BRK	1.00	1.00	1.00	CI	94.7	95.6	95.4	94.9	94.3	94.7	93.3	93.9	93.5

Orig: the original estimator; AI- \bar{X}_S : the adaptive interpolation estimator with \bar{X}_S and $\tau = 0.5$; RA: the rearranged estimator.

MAE: maximum absolute error over [0.1, 0.9]; RMISE: root mean integrated squared error over [0.1, 0.9]; BRK: number of knots or breakpoints. All are average measures, reported as relative to the original estimator. B: Empirical bias ($\times 1000$); SD: Empirical standard deviation ($\times 1000$); SE: Average standard error ($\times 1000$); CI: Empirical coverage (%) of 95% Wald confidence interval.

Three columns as grouped correspond to the estimated conditional quantile functions with non-constant covariate vectors $(1, 0)^T$, $(0, 1)^T$, and $(1, 1)^T$.

Table 3

Regression coefficient estimation under a quantile regression model with ten coefficients

Orig	τ													
	$\tau = 0.2$			$\tau = 0.5$				$\tau = 0.8$						
B	25	1	-13	-8	8	-9	-15	-8	-18	-13	-8	-11		
SD	620	442	450	448	493	344	347	330	460	307	310	312		
SE	701	466	465	463	536	360	358	357	477	318	321	319		
CI	96.8	96.6	96.2	95.3	96.2	95.0	95.6	96.4	95.2	94.3	94.6	95.6		
MAE	0.87	-27	14	-9	-4	7	-9	-15	-8	4	-22	-6		
RMISE	0.95	SD	593	414	434	431	493	343	347	329	431	287	297	287
BRK	0.01	CI	97.8	97.1	96.4	95.9	96.2	95.2	95.5	96.3	96.1	95.9	96.6	

Orig: the original estimator; $AI-\hat{\Sigma}$: the proposed adaptive interpolation estimator with $\hat{\Sigma}$ and $\tau \neq 0.5$.

MAE: maximum absolute error over $[0.1, 0.9]$; RMISE: root mean integrated squared error over $[0.1, 0.9]$; BRK: number of knots or breakpoints. All are average measures, reported as relative to the original estimator. B: Empirical bias ($\times 1000$); SD: Empirical standard deviation ($\times 1000$); SE: Average standard error ($\times 1000$); CI: Empirical coverage (%) of 95% Wald confidence interval.

Under each τ value are the first four estimated coefficients; the fourth is representative of the last seven.

Table 4

Regression coefficient estimation under an additive hazards model

		$t = 1$	$t = 2$	$t = 3$	$t = 4$						
Orig	B	1	-6	3	-13	5	-15	-12	-4		
	SD	52	119	104	232	187	444	354	863		
	SE	52	116	104	231	186	426	340	824		
	CI	93.8	94.2	95.0	94.2	95.4	94.8	94.0	94.9		
AI- $\hat{\Sigma}$	MAE	0.83	B	9	-19	11	-22	15	-20	1	36
	RMISE	0.94	SD	48	112	101	224	179	417	326	772
	BRK	0.31	CI	96.3	94.8	95.0	95.2	96.6	95.5	95.5	96.9
<u>Adapted estimators from curve monotization</u>											
AI-{ \mathbf{x} }	MAE	0.92	B	10	-10	14	-14	27	-7	44	49
	RMISE	0.99	SD	49	116	102	230	184	438	352	847
	BRK	0.94	CI	96.2	94.7	95.2	94.8	96.5	94.8	95.2	95.5
LY	MAE	0.92	B	9	-12	11	-17	19	-14	18	32
	RMISE	0.97	SD	49	115	101	228	183	434	344	834
	BRK	0.94	CI	96.3	94.8	95.5	94.8	96.3	95.0	94.8	95.9
RA	MAE	5.75	B	-29	-329	-5	-73	-16	-86	-64	-147
	RMISE	4.05	SD	607	4819	109	356	197	479	349	799
	BRK	1.96	CI	95.0	86.9	93.5	89.2	93.3	91.8	92.8	94.9
RA, limited	MAE	0.87	B	2	-7	3	-14	2	-17	-25	-34
	RMISE	0.96	SD	49	116	102	230	183	431	337	776
	BRK	1.93	CI	95.5	94.5	95.4	94.4	95.7	95.4	94.9	96.3

Orig: the original estimator; AI- $\hat{\Sigma}$: the proposed adaptive interpolation estimator with $\hat{\Sigma}$; AI-{ \mathbf{x} }: the adaptive interpolation method with $\hat{\Sigma}, S$ taking the singleton of a given covariate; LY: the Lin-Ying method; RA: the rearrangement method; RA, limited: the rearrangement limited to interval [0, 5].

MAE: maximum absolute error over [0, 4]; RMISE: root mean integrated squared error over [0, 4]; BRK: number of knots or breakpoints. All are average measures, reported as relative to the original estimator. B: Empirical bias ($\times 1000$); SD: Empirical standard deviation ($\times 1000$); SE: Average standard error ($\times 1000$); CI: Empirical coverage (%) of 95% Wald confidence interval.

Two columns under each t value correspond to the estimated intercept and slope.