Quantile Regression: 40 Years On

Roger Koenker

Department of Economics, University of Illinois, Urbana, IL 61801, USA; email: rkoenker@uiuc.edu

xxxxxx 0000. 00:1–25

Keywords

quantile regression, treatment effects, heterogeneity, causal inference

Abstract

Since Quetelet's work in the 19th century social science has iconified "the average man," that hypothetical man without qualities who is comfortable with his head in the oven, and his feet in a bucket of ice. Conventional statistical methods, since Quetelet, have sought to estimate the effects of policy treatments for this average man. But such effects are often quite heterogenous: medical treatments may improve life expectancy, but also impose serious short term risks; reducing class sizes may improve performance of good students, but not help weaker ones or vice versa. Quantile regression methods can help to explore these heterogeneous effects. Some recent developments in quantile regression methods are surveyed below.

Contents

1. Introduction

Quantiles offer a convenient way to summarize univariate probability distributions as exemplified by Tukey's ubiquitous boxplots. In contrast to moments, which characterize global features of the distribution and are consequently strongly influenced by tail behavior, quantiles are inherently local and are nearly impervious to small perturbations of distributional mass. We can move mass around above and below the median without disturbing it at all, provided of course that mass is not transferred from above the median to below, or vice-versa. This locality of the quantiles is highly advantageous for the same reasons that locally supported basis functions are advantageous in nonparametric regression, because it assures a form of robustness that is lacking in many conventional statistical procedures, notably those based on minimizing sums of squared residuals.

When there are covariates – and there are almost always covariates when econometric problems get interesting – we can't rely on sorting as a strategy for computing quantiles. Instead, fortunately, there is a simple, elegant optimization alternative. Univariate quantiles emerge as solutions to the piecewise linear expected loss problem,

$$\min_a \mathbb{E}\rho_\tau(Y - a),$$

where $\rho_\tau(u) = (\tau - I(u < 0))u$ and $\tau \in (0,1)$. When the distribution of Y admits a unique τth quantile we may differentiate,

$$\mathbb{E}\rho_\tau(Y - a) = \tau \int_a^\infty (y - a)dF_Y(y) + (\tau - 1)\int_{-\infty}^a (y - a)dF_Y(y)$$

to obtain the first order condition,

$$\begin{aligned} 0 &= \tau \int_\alpha^\infty dF_Y(y) + (\tau - 1)\int_{-\infty}^\alpha dF_Y(y) \\ &= F_Y(\alpha) - \tau, \end{aligned}$$

so $\alpha = F_Y^{-1}(\tau)$. When there are multiple values such that $F_Y(y) = \tau$, it is conventional to choose the smallest, i.e., $\alpha = \inf\{y : F_Y(y) \geq \tau\}$. Corresponding to these population quantiles are analogous expressions for the sample quantiles with F_Y replaced by the empirical distribution function $F_n(y) = n^{-1}\sum_{i=1}^n I(Y_i \leq y)$. Admissibility of the univariate sample quantile under the loss ρ_τ was considered by Fox and Rubin (1964), but the origin of such solutions under asymmetric linear loss go back at least to Edgeworth (1888a).

Regression estimators minimizing sums of absolute residuals also have a long history. Already in the 18th century, Boscovich, and somewhat later Laplace, advocated a form of bivariate regression that constrained the mean residual to be zero, and minimized the sum of absolute residuals to find the remaining slope parameter estimate. A century later Edgeworth (1888b) proposed removing the intercept constraint and determining both slope and intercept parameters by minimizing the sum of absolute residuals. He provided an effective algorithm for computing the estimator that anticipates modern simplex methods. For further details see Koenker (2017). Edgeworth's proposal languished until it was revived in the 1950's when it was recognized as a linear program. An early application of median regression in economics appears in the work of Arrow and Hoffenberg (1959) who found it convenient for estimating input-output coefficients subject to positivity constraints. Although there was a general recognition that the median, or ℓ_1, or LAD approach had the advantage that it was more resistant to outliers than the usual least squares estimator, a drawback of the approach in addition to the unfamiliarity of its computational methods was the absence of a formal inference apparatus.

Nor, as far as I am aware, was there any recognition that it might be interesting to consider quantile regression models other than the median. Gib Bassett and I began exploring this territory in the mid-1970's. We started with the observation that as long as the model "contained an intercept," that is that the linear span of the covariate/design matrix, X, included a constant vector, then solutions to the regression analogue of our elementary problem,

$$\min_{b \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - x_i^\top b),$$

had the property that roughly τn of the residuals, $r_i = y_i - x_i^\top \hat{b},\ i = 1, \cdots, n$ would be positive and $(1-\tau)n$ would be negative. This follows immediately from the subgradient condition requiring that at the optimum, $\hat{\beta}(\tau)$,

$$0 \in \partial_b \sum_{i=1}^n \rho_\tau(y_i - x_i^\top b)|_{b=\hat{\beta}(\tau)}.$$

Here, $\partial_b \rho_\tau(y_i - x_i^\top b) = -\psi_\tau(y_i - x_i^\top b)x_i$ with $\psi_\tau(u) = \tau - I(u < 0)$ for $u \neq 0$, and is set-valued, with $\partial_b \rho_\tau(y_i - x_i^\top b) = [-\tau, 1-\tau]x_i$, when the residual is zero. When the observations are in "general position" so no more than p observations lie on a hyperplane of of dimension p in the regression sample space, the subgradient condition implies that τ must lie between N/n and $(N+p)/n$ where N is the number of observations below the fitted hyperplane, i.e., having strictly negative residuals. This seemed to justify our conjecture that solutions $\hat{\beta}(\tau)$ of such problems could be considered analogues of the sample quantiles for the linear model, estimating the parameters of models that specified affine conditional quantile functions for $Y|X$.

Inference proved to be a somewhat harder nut to crack. We began be deriving a combinatorial expression for the finite sample density of $\hat{\beta}(\tau)$ based on the foregoing gradient optimality condition. Since this involved a summation over all $\binom{n}{p}$ elementary subset solutions that corresponded to exact fits of p observations, it didn't seem to be terribly practical at first. But eventually we were able to show that this density had a simple Gaussian limiting form that fully justified the regression quantile terminology that we had begun to use. In due course these results appeared in Koenker and Bassett (1978).

Since then many people have contributed to an effort that has gradually built an extensive toolbox for estimation and inference about conditional quantile models. In the remaining pages I will try to briefly survey these developments and suggest a few areas that seem ripe for future development.

2. Inference for Conditional Quantile Models

A fundamental precept of statistics is that estimates of effect magnitudes should be accompanied by some assessment of the precision of these estimates. In this section we will review a variety of methods that have evolved to address this task for quantile regression.

2.1. Binary Treatment Effects

The simplest quantile regression setting is the binary treatment response or two-sample model, where we have a treatment indicator, D_i that takes the value 1 for "treated" observations and 0 for "control" observations. In the classical mean treatment version interest focuses exclusively on the difference in the means of the two samples,

$$\mathbb{E}Y_i|D_i = \alpha + \beta D_i$$

This is typically justified by the location shift model expressed as,

$$Y_i = \alpha + \beta D_i + u_i$$

where the u_i's are either implicitly or explicitly assumed to be independent and identically distributed. Thus, effectively, the treatment is thought to shift the entire response distribution in lockstep by the amount, β.

In contrast, the quantile treatment effect model,

$$Q_{Y|D}(\tau|D) = \alpha(\tau) + \beta(\tau)D$$

recognizes that the distribution of the response can be arbitrarily different under the treatment and control regimes. In this formulation $\alpha(\tau)$ denotes the quantile function of the response for controls, $F^{-1}_{Y|D}(\tau)$, and $\beta(\tau)$ denotes the *difference* between the quantile functions of the treatment and control response: $F^{-1}_{Y|D=1}(\tau) - F^{-1}_{Y|D=0}(\tau)$. This QTE is closely related to the Lehmann (1974) proposal to generalize the mean treatment effect model by considering the horizontal difference between the treatment and control distribution functions, which he defined as the function $\Delta(y)$ such that,

$$F_{Y|D=1}(y) = F_{Y|D=0}(y + \Delta(y)).$$

Thus, the scalar mean treatment effect becomes a functional object capable of fully describing the difference between the treatment and control distributions.

To this point we have been silent about possible endogoneity/selection issues that might arise regarding treatment assignment, we will address such issues in Section 7 below. Since we are almost inevitably unable to observe both treatment and control response for individual subjects, we are consequently constrained to inference about marginal distributions. The presumption underlying the QTE is that a control subject with response $\alpha(\tau)$ will, if treated, have have response $\alpha(\tau) + \beta(\tau)$. This presumption is sometimes referred to as "rank invariance," as for example, in Heckman, Smith, and Clements (1997).

As long as treatment is randomly assigned estimation of the QTE is easily implemented. Given a sample $\{(y_i, d_i) : \; i = 1, \cdots, n\}$ we can simply solve for,

$$(\hat{\alpha}(\tau), \hat{\beta}(\tau)) = \text{argmin}_{(a,b)} \sum_{i=1}^{n} \rho_\tau(y_i - a - bd_i).$$

For this we don't even need any linear programming machinery since the problem separates into two distinct problems with solutions given by the ordinary sample quantile for the control and treatment samples. See Koenker (2005) for further details. As a consequence, inference about the QTE in the binary treatment model can also rely on classical large sample theory for the ordinary sample quantiles. Since the two samples are independent we have that $\hat{\beta}(\tau)$ has finite dimensional asymptotic distributions,

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \rightsquigarrow \mathcal{N}(0, \lambda_0 \Omega^0 + \lambda_1 \Omega^1),$$

where $\Omega_{ij}^k = (\tau_i - \tau_i \wedge \tau_j)/(f_k(F_k^{-1}(\tau_i)) f_k(F_k^{-1}(\tau_j)))$ for $k = 1, 2$ and $i, j : 1, \cdots, p$, and $\lambda_k = n/n_k$, provided that the relative sample sizes, n_k/n stay bounded away from zero and one as $n \to \infty$. Of course, this begs the question of how to estimate the matrices, Ω^k, since they involve the conditional density functions of the two samples. This has spawned a rather extensive literature, and a brief overview is provided in Koenker (2005).

The foregoing theory enables us to construct pointwise confidence bands for the QTE using the estimated covariance matrix. Uniform bands pose somewhat more of a challenge; one attractive approach would be to employ the asymptotic version of the Hotelling (1939) approach described in Koenker (2011). Various other resampling approaches have also been recently suggested notably by Belloni, Chernozhukov, and Kato (2016) and Hagemann (2016). A recent survey of resampling methods for quantile regression is provided by He (2017).

Given the traditional emphasis placed on location shift models of treatment response, e.g. Cox (1984), it is of some interest to explore tests of this hypothetical model. Such tests are closely related to classical goodness of fit tests involving estimated parameters. One approach to such testing, following Khmaladze (1981) is described in Koenker and Xiao (2002).

2.2. Multiple Treatments, Concomitant Covariates and Interactions

Expanding the binary treatment paradigm to permit multiple treatment options raises some new issues especially from the testing perspective, however QTEs can still be based on univariate sample quantile differences and therefore confidence regions can be based on theory essentially similar to that already described. A tantalizing problem of increasing significance in many fields is that of treatment assignment: Given an estimated model of treatment effects how should we go about assigning new subjects to various treatment regimes? Such questions, especially in the medical arena, require answers to thorny risk assessment questions where a distributional perspective on heterogeneous treatment effects can be crucial. A novel perspective on these issues is offered in recent work of Wang, Zhou, Song, and Sherwood (2016) based partially on Manski (2004).

When there are concomitant covariates in addition to the treatment indicator variables more new questions arise. If treatment assignment is fully randomized it is tempting to simply ignore these covariates; this is the viewpoint articulated by Freedman (2008), who argues that bias induced by misspecified introduction of extraneous covariate effects is likely to be more damaging than any benefits that may accrue from variance reduction. This argument has at least equal force in the quantile regression setting as it does for mean regression. Presumably randomization leaves us with treatment D that is stochastically independent of other covariates, say X, so further conditioning on X won't help and may hinder when the functional form of the X conditioning is ill chosen. Of course when treatment is assigned "on observables" X then the case for their inclusion is much more compelling. Kadane and Seidenfeld (1996) provide a nice discussion of this in the light of Student's (1931) infamous critique of the Lanarkshire milk experiment. Rather than relying on such selection-on-observables arguments several authors have opted instead for propensity score reweighting. An early example of this is the work of Lipsitz, Fitzmaurice, Molenberghs, and Zhao (1997) with later contributions by Firpo (2007) and others. The extensive recent work on so-called "doubly-robust" methods that combine these approaches could also be employed as recently suggested by Diaz (2016).

Somewhat neglected in the econometrics literature on treatment reponse and program evaluation is the potentially important role of interactions of covariates with treatment variables. Although interactions feature prominently in the classical analysis of variance literature and appear in some recent high-dimensional linear model research, econometrics has tended to focus attention on main effects of treatment. Interactions, if present, must play an essential role in post-analysis treatment assignment. More work needs to be done to develop better diagnostic tools to incorporate such effects. Cox (1984) offers an extensive agenda of open research topics many of which could be fruitfully extended to the quantile regression setting.

2.3. Method of Quantiles

It is not uncommon to face quantile regression settings with exclusively discrete covariates. In such cases, we can consider each cell of the covariate space, that is each distinct vector of covariates, as determining a separate subsample. As long as the sample sizes in each of these cells is reasonably large we can compute cell specific sample quantiles, each of which can be expected to be approximately Gaussian. When one or more of the discrete covariates arise from binning continuous covariates like age, or job tenure, and we are willing to consider imposing a linearity condition, or some weaker parametric restriction on these cell specific quantiles, it is natural to consider weighted least squares estimation of the restricted model. This is the approach proposed by Chamberlain (1994), and applied more recently by Bassett, Tam, and Knight (2002) and Knight and Bassett (2007).

Because conditional quantile functions completely characterize all that is observable about univariate conditional distributions they provide natural building blocks for structural models. Just as linear least squares estimation of reduced form models constitute a foundation for structural estimation of Gaussian linear simultaneous equation models, quantile regression provides a foundation for nonparametric structural models. This perspective has been elaborated in recent work of Matzkin (2015).

2.4. Nonlinear (in parameters) Quantile Regression

Once in a while we may be faced with specifications of conditional quantile models that are nonlinear in parameters,

$$Q_{Y_i}(\tau|x) = g(x, \theta(\tau))$$

which can be estimated in the immediately obvious manner,

$$\hat{\theta}(\tau) = \mathrm{argmin}_{\theta} \sum_{i=1}^{n} \rho_\tau(y_i - g(x_i, \theta)).$$

Prime examples of such circumstances are the Powell (1986) estimator of the (Tobit) censored regression model, and the Manski (1975) maximum score estimator of the binary response model. In both cases we have a linear in parameters latent response model that posits,

$$Q_{Y_i^*}(\tau|x) = x^\top \beta,$$

but the observable response, $Y_i = \max\{0, Y_i^*\}$ in the former case, and $Y_i = I(Y_i^* > 0)$ in the latter. Since $Q_{h(Y)}(\tau) = h(Q_Y(\tau))$ for any monotone transformation, h, see e.g. Koenker (2005) p. 39, it follows that the Powell estimator,

$$\hat{\beta}(\tau) = \mathrm{argmin}_{\beta} \sum \rho_\tau(y_i - \max\{0, x_i^\top \beta\})$$

and the Manski estimator,

$$\hat{\beta}(\tau) = \mathrm{argmin}_{\|\beta\|=1} \sum \rho_\tau(y_i - I(x_i^\top \beta > 0))$$

consistently estimates the parameters of the latent variable model, up to scale in the latter case.

Other parametric transformation models offer further examples. The venerable Box-Cox power family of transformations, adapted to the quantile regression setting asserts that,

$$Q_{h(Y,\lambda)}(\tau|X) = x^\top \beta(\tau)$$

where $h(y, \lambda) = (y^\lambda - 1)/\lambda$. Of course if λ is known we can easily estimate the linear parameters, $\beta(\tau)$, however joint estimation of $(\lambda(\tau), \beta(\tau))$ requires more effort. Machado and Mata (2000) propose estimating the nonlinear model,

$$Q_Y(\tau|X) = h_\lambda^{-1}(x^\top \beta(\tau)),$$

where $h_\lambda^{-1}(z) = (\lambda z + 1)^{1/\lambda}$, and Fitzenberger, Wilke, and Zhang (2009) suggest a modification to account for circumstances in which $\lambda x_i^\top \beta + 1 < 0$. More recently Mu and He (2007) have proposed an alternative estimator in which λ is estimated by minimizing a sum of squared cumsum residuals. Performance of these methods is sensitive to heterogeneity of the conditional density of the response as would be expected based of the large sample theory we have already sketched. A third option that doesn't appear to have been explored in the literature is to simply rescale the response by dividing by its geometric mean, $\tilde{y}_i = y_i/\bar{y}$, where $\bar{y} = (\prod y_i)^{1/n}$ and then estimate λ by solving,

$$\min_{\lambda, \beta} \sum \rho_\tau(h(\tilde{y}_i, \lambda) - x_i^\top \beta).$$

The rescaling of the response accounts for the Jacobian term from the Box-Cox transformation of the response, as in the conventional mean regression setting. This formulation may provide a more homogeneous conditional density in some applications. Optimization over the scalar λ is easily handled via grid search or other naive methods. For more complicated nonlinear in parameter models it is possible to use iterative versions of the interior point methods that underlie linear in parameter fitting as described in Koenker and Park (1996).

3. Nonparametric Quantile Regression

There is an extensive literature on nonparametric quantile regression that relaxes the strict linearity in covariate assumptions of the foregoing methods while preserving the convenient linear in parameters structure that facilitates efficient computation. Chaudhuri (1991) considers the asymptotic behavior of locally polynomial quantile regression estimators and establishes conditions under which these estimators achieve optimal rates of convergence. Subsequently, work of Lee (2003) and Lee, Mammen, and Park (2010) extended the locally polynomial approach to partially linear and additive models respectively.

As nonparametric quantile regression models become more conplex local fitting and backfitting to accommodate new components become more burdensome and the literature has evolved toward sieve methods. See, for example the influential early work of Stone (1994) and the survey of Chen (2007). Basis function expansions can be readily adapted to particular applications and more easily incorporate partially linear and additive components. The obvious challenge is the control the parametric dimension of the resulting models. Penalty methods, particularly the ℓ_1 penalty of Donoho, Chen, and Saunders (1998) and Tibshirani (1996), have emerged as critical tools for dimension reduction. The lasso is especially convenient in the quantile regression setting since it maintains the linear programming structure of the original problem. This was a primary motivation for the use of total variation roughness penalties in Koenker, Ng, and Portnoy (1994) and Koenker and Mizera (2004) where ℓ_1 penalties were imposed on linear transforms of model parameters, effectively controlling total variation of the derivatives of the fitted functions. A crucial aspect of the computational strategy underlying these methods is the sparse linear algebra employed to represent high dimensional design matrices and to solve systems of linear equations required at each iteration of the interior point algorithms used for fitting. This is particularly evident in applications like that of Koenker (2011) where multiple additive components result in several thousand model parameters. In such cases there may be several Lagrangian parameters controlling the additive nonparametric components as well as a more conventional lasso λ that controls the effective number of active linear covariate effects.

A variety of proposals have been made for how to choose these penalty parameters, but I think that it is fair to say that no consensus has been reached. In prior work I have recommended some form of information criterion in which model dimension is represented by an estimate of the number of observations interpolated by the fitted model. This is a variant of the Meyer and Woodroofe (2000) divergence measure of dimension since,

$$\mathrm{div}(\hat{y}) = \sum_{i=1}^{n} \partial \hat{y}_i / \partial y_i,$$

has summands that take the value one when y_i is interpolated and zero otherwise in the quantile regression setting. However, the proposal of Belloni and Chernozhukov (2011) which constructs a reference distribution for λ's based on a pivotal representation of the gradient condition seems to be a very attractive alternative approach.

Even more formidable than λ-selection is the task of post-selection inference in these high dimensional nonparametric settings. Consequently, this topic has spawned considerable recent research and controversy. Most of this work has focused on resampling methods as exemplified in the quantile regression context by Belloni, Chernozhukov, and Kato (2015) and Belloni, Chernozhukov, and Kato (2016). I believe that the Hotelling tube methods described in Koenker (2011) offer an attractive alternative for some applications. Simulation evidence on their performance for construction of uniform confidence bands for univariate nonparametric components is provided there as well as discussion of an application to modeling sources of malnutrition in India.

4. Time-series Models

Econometric timeseries analysis has traditionally relied on Gaussian models that exclusively employ first and second moment information. However it is now widely recognized that asymmetries and heavy tail behavior, features that are essentially invisible when estimating Gaussian models, can be revealed with the aid of quantile regression methods. Koenker and Xiao (2006) consider quantile autoregressive (QAR) models of the form,

$$Q_{Y_t}(\tau|\mathcal{F}_t) = \alpha_0(\tau) + \sum_{i=1}^{q} \alpha_i(\tau) Y_{t-i}.$$

When the $\alpha_i(\tau)$ do not depend upon τ for $i = 1, \cdots, q$, we have the familiar iid error autoregression with errors having quantile function α_0. More generally, we have a random coefficient QAR(q) model with Y_t generated as,

$$Y_t = \alpha_0(U_t) + \sum_{i=1}^{q} \alpha_i(U_t) Y_{t-i}.$$

with $U_t \sim U[0,1]$. This is a rather special random coefficient model, however, since all its coefficients are driven by the same iid uniform random variables. In the terminology of Schmeidler (1986) the coefficients are comonotonic. The simplest case of the QAR(1) model,

$$Y_t = \alpha_0(U_t) + \alpha_1(U_t) Y_{t-1}.$$

is instructive. Let $\mu_i = \int_0, 1\alpha_i(t)dt$ and $\omega_i^2 = \int_0^1 \alpha_i^2(t)dt$ with $\omega_0 < \infty$ and $\omega_1 < 1$, then Y_t is covariance stationary with $n^{-1}\sum(Y_t - \mu_y)^2 \rightsquigarrow \mathcal{N}(0, \omega_y^2)$ where $\mu_y = \mu_0/(1-\mu_1)$, and $\omega_y^2 = \omega_0^2(1+\mu_1)/((1-\mu_0)(1-\omega_1^2))$.

To illustrate, suppose that $\alpha_1(\tau) = \min\{\frac{1}{2} + 5\tau, 1\}$ and $\alpha_0 = \Phi^{-1}(\tau)$. Simulating the model by drawing a sequence of iid random uniforms we see that runs of $U_t > 0.1$ behave precisely as if the series were a standard Gaussian unit root model. However, as soon as we see a $U_t < 0.1$, the model's mean reversion tendency kicks in, and stationarity is salvaged. This simple example demonstrates the capability of QAR models to mimic some features of common nonstationary time series while preserving essential features of stationarity.

In Figure 1 we illustrate the estimated $\hat{\alpha}_1(\tau)$ process from the augmented Dickey-Fuller type model,

$$Q_{Y_t}(\tau|\mathcal{F}_t) = \alpha_0(\tau) + \alpha_1(\tau) Y_{t-1} + \sum_{j=1}^{4} \delta_j(\tau) \Delta Y_{t-j}.$$

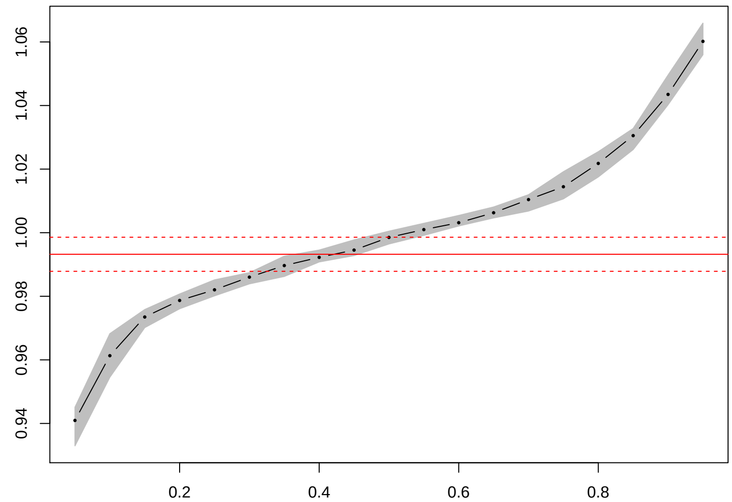

Figure 1

Estimated QAR $\hat{\alpha}_1(\tau)$ coefficient for three-month US Treasury bills based on monthly data 1971-2015.

where Y_t is the three month US Treasury bill rate observed monthly over the period 1971-2015. The horizontal solid line represents the least squares estimate of 0.99 strongly suggesting unit root behavior. Evidence from the QAR estimates clearly contradicts the constant coefficient unit root hypothesis, although in this case the explosive behavior of $\hat{\alpha}_1(\tau)$ for $\tau > 0.5$ compensates for the mean reversion tendency when $\tau < 0.5$ and we are left on the edge of the QAR stationarity conditions in this example.

An obvious critique of the linear QAR model comes from the observation that when the QAR slope coefficient depends upon τ there must be a sub-region of the support of Y_t for which the ordering of quantiles is reversed. Thus, the linear in lagged Y_t's formulation must be regarded, at best, as a local approximation. See, for example, the discussion and response of Koenker and Xiao (2006). One remedy for this predicament is to resort to nonlinear formulations of the QAR model in lagged Y_t's. Chen, Koenker, and Xiao (2009) explore one approach to models of this type based on copula specifications.

Complementary to the time domain formulations of the QAR model is the relatively recent development of frequency domain methods. Building on earlier work of Li (2008) and Li (2012), Hagemann (2011) and Kley, Volgushev, Dette, and Hallin (2016) have proposed variants of quantile spectral analysis. The initial proposal of Li considered the harmonic quantile regression model,

$$Q_{Y_t}(\tau, \omega_j) = \alpha_0(\tau, \omega_j)\alpha_1(\tau, \omega_j)\cos(t\omega_j) + \alpha_2(\tau, \omega_j)\sin(t\omega_j)$$

estimable at the Fourier frequencies, $\omega_j = 2\pi j/n$. Alternatively, we can base the analysis on the periodogram of the level crossing process, $Z_t(\tau) = I(X_t < F_{X_n}^{-1}(\tau))$ where $F_{X_n}^{-1}(\tau)$ denotes the τth unconditional sample quantile of the observed X_t's. Both approaches allow researchers to explore frequency domain features of time series by focusing attention on local

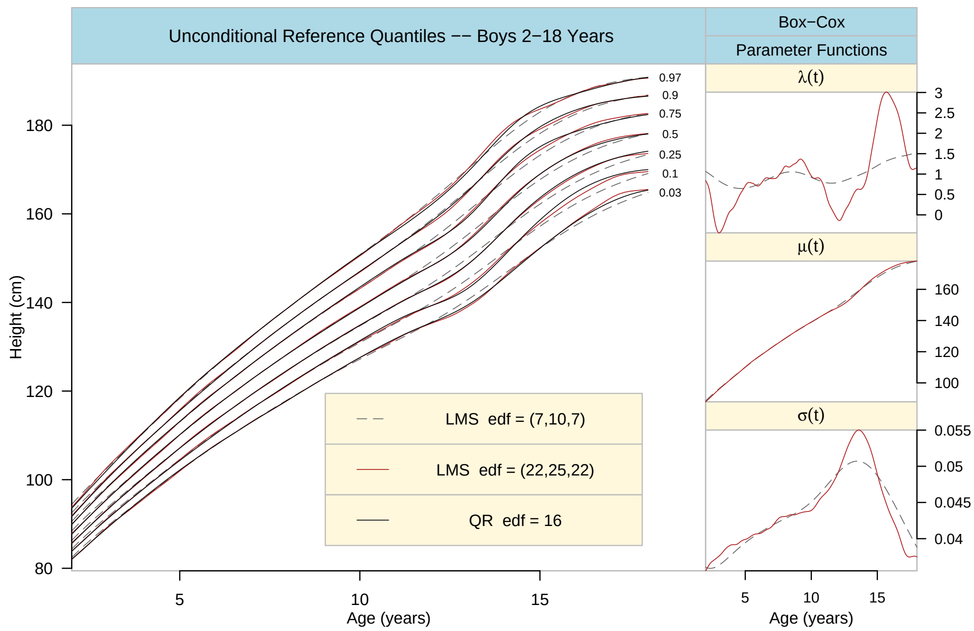

Figure 2

Comparison of LMS and QR Growth Curves: The figure illustrates three families of growth curves, two estimated with the LMS methods of Cole and Green, and one using quantile regression methods.

behavior at or near specific quantiles, recognizing that cyclic behavior can be quite different in the upper tail of the distribution than in the middle or in the lower tail. Extensions to locally stationary time series and cross-spectral relationships among variables are topics of active current research.

5. Longitudinal Data

Longitudinal, or panel, data poses numerous challenges for anyone contemplating extending the quantile regression paradigm. My first encounter with these challenges involved estimation of reference growth charts for height based on a sample of Finnish children and reported in Wei, Pere, Koenker, and He (2005). Our objective was to develop a practical approach to estimating growth charts based on quantile regression methods and illustrate its use on a reference sample of 2305 Finnish children observed, on average, 20 times between the ages of 0 and 20. At the time the state of the art for estimating such charts was the LMS method of Cole and Green (1992), which assumed that heights at each age, $Y(t)$, could be transformed to normality by the classical Box-Cox power transformation, that is that $(Y(t)^{\lambda(t)} - 1)/\lambda(t) \sim \mathcal{N}(\mu(t), \sigma^2(t))$, or at least approximately so. To impose smoothness on the resulting quantile growth curves penalty functions of the usual cubic smoothing spline type were appended to the Box-Cox log likelihood.

In Figure 2 we illustrate a comparison of the quantile regression estimates based on a

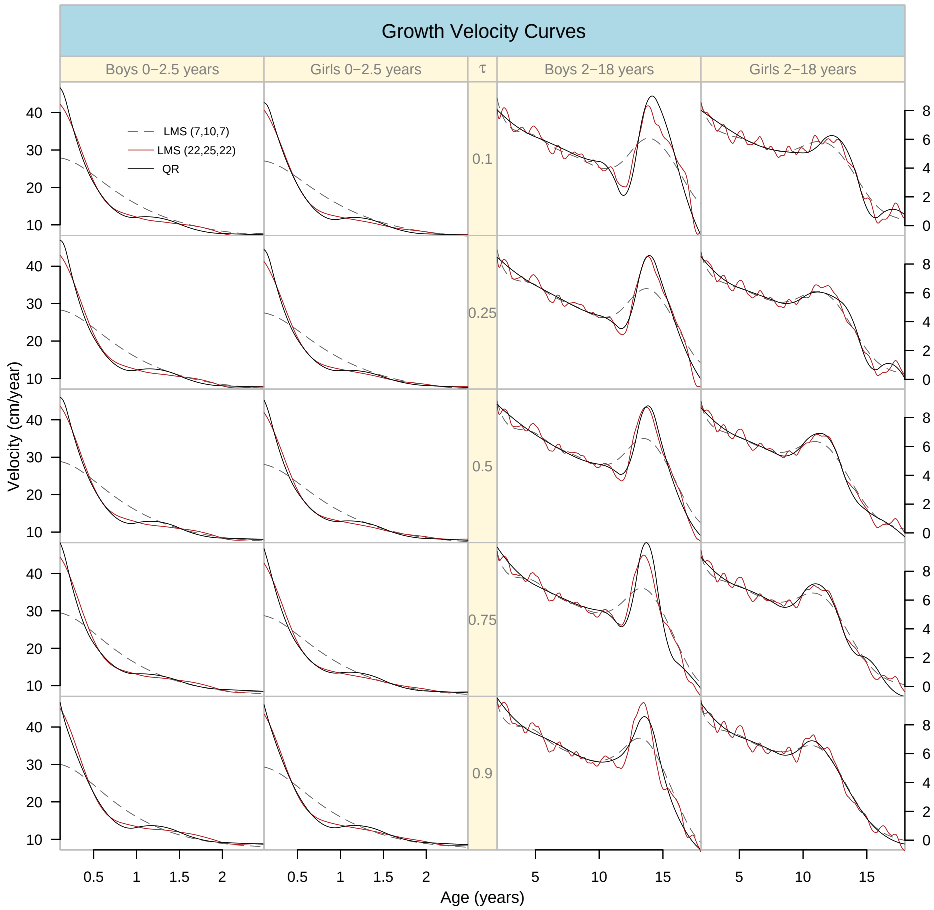

Figure 3

Comparison of LMS and QR Growth Velocity Curves: The figure illustrates three families of growth curves of the prior figure, now representing the estimated velocity of growth.

B-spline basis expansion with 16 knots and two variants of the Cole-Green estimates: one with penalty parameters chosen with the default setting resulting in effective dimension of the $(\lambda(t), \mu(t), \sigma(t))$ functions of $(7, 10, 7)$, the other with a less restrictive choice of the penalty parameters with effective dimension $(22, 25, 22)$. The latter closely mimicked the QR estimates. On the right side of this figure we illustrate the two age dependent paths of the $(\lambda(t), \mu(t), \sigma(t))$ estimates. The default default λ's tend to oversmooth, whereas the more flexible model, while quite nicely matching the QR results yields a rather erratic estimate of the underlying Box-Cox parameter paths. To reenforce this message we illustrate in Figure 3 the corresponding growth velocity curves. Separate estimates were made for infants between the ages of 0 and 2, and older children ages 2-18, as well as distinguishing girls and boys. For the younger children there is generally good agreement between the QR estimates and the more flexible Cole-Green estimates, however for the older children the Cole-Green estimates exhibit substantially more variability than the QR estimates of velocity. As has been already emphasized an advantage of the QR sieve approach is that it makes it relatively easy to condition on additional covariates. In the growth curve setting this was illustrated by incorporating an AR(1) component and midparent's height as additional explanatory variables.

The econometric literature on panel data has focused considerable attention on unobserved individual specific effects. Ever since the influential work of Neyman and Scott (1948) we have struggled to come to terms with these effects. In a rather naive attempt to introduce them into longitudinal settings for quantile regression, Koenker (2004) considered the model,

$$Q_{Y_{it}}(\tau|x_{it}) = \alpha_i + x_{it}^\top \beta(\tau).$$

The proposed estimation strategy was to minimize the penalized QR objective,

$$R(\alpha, \beta) = \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \rho_{\tau_j}(y_{it} - \alpha_i + x_{it}^\top \beta(\tau_j)) + \lambda \|\alpha\|_1.$$

Two features of the estimator were intended to control the ill effects of the α_i's: first, they were assumed to be independent of τ, thus representing a pure location shift of the conditional distribution of the response, and second, their ℓ_1 norm was controlled by the penalty parameter λ. We can interpret this procedure in several ways, none terribly compelling. If we view the α_i's as individual specific fixed effects, then the penalty term is simply a shrinkage scheme: reduced variability of the vector α may help to improve the precision of the estimates of primary interest, β. If instead, we view the α_i's as random, then the Bayesian interpretation of the penalty term suggests that shrinkage could be justified by a Laplace (double exponential) prior on the vector α. Why double exponential? Aside from the obvious computational convenience there seems to be little credible motivation. As we know from the extensive lasso literature the ℓ_1 penalty tends to behave like a hard threshholding rule, shrinking some of the α_i's all the way to zero, while leaving others alone. This may be desirable in some applications, but perhaps not in others.

Subsequent literature has elaborated on this approach as well as introducing a variety of new alternatives. Lamarche (2010) has clarified the role of the shrinkage parameter λ and suggested straegies for choosing it. Galvao (2011) has explored dynamic variants of the model and proposed insturmental variable methods for estimation, and Kato, Galvao, and Montes-Rojas (2012) have substantially improved upon prior results on rate requirements for asymptotic inference.

A considerably more sophisticated approach to quantile regression methods for panel data has been recently proposed by Arellano and Bonhomme (2016). Their approach may be interpreted as an elaboration of the Chamberlain (1984) correlated random effects approach to classical mean regression methods for panel data. Rather than specifying individual specific parametric effects, Arellano and Bonhomme (2016) posit latent variables that enter the model as if they were observable covariates. Estimation proceeds by a variant of the EM algorithm in which the latent covariates are imputed by methods similar to those introduced by Wei and Carroll (2009) for dealing with quantile regression models with measurement error in the covariates. This approach has several advantages over prior panel methods, not the least of which is that there is a clear conditional quantile interpretation, albeit one that conditions on the latent covariates.

6. Duration Models

Quantile regression offers an attractive modeling strategy for duration, or survival, data where interest focuses on restricted regions of the conditional distribution and censoring renders identificaiton of mean effects problematic. My first encounter with these issues was working on Koenker and Geling (2001) where we had the luxury of having a sample of 1.2 million *uncensored* observations of medfly lifetimes. A simple logarithmic accelerated failure time model allowed us to consider extreme tail behavior and covariate effects in a more comprehensive manner than was possible with other standard modeling strategies. A striking example of the latter virtue was the gender cross-over in survival functions for medflies. In most parametric and semiparametric survival models including the Cox proportional hazard model covariates exert a scalar shift effect on hazards or survival probabilities that must be either positive or negative over the entire time scale. Such models cannot accommodate effects like gender cross-over in which a treatment, or a characteristic like gender, has a positive impact on survival at early ages, but then becomes a negative influence at more advanced ages.

Of course it is highly unusual to encounter duration data that doesn't exhibit some form of censoring. As we have already noted, Powell (1986) showed that quantile regression could be adapted to various forms of fixed censoring, thereby relaxing the restrictive conditions imposed by earlier Gaussian likelihood methods. Random censoring of the type typically encountered in biostatistics resisted quantilification until Portnoy (2003) proposed a shrewd recursive scheme that generalized the well known Kaplan-Meier estimator to the regression setting. Observing, as in Efron (1967), that Kaplan-Meier redistributes mass to the right for right censored response, Portnoy proposed a similar procedure in regression leading to a sequence of weighted quantile regression estimates. Somewhat later, Peng and Huang (2008) using a martingale estimating equation formulation like that underlying the Nelson-Aalen estimator, proposed a closely related procedure. Subsequent work has greatly expanded the applicability of these methods, to competing risks, recurrent events, double censoring and other settings. See the recent survey papers of Ying (2017), Peng (2017) and Li and Peng (2017) for further details.

Noting that censoring shares many features with recent work on missing data models, Yang, Narisetty, and He (2016) have recently proposed a clever data augmentation approach that encompasses a wide variety of censored quantile regression models. This approach seems very promising especially for settings like interval censoring that seem otherwise quite intractable.

7. Causal Models and Instrumental Variables

Causal inference, not to be confused with casual empiricism, has been a longstanding focus of econometrics. Indeed it is sometimes claimed that causal modeling is what distinguishes econometric analysis from the "merely descriptive" subject of statistics. This has always seemed to me to be a little self-serving, but there is an element of truth to it. It has become increasingly difficult to maintain this monopoly, as not only the statisticians but also the computer scientists have taken up the banner of causal inference.

Once one abandons the comfortable world of mean effects and linear structural models and embraces the diversity inherent in nonlinear, nonseparable models of distributional effects many open problems present themselves. To my mind the most compelling general approach to these problems is that set out by Chesher (2003). This approach draws on the Rosenblatt (1952) transform and may also be viewed as an extension of the influential, if somewhat controversial, work of Strotz and Wold (1960) on recursive, or triangular, models of linear structural equations.

If X is a scalar random variable with absolutely continuous distribution function F, then $Z = F(X) \sim U[0,1]$, that is Z is uniformly distributed on the unit interval. Rosenblatt noted that this simple idea could be extended to k-variate random variables having absolutely continuous distribution $F(x_1, \cdots, x_k)$ by defining the transformation $z = Tx$ by recursive conditioning,

$$
\begin{aligned}
z_1 &= P\{X_1 \leq x_1\} \equiv F_1(x_1) \\
z_2 &= P\{X_2 \leq x_2 | X_1 = x_1\} \equiv F_2(x_2|x_1) \\
&\vdots \\
z_k &= P\{X_k \leq x_k | X_1 = x_1, \cdots, X_{k-1} = x_{k-1}\} \equiv F_k(x_k|x_1, \cdots, x_{k-1}),
\end{aligned}
$$

so the random vector, $Z = TX$ is uniformly distributed on the unit cube in $\mathbb{R}^k$. This leads us immediately to recursively conditioned quantile functions, $Q_{Y_1|X)}(\tau|x)$, $Q_{Y_2|Y_1,X)}(\tau|y_1, x)$, $\ldots$, $Q_{Y_k|Y_1,\ldots Y_{k-1},X)}(\tau|y_1, \cdots, y_{k-1}, x)$, as an equivalent way to characterize the distribution. Of course there are $k!$ ways of doing this, one for each ordering of the Y's so we require a causal ordering of the response vector. Chesher provides an elegant nonparametric elaboration of this approach with general identification and estimation results. Ma and Koenker (2006) considers more restrictive parametric formulations, and suggests a control variate estimation strategy. Wei (2008) illustrates the approach for estimating bivariate children's growth contours for height and weight where there is a compelling biological argument for the causal precedence of height.

When some components of Y are discrete point identification generally fails, and one must resort to bounds analysis and set valued identification results as demonstrated in Chesher (2005). Since many econometric applications involve discrete endogonous variables alternative approaches framed in terms of instrumental variables have proven to be very influential. The first of these approaches was that of Abadie, Angrist, and Imbens (2002) who considered the very typical case of binary treatment with a binary instrumental variable. This setting is often encountered in experimental settings where treatment is offered to a randomly selected group, but participants cannot be compelled to accept the treatment, so there is voluntary compliance. This is usually described as the "intent to treat" model.

Abadie, Angrist, and Imbens (2002) adopt the potential outcome framework underlying earlier work on local average treatment effects as in Imbens and Angrist (1994). Given a vector of conditioning covariates, X, and a binary treatment indicator D, suppose we have a binary instrumental variable, Z, independent of outcome and treatment status conditional on X. Let D_1 denote (random) treatment status when $Z = 1$, and D_0 when $Z = 0$, and assume that $\mathbb{P}(D_1 \geq D_0|X) = 1$. The objective is to identify the treatment effect on the compliers, that is on the subpopulation with $D_1 > D_0$, those that actually switch to the treatment option when given the opportunity to do so. Compliers are not individually recognizable in the sample, but they are probabilistically recognizable via the following subterfuge. Let

$$\kappa(D, Z, X) = 1 - \frac{D(1-Z)}{1-\pi_0(X)} - \frac{Z(1-D)}{\pi_0(X)}$$

with $\pi_0(X) = \mathbb{P}(Z = 1|X)$. When $D = Z$, $\kappa = 1$, otherwise it is negative. Weighting the usual quantile regression objective function by consistent estimates of the κ's yields a consisten estimator of the quantile treatment effect, $\hat{\alpha}$, solving,

$$\min_{\alpha,\beta} \sum_{i=1}^{n} \hat{\kappa}_i \rho_\tau(y_i - \alpha d_i - x_i^\top \beta).$$

Negative weights are, however, problematic from a computational point of view since they contribute concave summands to what is otherwise a nice, convex objective. So Abadie, Angrist, and Imbens (2002) propose replacing the $\hat{\kappa}_i$'s by estimates of the modified, conditional weights,

$$\kappa_\nu = \mathbb{E}(\kappa|Y, D, X) = 1 - \frac{D(1-\nu_0)}{1-\pi_0(X)} - \frac{\nu_0(1-D)}{\pi_0(X)}$$

where $\nu_0 = \mathbb{E}(Z|Y, D, X) = \mathbb{P}(Z = 1|Y, D, X)$, which is shown under the specified conditions to satisfy $\kappa_\nu = \mathbb{P}(D_1 > D_0|Y, D, X) > 0$. Obviously, there are still significant challenges in estimating the weights, κ_ν, but at least their positivity restores the convexity of the objective function facilitating the minimization step.

In a series of papers Chernozhukov and Hansen (2004, 2005, 2006, 2008) have introduced a broader framework for instrumental variable methods for quantile regression. Their approach has been thoroughly reviewed in Chernozhukov and Hansen (2013), so I will be rather brief here. Again, the potential outcomes formalism is adopted and for each potential outcome, Y_d for $d \in \mathcal{D}$, there is a quantile function $Q_{Y_d}(\tau|d, x) = q(d, x, \tau)$ so $Y_d = q(d, x, U_d)$ with $U_d \sim U[0, 1]$. Conditional on X and for each $d \in \mathcal{D}$, U_d is independent of the instrumental variable, Z. The treatment, D, is determined as $D = \delta(Z, X, V)$ for some random vector, V, and conditional on (X, Z, V), the U_d are iid. This last condition, which Chernozhukov and Hansen (2013) refer to as "rank similarity." may be viewed as a somewhat relaxed version of the rank invariance condition we have described earlier as underlying more naive interpretations of the QTE. Rank similarity does not require that subjects have the same rank under each treatment regime, but only that "the expectation of any function of the rank U_d does not vary across the treatment states." Charnozhukov and Hansen (2013, p 65). This precludes systematic differences in subjects' ranks across treatments as would occur for example if a medical treatment rather miraculously made the most frail patients most robust, and vice versa, or helped only the most frail and most robust leaving intermediate cases unimproved.

Under the foregoing assumptions the moment condition, $\mathbb{P}\{Y \leq q(D, X, \tau)|X, Z\} = \tau$ can be employed to construct an estimator of the structural QTE. In practice such moment conditions, lacking both smoothness and convexity, are rather awkward; fortunately, a more tractable alternative is provided. Consider the linear specification $q(D, X, \tau) = D^\top \alpha(\tau) + X^\top \beta(\tau)$. By the independence, or exclusion, assumption on Z, we may consider estimating the model,

$$Q_{Y - D^\top \alpha|X,Z}(\tau|X, Z) = X^\top \beta(\tau) + Z^\top \gamma(\tau, \alpha),$$

for various α and trying to minimize $\|\gamma(\tau, \alpha)\|_W = \gamma(\tau, \alpha)^\top W \gamma(\tau, \alpha$ with respect to α. Ideally, one would like to choose W to be the inverse of a reasonable estimate of the covariance matrix of $\hat{\gamma}$. It is a useful exercise to show that the mean regression analogue of this procedure is equivalent to classical two stage least squares. Weaker forms of the Chernozhukov and Hansen conditions lead to moment inequality conditions and back toward the theory of Chesher (2005).

8. Errors in Variables, Missing Data and Sample Selection

Going back to the earliest days of econometrics it has been recognized that conventional least squares estimates can be badly biased when covariates are measured with error. Wald (1940) and Durbin (1954) review this early literature and describe instrumental variable methods intended to ameliorate these effects. I will briefly describe some recent developments for treating measurement errors of this type in quantile regression methods.

Wei and Carroll (2009) consider the linear quantile regression model,

$$Q_Y(\tau|x) = x^\top \beta_0(\tau), \tag{1}$$

however, x, is not observed, instead we observe a surrogate, w, that satisfies the condition that $f_Y(y|x,w) = f_Y(y|x)$. This surrogacy condition implies that w is uninformative about Y conditional on X. Since the usual estimating equations,

$$n^{-1} \sum \psi_\tau(y_i - x_i^\top \beta) x_i = 0,$$

are unavailable, we must instead consider the revised equations,

$$n^{-1} \sum \int_x \psi_\tau(y_i - x^\top \beta) x f(x|y_i, w_i) dx = 0,$$

which replaces the x_i by their conditional expectations. We are unlikely to know anything about the conditional density $f(x|y_i, w_i)$ a priori, however under the surrogacy condition,

$$f(x|y_i, w_i) = \frac{f(y_i|x) f(x|w_i)}{\int_x f(y_i|x) f(x|w_i) dx}.$$

Assuming that the linear conditional quantile model holds for all $\tau \in (0,1)$, we can express $f(y|x)$ in terms of the difference quotient,

$$f(y|x) = \lim_{h \to 0} \frac{h}{x^\top (\beta_0(\tau_y + h) - \beta_0(\tau_y))},$$

where $\tau_y = \{\tau \in (0,1) | x^\top \beta(\tau) = y\}$. Estimation proceeds by what Wei and Carroll refer to as a nonparametric analogue of the EM algorithm. There is assumed to be a reliable estimator of $f(x|w)$, perhaps based on replicated observations and a parametric model of the measurement error. Then $\theta \equiv (\beta(\tau_1), \cdots, \beta(\tau_m))$ is initialized using the naive estimator that simply replaces x by its surrogate w, in (1). Weights are then constructed on a grid of x values, and θ is reestimated from the weighted quantile regression objective function at each of the τ_k's, and the process is repeated until convergence is achieved.

Gridding for $\tau \in (0,1)$ isn't worrisome, but the gridding on x may be more so, especially when the dimension of x is large, say bigger than one. What matters of course is the "affected" dimension of x, the number of coordinates subject to measurement error; this is accounted for in $f(x|w)$, which admits the possibility that some coordinates of x are accurately measured.

The approach of Wei and Carroll (2009) reveals an important but somewhat paradoxical feature of the quantile regression paradigm. It is usually emphasized that a cardinal virtue of these methods is that they are local, relying only on data near a particular conditional quantile, and undisturbed by what may be going on elsewhere in the conditional distribution. But when we assert that (1) holds for all $\tau \in (0,1)$, we have taken a leap of faith into the quagmire of global semiparametric models. Although Wei and Carroll (2009) restrict estimation of their original model to a discrete grid of $\tau \in (0,1)$, they explicitly tie these estimates together with the assumption that the coordinates $\beta_j(\tau)$ can be approximated by linear splines with algebraic tail behavior. Indeed, when we compute the weights we require estimates of the global conditional density. Obviously, there are several other strategies that might be employed to produce alternative estimates of such conditional densities. What, if anything, makes quantile regression advantageous for this purpose? I would argue that the main advantages are the linear parameterization, perhaps in some form of basis expansion, and the efficient computation that this facilitates. Given a family of independently estimated conditional quantile functions it is also easy to impose further structure such as smoothness, or particular forms of tail behavior as illustrated in the Wei-Carroll approach. These advantages are also apparent in the recent work of Chernozhukov, Fernndez-Val, and Galichon (2010), Wang, Li, and He (2012) and Arellano and Bonhomme (2016).

Several other approaches to estimation of quantile regression models with errors in variables have been proposed. An analogue of orthogonal least squares regression is considered in He and Liang (2000). Wang, Stefanski, and Zhu (2012) propose a modification of the usual quantile regression objective function adapted to the Gaussian measurement error model. And Schennach (2008) constructs an elegant general approach to deal with nonparametric measurement error employing deconvolution methods.

It is common especially in applications based on survey data to encounter missing covariates. Two general approaches have emerged for dealing with this eventuality: reweighting a la propensity score methods, and multiple imputation. Both approaches have been explored in the quantile regression setting, the former approach requires reliable estimation of a model for missingness, the latter requires a model for the conditional distribution of the missing observations. Both approaches and their combination are discussed in Wei (2017), and the references provided there.

Sample selection is a potentially serious source of bias in many applications as stressed in the seminal work of Heckman (1974) and Gronau (1974). While the parametric approach of Heckman (1979) has been enormously influential, attention as gradually shifted towards models with less stringent conditions. The Frechet bounds analysis of Manski (1993) constitutes an attractive, if pessimistic, option, while the additive control variate approach of Buchinsky (2001) is more pragmatic. In recent work Arellano and Bonhomme (2015) have sought to bridge this gap and proposed an intriguing copula based approach that ultimately relies on a novel modification of the usual quantile regression objective function in which each observation is assigned an individual specific $\hat{\tau}_i$ that depends upon the estimated copula linking the outcome and selection models evaluated at the covariate vector of the selection model. Given these $\hat{\tau}_i$'s, the efficient computational machinery of quantile regression can be employed, but specification and estimation of the copula and the selection model remain serious challenges.

9. Multivariate and Functional Data

It is hardly surprising that the problem of extending the simple idea of estimating conditional quantile functions for a univariate response to multivariate response has proven difficult. Even without any conditioning covariates, it is unclear how one should go about "inverting" a distribution function $F : \mathbb{R}^d \to (0,1)$. Even the notion of the multivariate median is controversial. We have already described one proposal based on the Rosenblatt transform that relies on a causal ordering of the coordinates of the response. This formulation gives us a way, for example, to ask: How do changes in the τ_1 quantile of height impact changes in the τ_2 quantile of weight? But what if we would like to reverse the causal direction? Maybe we should just resist this temptation, but the question is almost irresistible so let's briefly describe some current options.

Suppose for the moment we have no covariates, only our $Y \in \mathbb{R}^d$ response. Let u be a d-vector with $\|u\| = 1$, and denote $Y_u = u^\top Y$ and its orthogonal complement as $Y_u^\perp$. Then,

$$\gamma(\tau, u) = \text{argmin}_{\gamma=(\alpha,\beta)} \mathbb{E}\rho_\tau(Y_u - \alpha - \beta^\top Y_u^\perp)$$

defines a family of hyperplanes indexed by (τ, u). Of course since $\gamma(\tau, u) = \gamma(1-\tau, -u)$ we need only consider $\tau \in (0, 1/2]$. Kong and Mizera (2012) consider contour sets determined by the empirical analogues of these directional conditional quantile hyperplanes and show that the resulting polyhedral contour sets correspond to Tukey halfspace depth contours. This formulation easily accommodates additional linear covariates as shown in Hallin, Paindaveine, and Šiman (2010) who also provide a nice computational refinement based on parametric linear programming. Further extensions to nonparametric formulations are provided in Hallin, Lu, Paindaveine, and Šiman (2015).

In an exciting new development Carlier, Chernozhukov, and Galichon (2016) have proposed a vector quantile regression notion motivated by classical Monge-Kantorovich optimal transport theory. Their approach maintains two essential properties of the univariate conditional quantile functions: namely that the map $(u, x) \mapsto Q_{Y|X}(u|x)$ be monotone in u, and satisfy the representation,

$$Y = Q_{Y|X}(U|X), \quad U|X \sim U(0,1)^d.$$

For $d = 1$ this is all very familiar, but how should we interpret monotonicity in u, for $u \in \mathbb{R}^d$? Carlier, Chernozhukov, and Galichon (2016), employing earlier results of McCann and Brenier, show that there is a unique mapping satisfying these conditions with monotonicity in u interpreted as the requirement that the map $u \mapsto Q_{Y|X}(u|x)$ is the gradient of a convex function, so,

$$(Q_{Y|X}(u|x) - Q_{Y|X}(u'|x))(u - u') \geq 0,$$

for all u and u' in $(0,1)^d$ and $x \in \mathcal{X}$, the support of the conditioning covariates.

As in the univariate setting it is convenient to consider linear parameterizations,

$$Q_{Y|X}(u|x) = \beta_0(u)^\top f(x),$$

so we would have the representation, $Y = \beta_0(U)^\top f(X)$, with $U|X \sim U(0,1)^d$. Here $\beta(u)$ is a p by d matrix of coefficients, and $f(X)$ is a p vector of conditioning covariates, possibly including basis expansion and interaction terms. This formulation leads to a linear programming problem for computing $\beta(u)$ for both population and sample settings.

The inherent ambiguity of the multivariate quantile problem arises from the fact that there are many maps Q such that if $U \sim U(0,1)^d$ then $Q(U|x) \sim F_{Y|X=x}$, any one of which determine a *transport* from the d dimensional uniform distribution function F_U to $F_{Y|X=x}$. Among these choices, the vector quantile regression selection of Carlier, Chernozhukov, and Galichon (2016) is the one that minimizes the Wasserstein distance $\mathbb{E}\|Q(U|X) - U\|^2$. Many interesting questions remain about this appealing formulation. A valuable introduction to optimal transport from a broad economic perspective may be found in the monograph of Galichon (2016).

Functional data analysis has become increasingly important since the appearance of Ramsay and Silverman (1997), but it is received only limited attention in the quantile regression setting, despite the fact that applications to growth curves, pollution concentrations and shape analysis seem to cry out for more flexible methods. A notable exception to this neglect is the work of Kato (2012) who considers function valued covariates and scalar response as an ill-posed inverse problem regularized by truncation of a principal component derived basis expansion. Function valued response poses new challenges closely related to the problems of vector response. A novel recent attack on these issues is described in Choudhury and Chaudhuri (2017).

10. Computational Methods

More data and more complex models have put increased stress on computational resources thoughout statistics and led to many innovations including computational methods for quantile regression. Early simplex based methods of linear programming have gradually given way to interior point methods as described in Portnoy and Koenker (1997), and sparse algebra has further expanded the scope of these methods allowing models with several thousand parameters to be efficiently estimated as illustrated in Koenker (2011). But new demands exceed the capabilities of even these methods and the advent of distributed computing has shifted attention toward gradient descent methods. Koenker (2017) briefly describes one approach to such strategies for quantile regression based on the alternating direction method of multipliers (ADMM) approach of Parikh and Boyd (2014).

I have tried to maintain a comprehensive package of quantile regression software for the R language, R Core Team (2016), acessible on the CRAN network as **quantreg**, Koenker (2016), And other R contributors have extended its capabilities. SAS has now cloned some of the features of the **quantreg** package, including some of the survival analysis methods. Stata has a more limited quantile regression capabilities based on the original simplex algorithm implementation described in Koenker and d'Orey (1987). Implementations of the interior point algorithm described in Portnoy and Koenker (1997) for Matlab and Ox are available from my website.

11. Conclusion

Gaussian models and methods have encouraged the misconception that all things empirical are revealed by conditional means, and perhaps one or two more moments. Quantile regression offers a set of complementary methods designed to explore data features invisible to the inveiglements of least-squares. As data sources become richer and awareness of the importance of heterogeneity increases, quantile regression methods have become more relevant. The scope of quantile regression methods has broadened considerably in recent years, thanks to the efforts of numerous researchers. I hope that this constructive process will continue.

LITERATURE CITED

ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental Variables Estimates of Subsidized Training on the Quantile of Trainee Earnings," *Econometrica*, 70, 91–117.

ARELLANO, M., AND S. BONHOMME (2015): "Quantile Selection Models: With an Application to Understanding Changes in Wage Inequality," preprint.

——— (2016): "Nonlinear Panel Data Estimation via Quantile Regression," *Econometrics Journal*.

ARROW, K., AND M. HOFFENBERG (1959): *A Time Series Analysis of Interindustry Demands*. North-Holland, Amsterdam.

BASSETT, GILBERT W., J., M.-Y. S. TAM, AND K. KNIGHT (2002): "Quantile models and estimators for data analysis," *Metrika*, 55, 17–26.

BELLONI, A., AND V. CHERNOZHUKOV (2011): "ℓ_1-penalized quantile regression in high-dimensional sparse models," *The Annals of Statistics*, 39, 82–130.

BELLONI, A., V. CHERNOZHUKOV, AND K. KATO (2015): "Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems," *Biometrika*, 102, 77–94.

——— (2016): "Valid post-selection inference in High-Dimensional Approximately Sparse Quantile Regression Models," arXiv:1312.7186v4.

BUCHINSKY, M. (2001): "Quantile regression with sample selection: Estimating women's return to education in the U.S.," *Empirical Economics*, 26, 87–113.

CARLIER, G., V. CHERNOZHUKOV, AND A. GALICHON (2016): "Vector quantile regression: An optimal transport approach," *Ann. Statist.*, 44, 1165–1192.

CHAMBERLAIN, G. (1984): "Panel data," in *Handbook of Econometrics*, vol. 2, pp. 1247–1318. Elsevier.

CHAMBERLAIN, G. (1994): "Quantile regression, censoring and the structure of wages," in *Advances in Econometrics*, ed. by C. Sims. Elsevier: New York.

CHAUDHURI, P. (1991): "Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation," *The Annals of Statistics*, 19, 760–777.

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6, Part B, pp. 5549 – 5632. Elsevier.

CHEN, X., R. KOENKER, AND Z. XIAO (2009): "Copula-based nonlinear quantile autoregression," *Econometrics Journal*, 12, S50–S67.

CHERNOZHUKOV, V., I. FERNNDEZ-VAL, AND A. GALICHON (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125.

CHERNOZHUKOV, V., AND C. HANSEN (2005a): "The Effects of 401(k) Participation on the wealth distribution: An Instrumental Quantile Regression Analysis," *Review of Economics and Statistics*, 73, 735–751.

——— (2005b): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–262.

——— (2006): "Instrumental Quantile Regression Inference for structural and treatment effect models," *J. of Econometrics*, 132, 491–525.

——— (2008): "Instrumental Variable Quantile Regression: A Robust Inference Approach," *J. of Econometrics*, 142, 379–398.

——— (2013): "Quantile Models with Endogeneity," *Annual Reviews in Economics*, 5, 57–81.

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

——— (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 1525–1550.

CHOUDHURY, J., AND P. CHAUDHURI (2017): "Quantile Regression with Measurement Error and Missing Data," in *Handbook of Quantile Regression*, ed. by V. Chernozhukov, X. He, R. Koenker, and L. Peng. Chapman-Hall.

Cole, T. J., and P. Green (1992): "Smoothing Reference Centile Curves: The LMS Method and Penalized Likelihood," *Statistics in Medicine*, 11, 1305–1319.

Cox, D. R. (1984): "Interaction," *International Statistical Review*, 52, 1–24.

Diaz, I. (2016): "Efficient estimation of quantiles in missing data models," `arXiv:1512.08110`.

Donoho, D., S. Chen, and M. Saunders (1998): "Atomic decomposition by basis pursuit," *SIAM J. of Scientific Computing*, 20, 33–61.

Durbin, J. (1954): "Errors in Variables," *Review of the International Statistical Institute*, 22, 23–32.

Edgeworth, F. (1888a): "A Mathematical Theory of Banking," *J. Royal Stat Soc*, 51, 113–127.

——— (1888b): "On a new method of reducing observations relating to several quantities," *Philosophical Magazine*, 25, 184–191.

Efron, B. (1967): "The Two Sample Problem with Censored Data," in *Proc. 5th Berkeley Sympt. Math. Statist. Prob.* Prentice-Hall: New York.

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259–276.

Fitzenberger, B., R. Wilke, and X. Zhang (2009): "Implementing BoxCox Quantile Regression," *Econometric Reviews*, 29, 158–181.

Fox, M., and H. Rubin (1964): "Admissibility of Quantile Estimates of a Single Location Parameter," *Annals of Mathematical Statistics*, 35, 1019–1030.

Freedman, D. A. (2008): "On regression adjustments in experiments with several treatments," *Annals Applied Stat.*, 2, 176–196.

Galichon, A. (2016): *Optimal Transport Methods in Economics*. Princeton U. Press.

Galvao, A. F. (2011): "Quantile regression for dynamic panel data with fixed effects," *Journal of Econometrics*, 164, 142–157.

Gronau, R. (1974): "Wage Comparisons–A Selectivity Bias," *Journal of Political Economy*, 82, 1119–1143.

Hagemann, A. (2011): "Robust Spectral Analysis," arXiv:1111.1965.

——— (2016): "Cluster-robust bootstrap inference in quantile regression models," *Journal of the American Statistical Association*.

Hallin, M., Z. Lu, D. Paindaveine, and M. Šiman (2015): "Local bilinear multiple-output quantile/depth regression," *Bernoulli*, 21, 1435–1466.

Hallin, M., D. Paindaveine, and M. Šiman (2010): "Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth," *Ann. Statist.*, 38, 635–669.

He, X. (2017): "Resampling Methods for Quantile Regression," in *Handbook of Quantile Regression*. Chapman-Hall.

He, X., and H. Liang (2000): "Quantile regression estimates for a class of linear and partially linear errors-in-variables models," *Statistica Sinica*, 10(1), 129–140.

Heckman, J. (1974): "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42, 679–694.

Heckman, J. J. (1979): "Sample Selection Bias As a Specification Error," *Econometrica*, 47, 153–161.

Heckman, J. J., J. Smith, and N. Clements (1997): "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *The Review of Economic Studies*, 64, 487–535.

Hotelling, H. (1939): "Tubes and Spheres in n-spaces, and a class of statistical problems," *Am J. Math*, 61, 440–460.

Imbens, G., and J. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Kadane, J. B., and T. Seidenfeld (1996): "Statistical Issues in the Analysis of Data Gathered in the New Designs," in *Bayesian Methods and Ethics in a Clinical Trial Design*, ed. by J. B.

Kadane, pp. 115–125. Wiley.

Kato, K. (2012): "Estimation in functional linear quantile regression," *Ann. Statist.*, 40, 3108–3136.

Kato, K., A. F. Galvao, and G. V. Montes-Rojas (2012): "Asymptotics for panel quantile regression models with individual effects," *Journal of Econometrics*, 170, 76–91.

Khmaladze, E. V. (1981): "Martingale Approach in the Theory of Goodness-of-fit Tests," *Theory of Probability and its Applications*, 26, 240–257.

Kley, T., S. Volgushev, H. Dette, and M. Hallin (2016): "Quantile spectral processes: Asymptotic analysis and inference," *Bernoulli*, 22, 1770–1807.

Knight, K., and G. Bassett (2007): "Second order improvements of sample quantiles using subsamples," Technical Report: Department of Statistics, U. of Toronto.

Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

Koenker, R. (2005): *Quantile Regression*. Cambridge University Press.

——— (2011): "Additive models for quantile regression: Model selection and confidence bandaids," *Brazilian J. Probab. Stat.*, 25, 239–262.

——— (2016): *quantreg: Quantile Regression*R package version 5.27.

Koenker, R. (2017): "Computational Methods for Quantile Regression," in *Handbook of Quantile Regression*. Chapman-Hall.

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

Koenker, R., and V. d'Orey (1987): "Computing Regression Quantiles," *Applied Statistics*, 36, 383–393.

Koenker, R., and O. Geling (2001): "Reappraising Medfly Longevity: A quantile regression survival analysis," *J. of Am. Stat. Assoc.*, 96, 458–468.

Koenker, R., and I. Mizera (2004): "Penalized Triograms: Total Variation Regularization for Bivariate Smoothing," *J. Royal Stat. Society (B)*, 66, 145–163.

Koenker, R., P. Ng, and S. Portnoy (1994): "Quantile Smoothing Splines," *Biometrika*, 81, 673–80.

Koenker, R., and B. Park (1996): "An interior point algorithm for nonlinear quantile regression," *Journal of Econometrics*, 71, 265–283.

Koenker, R., and Z. Xiao (2002): "Inference on the quantile regression process," *Econometrica*, 70, 1583–1612.

——— (2006): "Quantile Autoregression, with Discussion and Rejoinder," *Journal of the American Statistical Association*, 101, 980–1006.

Kong, L., and I. Mizera (2012): "Quantile tomography: using quantiles with multivariate data," *Statistica Sinica*, 22, 1589–1610.

Lamarche, C. (2010): "Robust penalized quantile regression estimation for panel data," *Journal of Econometrics*, 157, 396–408.

Lee, S. (2003): "Efficient semiparametric estimation of a partially linear quantile regression model," *ET*, pp. 1–31.

Lee, Y. K., E. Mammen, and B. U. Park (2010): "Backfitting and smooth backfitting for additive quantile models," *Ann. Statist.*, 38, 2857–2883.

Lehmann, E. (1974): *Nonparametrics: Statistical Methods based on Ranks*. Holden-Day.

Li, R., and L. Peng (2017): "Survival Analysis with Computing Risks," in *Handbook of Quantile Regression*. Chapman-Hall.

Li, T.-H. (2008): "Laplace Periodogram for Time Series Analysis," *Journal of the American Statistical Association*, 103, 757–768.

——— (2012): "Quantile Periodograms," *Journal of the American Statistical Association*, 107, 765–776.

Lipsitz, S., G. Fitzmaurice, G. Molenberghs, and L. Zhao (1997): "Quantile regression methods for longitudinal data with Drop-outs: Application CD4 cell counts of patients infected with

HIV," *Applied Statistics*, pp. 463–476.

Ma, L., and R. Koenker (2006): "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134, 471–506.

Machado, J., and J. Mata (2000): "Box-Cox Quantile regression and the distribution of firm sizes," *Journal of Applied Econometrics*, pp. 253–274.

Manski, C. (1975): "Maximum score estimation of the stochastic utility model of choice," *J. of Eonometrics*, 3, 205–228.

Manski, C. (1993): "The Selection Problem in Econometrics and Statistics," in *Handbook of Statistics*, ed. by C. R. G.S. Maddala, and H. Vinod. Elsevier: New York.

Manski, C. F. (2004): "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 1221–1246.

Matzkin, R. L. (2015): "Estimation of Nonparametric Models With Simultaneity," *Econometrica*, 83, 1–66.

Meyer, M., and M. Woodroofe (2000): "On the degrees of freedom in shape-restricted regression," *Annals of Statistics*, 28, 1083–1104.

Mu, Y., and X. He (2007): "Power Transformation toward a Linear Regression Quantile," *Journal of the American Statistical Association*, 102, 269–279.

Neyman, J., and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

Parikh, N., and S. Boyd (2014): "Proximal Algorithms," *Foundations and Trends in Optimization*, 1, 123–231.

Peng, L. (2017): "Quantile Regression for Survival Analysis," in *Handbook of Quantile Regression*. Chapman-Hall.

Peng, L., and Y. Huang (2008): "Survival Analysis With Quantile Regression Models," *Journal of the American Statistical Association*, 103, 637–649.

Portnoy, S. (2003): "Censored Quantile Regression," *Journal of the American Statistical Association*, 98, 1001–1012.

Portnoy, S., and R. Koenker (1997): "The Gaussian Hare and the Laplacian Tortoise: Computability of squared-error versus absolute-error estimators, with discusssion," *Statistical Science*, 12, 279–300.

Powell, J. L. (1986): "Censored Regression Quantiles," *J. of Econometrics*, 32, 143–155.

R Core Team (2016): *R: A Language and Environment for Statistical Computing*R Foundation for Statistical Computing.

Ramsay, J., and B. Silverman (1997): *Functional Data Analysis*. Springer.

Rosenblatt, M. (1952): "Remarks on a Multivariate Transformation," *Ann. Math. Statist.*, 23, 470–472.

Schennach, S. M. (2008): "Quantile Regression with Mismeasured Covariates," *Econometric Theory*, 24, 1010–1043.

Schmeidler, D. (1986): "Integral representation without additivity," *Proceedings of Am Math Society*, 97, 255–261.

Stone, C. J. (1994): "The Use of Polynomial Splines and Their Tensor Products in Multivariate Function Estimation," *Annals of Statistics*, 22, 118–171.

Strotz, R., and H. Wold (1960): "A Triptych on Causal Systems," *Econometrica*, 28, 417–463.

Student (1931): "The Lanarkshire Milk Experiment," *Biometrika*, 23, 398–406.

Tibshirani, R. (1996): "Regression Shrinkage and Selection Via the Lasso," *J. Royal Stat. Soc. (B)*, 58, 267–288.

Wald, A. (1940): "The fitting of straight lines if both variables are subject to error," *Annals of Mathematical Statistics*.

Wang, H. J., D. Li, and X. He (2012): "Estimation of High Conditional Quantiles for Heavy-Tailed Distributions," *Journal of the American Statistical Association*, 107, 1453–1464.

Wang, H. J., L. A. Stefanski, and Z. Zhu (2012): "Corrected-loss estimation for quantile regression with covariate measurement errors," *Biometrika*, 99, 405–421.

Wang, L., Y. Zhou, R. Song, and B. Sherwood (2016): "Quantile Optimal Treatment Regimes," Technical Report: Department of Statistics, U. of Minnesota.

Wei, Y. (2008): "An Approach to Multivariate Covariate-Dependent Quantile Contours With Application to Bivariate Conditional Growth Charts," *Journal of the American Statistical Association*, 103, 397–409.

——— (2017): "Quantile Regression with Measurement Error and Missing Data," in *Handbook of Quantile Regression*, ed. by V. Chernozhukov, X. He, R. Koenker, and L. Peng. Chapman-Hall.

Wei, Y., and R. J. Carroll (2009): "Quantile Regression With Measurement Error," *Journal of the American Statistical Association*, 104, 1129–1143.

Wei, Y., A. Pere, R. Koenker, and X. He (2005): "Quantile Regression for Reference Growth Charts," *Statistics in Medicine*, 25, 1369–1382.

Yang, X., N. Narisetty, and X. He (2016): "A New Approach to Censored Quantile Regression Estimation," University of Michigan, Department of Statistics Technical Report.

Ying, Z. (2017): "Analysis of Survival Data and Quantile Regression," in *Handbook of Quantile Regression*. Chapman-Hall.