Annual Review of Economics

Quantile Regression: 40 Years On

Roger Koenker

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

Annu. Rev. Econ. 2017. 9:155–76

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-063016-103651

JEL codes: C21, C22, C31, C32, C41

Keywords

quantile regression, treatment effects, heterogeneity, causal inference

Abstract

Since Quetelet's work in the nineteenth 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. However, such effects are often quite heterogeneous: Medical treatments may improve life expectancy but also impose serious short-term risks; reducing class sizes may improve the 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 in this review.

1. INTRODUCTION

Quantiles offer a convenient way to summarize univariate probability distributions, as exemplified by Tukey's ubiquitous box plots. 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 cannot 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

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

to obtain the first-order condition

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

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 $F_n(y) = n^{-1}\sum_{i=1}^n I(Y_i \leq y)$. Fox & Rubin (1964) considered admissibility of the univariate sample quantile under the loss ρ_τ, but the origins 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 eighteenth 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 1950s, when it was recognized as a linear program. An early application of median regression in economics appears in the work of Arrow & Hoffenberg (1959), who found it convenient for estimating input–output coefficients subject to positivity constraints. Although there was a general recognition that the median, also known as the ℓ_1 or least absolute deviations (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-1970s. 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 analog 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, $r_i = y_i - x_i^\top \hat{b},\ i = 1, \ldots, n$, 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, we find that $\partial_b \rho_\tau(y_i - x_i^\top b) = -\psi_\tau(y_i - x_i^\top b)x_i$, 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 that no more than p observations lie on a hyperplane 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 analogs 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 by deriving a combinatorial expression for the finite sample density of $\hat{\beta}(\tau)$ based on the foregoing gradient optimality condition. Because this involved a summation over all $\binom{n}{p}$ elementary subset solutions that corresponded to exact fits of p observations, it did not seem to be terribly practical at first. Eventually, however, 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 & 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 remainder of this review, I attempt 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, I 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 there is 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 quantile treatment effect (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, I have been silent about possible endogoneity or selection issues that might arise regarding treatment assignment; I address such issues in Section 7. Because 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 response $\alpha(\tau) + \beta(\tau)$. This presumption is sometimes referred to as rank invariance, as in, for example, the work of Heckman et al. (1997).

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

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

For this, we do not even need any linear programming machinery because 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. Because the two samples are independent, we find 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 $\Omega^k_{ij} = (\tau_i - \tau_i \wedge \tau_j)/\{f_k[F_k^{-1}(\tau_i)]f_k[F_k^{-1}(\tau_j)]\}$ for $k = 1, 2$ and $i, j : 1, \ldots, p$ and $\lambda_k = n/n_k$, provided that the relative sample sizes n_k/n stay bounded away from 0 and 1 as $n \to \infty$. Of course, this begs the question of how to estimate the matrices Ω^k because they involve the conditional density functions of the two samples. This has spawned a rather extensive literature, and a brief overview is provided by 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 by Koenker (2011). Various other resampling approaches have also recently been suggested, notably by Belloni et al. (2016) and Hagemann (2017). 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 by Koenker & 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 the recent work of Wang et al. (2016), which is based partially on the work of 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 in that of mean regression. Presumably, randomization leaves us with treatment D that is stochastically independent of other covariates, say, X, so further conditioning on X will not help and may hinder the creation of biases 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 & 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 approach is the work of Lipsitz et al. (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 response and program evaluation is the potentially important role of the 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 the main effects of treatment. Interactions, if present, must play an essential role in postanalysis 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 size 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 et al. (2002) and Knight & 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 constitutes 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 by 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,

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

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

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

and the Manski estimator

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

consistently estimate 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, then we can easily estimate the linear parameters $\beta(\tau)$; however, joint estimation of $[\lambda(\tau), \beta(\tau)]$ requires more effort. Machado & 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 et al. (2009) suggest a modification to account for circumstances in which $\lambda x_i^\top \beta + 1 < 0$. Mu & He (2007) propose an alternative estimator in

which λ is estimated by minimizing a sum of squared cumulative sum residuals. Performance of these methods is sensitive to heterogeneity of the conditional density of the response, as would be expected based on the large-sample theory sketched above. An additional option that does not 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 by Koenker & 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 by Lee (2003) and Lee et al. (2010) extends the locally polynomial approach to partially linear and additive models, respectively.

As nonparametric quantile regression models have become more complex, local fitting and backfitting to accommodate new components have become more burdensome and the literature has evolved toward sieve methods. The reader is referred, for example, to 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 to control the parametric dimension of the resulting models. Penalty methods, particularly the ℓ_1 penalty of Donoho et al. (1998) and Tibshirani (1996), have emerged as critical tools for dimension reduction. The lasso is especially convenient in the quantile regression setting because it maintains the linear programming structure of the original problem. This was a primary motivation for the use of total variation roughness penalties by Koenker et al. (1994) and Koenker & Mizera (2004), who impose ℓ_1 penalties 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.

Various proposals have been made for how to choose these penalty parameters, but I think 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 & Woodroofe (2000)

divergence measure of dimension because

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

has summands that take the value 1 when y_i is interpolated and 0 otherwise in the quantile regression setting. However, the proposal of Belloni & 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 postselection 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 et al. (2015, 2016). I believe that the Hotelling tube methods described by Koenker (2011) offer an attractive alternative for some applications. Koenker (2011) provides simulation evidence on the performance of these methods for the construction of uniform confidence bands for univariate nonparametric components, as well as discussion of an application to modeling sources of malnutrition in India.

4. TIME-SERIES MODELS

Econometric time-series 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 & 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 $i = 1, \ldots, q$, we have the familiar independent and identically distributed (i.i.d.) 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, because all its coefficients are driven by the same i.i.d. 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 $\mu_i = \int_0^1 \alpha_i(t)\mathrm{d}t$ and $\omega_i^2 = \int_0^1 \alpha_i^2(t)\mathrm{d}t$ 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 $\alpha_0 = \Phi^{-1}(\tau)$. If we simulate the model by drawing a sequence of i.i.d. 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.

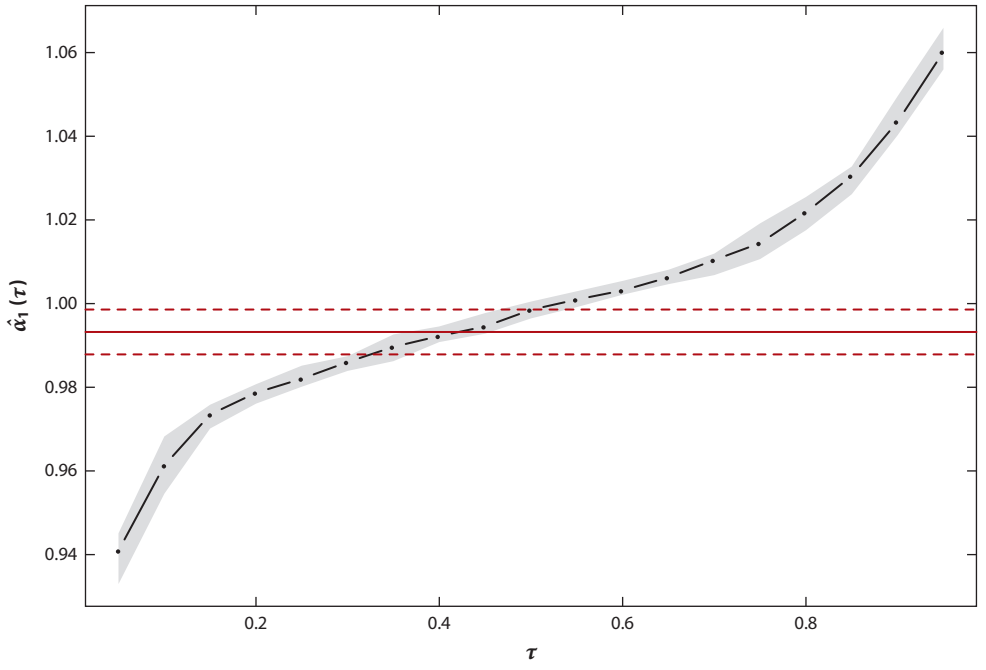

Figure 1

Estimated quantile autoregressive (QAR) $\hat{\alpha}_1(\tau)$ coefficient for 3-month US Treasury bills based on monthly data for the period 1971–2015. The horizontal solid line represents the least squares estimate of 0.99, strongly suggesting unit root behavior.

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

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

where Y_t is the 3-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.

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 subregion 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, Koenker & Xiao 2006, discussion and response). One remedy for this predicament is to resort to nonlinear formulations of the QAR model in lagged Y_t's. Chen et al. (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 by Li (2008, 2012), Hagemann (2011) and Kley et al. (2016) propose 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, $Z_t(\tau) = I(X_t < F^{-1}_{X_n}(\tau))$, where $F^{-1}_{X_n}(\tau)$ 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 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 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 pose 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, as reported by Wei et al. (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 2,305 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 lambda-mu-sigma (LMS) method of Cole & 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. Penalty functions of the usual cubic smoothing spline type were appended to the Box-Cox log likelihood to impose smoothness on the resulting quantile growth curves.

Figure 2 illustrates a comparison of the quantile regression estimates based on a 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 $(7, 10, 7)$, and the other with a less restrictive choice of the penalty parameters with effective dimension $(22, 25, 22)$. The latter closely mimicked the quantile regression estimates. The right side of the figure illustrates the two age-dependent paths of the $[\lambda(t), \mu(t), \sigma(t)]$ estimates. The default λ's tend to oversmooth, whereas the more flexible model, although quite nicely matching the quantile regression results, yields a rather erratic estimate of the underlying Box-Cox parameter paths. To reinforce this message, **Figure 3** illustrates the corresponding growth velocity curves. Separate estimates are made for infants between the ages of 0 and 2 and older children between the ages of 2 and 18; girls and boys are also distinguished. For the younger children, there is generally good agreement between the quantile regression estimates and the more flexible Cole-Green estimates; however, for the older children, the Cole-Green estimates exhibit substantially more variability than the quantile regression estimates of velocity. As emphasized above, an advantage of the quantile regression sieve approach is that it makes it relatively easy to condition on additional covariates. In the growth curve setting, this fact was illustrated by incorporating an AR(1) component and mean parental 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 & 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) considers the model

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

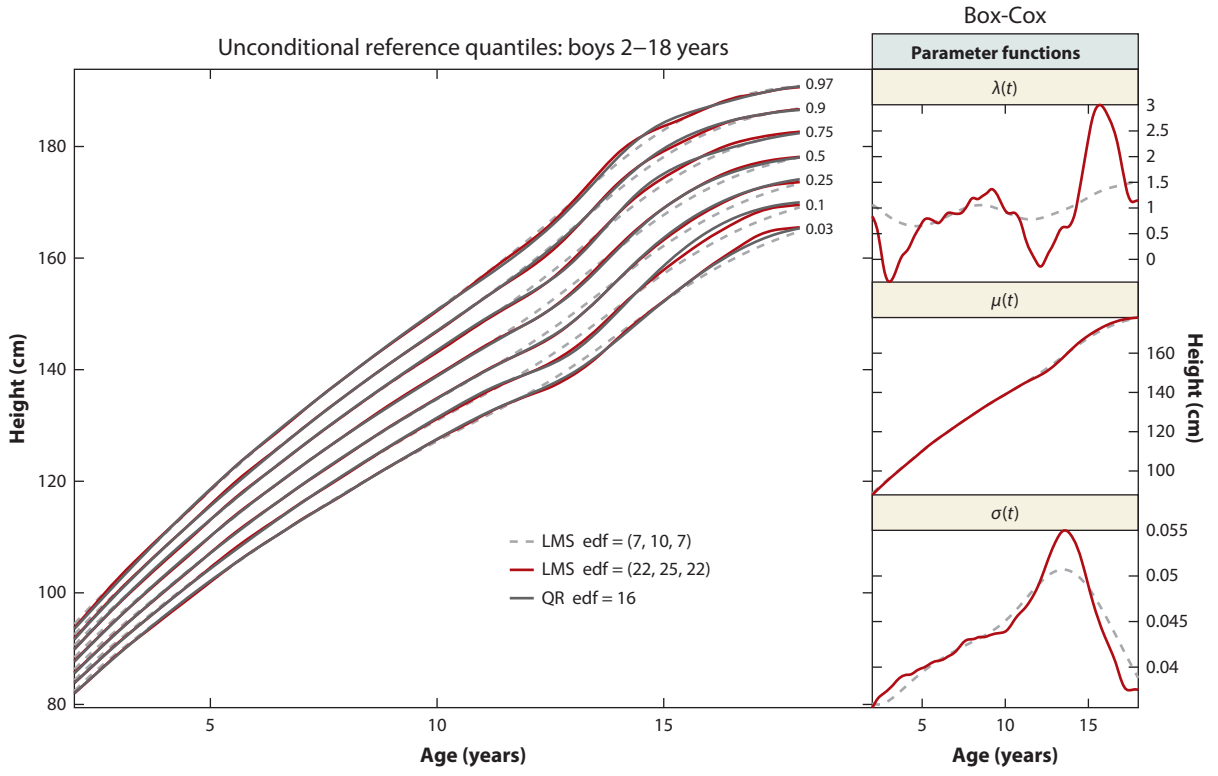

Figure 2

Comparison of LMS and QR growth curves. The figure illustrates three families of growth curves, two estimated with the LMS methods of Cole & Green (1992) and one using QR methods. Abbreviations: edf, effective degrees of freedom; LMS, lambda-mu-sigma; QR, quantile regression.

The proposed estimation strategy was to minimize the penalized quantile regression 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 should the prior be 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 thresholding 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.

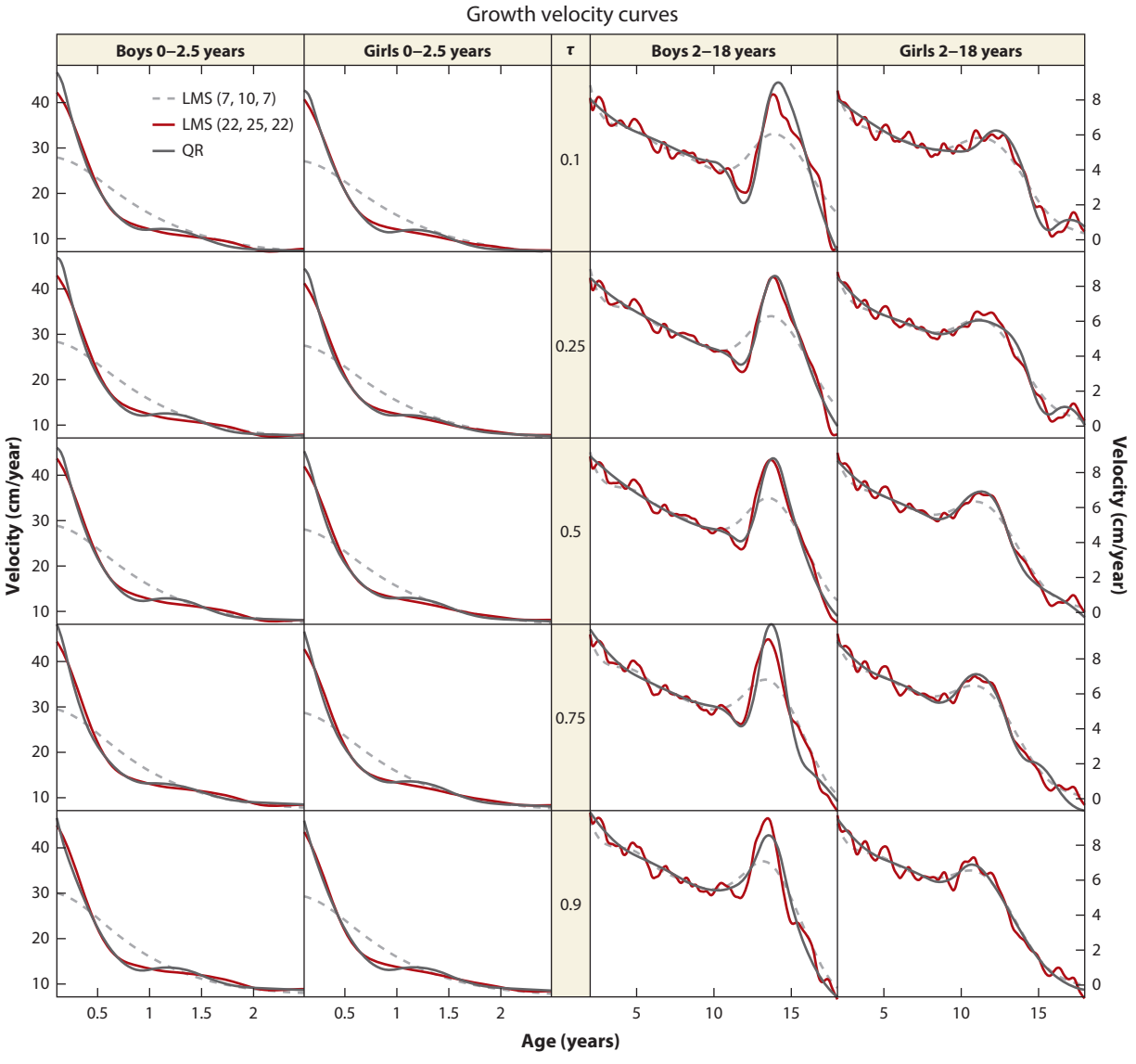

Figure 3

Comparison of LMS and QR growth velocity curves. The figure illustrates three families of growth curves, two estimated with the LMS methods of Cole & Green (1992) and one using QR methods (see **Figure 2**), representing the estimated velocity of growth. Abbreviations: LMS, lambda-mu-sigma; QR, quantile regression.

Subsequent literature has elaborated on this approach as well as introducing a variety of new alternatives. Lamarche (2010) clarifies the role of the shrinkage parameter λ and suggests strategies for choosing it. Galvao (2011) explores dynamic variants of the model and proposes instrumental variable methods for estimation, and Kato et al. (2012) substantially improve 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 & Bonhomme (2017a). 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 & Bonhomme (2017a) posit latent variables that enter the model as if they were observable covariates. Estimation proceeds by a variant of the expectation–maximization (EM) algorithm in which the latent covariates are imputed by methods similar to those introduced by Wei & 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. In my first encounter with these issues (Koenker & Geling 2001), we had the luxury of having a sample of 1.2 million observations of medfly (*Ceratitis capitata*) 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 this virtue was the gender crossover 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 crossover in which a treatment or a characteristic, such as gender, has a positive impact on survival at early ages but becomes a negative influence at more advanced ages.

Of course, it is highly unusual to encounter duration data that do not exhibit some form of censoring. As noted above, Powell (1986) shows that quantile regression can 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 quantification until Portnoy (2003) proposed a shrewd recursive scheme that generalized the well-known Kaplan-Meier estimator to the regression setting. Observing, as does Efron (1967), that the Kaplan-Meier estimator redistributes mass to the right for right censored response, Portnoy proposes a similar procedure in regression, leading to a sequence of weighted quantile regression estimates. Somewhat later, Peng & Huang (2008), using a martingale estimating equation formulation like that underlying the Nelson-Aalen estimator, propose a closely related procedure. Subsequent work has greatly expanded the applicability of these methods to competing risks, recurrent events, double censoring, and other settings (for further details, see Li & Peng 2017, Peng 2017, Ying & Sit 2017).

Noting that censoring shares many features with recent work on missing data models, Yang et al. (2016) propose a clever data augmentation approach that encompasses a wide variety of censored quantile regression models. This approach seems very promising, especially for settings, such as 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 long-standing 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 & 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 (1952) notes that this simple idea could be extended to k-variate random variables having absolutely continuous distribution $F(x_1, \ldots, x_k)$ by defining the transformation $z = Tx$ by recursive conditioning,

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

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, $Q_{Y_1|X}(\tau|x)$, $Q_{Y_2|Y_1,X}(\tau|y_1, x), \ldots,$ $Q_{Y_k|Y_1,\ldots,Y_{k-1},X}(\tau|y_1, \ldots, y_{k-1}, x)$, 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 (2003) provides an elegant nonparametric elaboration of this approach with general identification and estimation results. Ma & Koenker (2006) consider more restrictive parametric formulations and suggest a control variate estimation strategy. Wei (2008) illustrates the approach for estimating bivariate children's growth contours for height and weight when 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 by Chesher (2005). Because many econometric applications involve discrete endogenous variables, alternative approaches framed in terms of instrumental variables have proven to be very influential. The first of these approaches was that of Abadie et al. (2002), who consider 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 et al. (2002) adopt the potential outcome framework underlying earlier work on local average treatment effects, as in the work of Imbens & 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 $\mathbb{P}(D_1 \ge D_0|X) = 1$. The objective is to identify the treatment effect on the compliers, that is, on the subpopulation with $D_1 > D_0$ 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 $D = Z$, we obtain $\kappa = 1$; otherwise κ is negative. Weighting the usual quantile regression objective function by consistent estimates of the κ's yields a consistent

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 because they contribute concave summands to what is otherwise a nice, convex objective. Thus, Abadie et al. (2002) propose replacing the $\hat{\kappa}_i$'s by estimates of the modified, conditional weights,

$$\kappa_v = \mathbb{E}(\kappa | Y, D, X) = 1 - \frac{D(1-v_0)}{1-\pi_0(X)} - \frac{v_0(1-D)}{\pi_0(X)},$$

where $v_0 = \mathbb{E}(Z|Y, D, X) = \mathbb{P}(Z = 1|Y, D, X)$, which is shown under the specified conditions to satisfy $\kappa_v = \mathbb{P}(D_1 > D_0|Y, D, X) > 0$. Obviously, there are still significant challenges in estimating the weights κ_v, but at least their positivity restores the convexity of the objective function facilitating the minimization step.

In a series of papers, Chernozhukov & Hansen (2004, 2005, 2006, 2008) introduce a broader framework for instrumental variable methods for quantile regression. This approach is thoroughly reviewed by Chernozhukov & 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 i.i.d. This last condition, which Chernozhukov & Hansen (2013) refer to as "rank similarity," may be viewed as a somewhat relaxed version of the rank invariance condition 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" (Chernozhukov & Hansen 2013, p. 65). This precludes systematic differences in subjects' ranks across treatments, which would occur, for example, if a medical treatment rather miraculously made the most frail patients the 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 try 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 analog of this procedure is equivalent to classical two-stage least squares. Weaker forms of the Chernozhukov-Hansen conditions lead to moment inequality conditions, as in the theory of Chesher (2005).

8. ERRORS IN VARIABLES, MISSING DATA, AND SAMPLE SELECTION

Since 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. In this section, I briefly describe some recent developments for treating measurement errors of this type in quantile regression methods.

Wei & Carroll (2009) consider the linear quantile regression model

$$Q_Y(\tau|x) = x^\top \beta_0(\tau); \qquad 1.$$

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. Because the usual estimating equation

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

is unavailable, we must instead consider the revised equation

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

which replaces the x_i by its 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, we find that

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

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 & Carroll (2009) refer to as a nonparametric analog 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 $\theta \equiv [\beta(\tau_1), \ldots, \beta(\tau_m)]$ is initialized using the naive estimator, which simply replaces x by its surrogate w, in Equation 1. Weights are then constructed on a grid of x values, 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 $\tau \in (0,1)$ is not worrisome, but the gridding on x may be more so, especially when the dimension of x is large, say, larger 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 & Carroll (2009) reveals an important but somewhat paradoxical feature of the quantile regression paradigm. 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. However, when we assert that Equation 1 holds for all $\tau \in (0,1)$, we have taken a leap of faith into the quagmire of global semiparametric models. Although Wei & 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 of quantile regression 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 et al. (2010), Wang et al. (2012), and Arellano & Bonhomme (2017a).

Several other approaches to estimation of quantile regression models with errors in variables have been proposed. An analog of orthogonal least squares regression is considered by He & Liang (2000). Wang et al. (2012) propose a modification of the usual quantile regression objective function adapted to the Gaussian measurement error model. 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 à 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, and the latter requires a model for the conditional distribution of the missing observations. Both approaches and their combination are discussed by Wei (2017) and the references provided therein.

Sample selection is a potentially serious source of bias in many applications, as stressed in the seminal work of Heckman (1974) and Gronau (1974). Although the parametric approach of Heckman (1979) has been enormously influential, attention has gradually shifted toward models with less stringent conditions. The Fréchet bounds analysis of Manski (1993) constitutes an attractive, if pessimistic, option, whereas the additive control variate approach of Buchinsky (2001) is more pragmatic. In recent work, Arellano & Bonhomme (2017b) seek to bridge this gap and propose 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 to ask, for example: How do changes in the τ_1 quantile of height impact changes in the τ_2 quantile of weight? However, what if we would like to reverse the causal direction? Maybe we should resist this temptation, but the question is almost irresistible, so I briefly describe some current options.

Suppose for the moment that 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, because $\gamma(\tau, u) = \gamma(1 - \tau, -u)$, we need only consider $\tau \in (0, 1/2]$. Kong & Mizera (2012) consider contour sets determined by the empirical analogs 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 by Hallin et al. (2010), who also provide a

nice computational refinement based on parametric linear programming. Further extensions to nonparametric formulations are provided by Hallin et al. (2015).

In an exciting new development, Carlier et al. (2016) propose 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

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

This is all very familiar for $d = 1$, but how should we interpret monotonicity in u for $u \in \mathbb{R}^d$? Carlier et al. (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 that

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

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

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

thus, 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 these maps determines 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 et al. (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 & Silverman's (1997) work, but it has 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 by Chowdhury & Chaudhuri (2017).

10. COMPUTATIONAL METHODS

More data and more complex models have put increased stress on computational resources throughout statistics and have 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 by Portnoy & 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 by Koenker (2011). However, 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 & Boyd (2014).

I have tried to maintain a comprehensive package of quantile regression software for the R language (R Development Core Team 2016), accessible 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 more limited quantile regression capabilities based on the original simplex algorithm implementation described by Koenker & d'Orey (1987). Implementations of the interior point algorithm described by Portnoy & Koenker (1997) for Matlab and Ox are available from my website (**http://www.econ.uiuc.edu/~roger/**).

11. CONCLUSION

Gaussian models and methods have encouraged the misconception that all things empirical are revealed by conditional means and perhaps by 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.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The author would like to thank his collaborators and the many independent scholars who have contributed to the development of quantile regression methods, and to express his appreciation to the National Science Foundation, whose largesse has supported his research over an extended period.

LITERATURE CITED

Abadie A, Angrist J, Imbens G. 2002. Instrumental variables estimates of subsidized training on the quantile of trainee earnings. *Econometrica* 70:91–117

Arellano M, Bonhomme S. 2017a. Nonlinear panel data estimation via quantile regression. *Econom. J.* 19:C61–94

Arellano M, Bonhomme S. 2017b. Quantile selection models with an application to understanding changes in wage inequality. *Econometrica* 85:1–28

Arrow K, Hoffenberg M. 1959. *A Time Series Analysis of Interindustry Demands*. Amsterdam: North-Holland

Bassett GWJ, Tam MYS, Knight K. 2002. Quantile models and estimators for data analysis. *Metrika* 55:17–26

Belloni A, Chernozhukov V. 2011. ℓ_1-penalized quantile regression in high-dimensional sparse models. *Ann. Stat.* 39:82–130

Belloni A, Chernozhukov V, Kato K. 2015. Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* 102:77–94

Belloni A, Chernozhukov V, Kato K. 2016. Valid post-selection inference in high-dimensional approximately sparse quantile regression models. arXiv:1312.7186v4 [math.ST]

Buchinsky M. 2001. Quantile regression with sample selection: estimating women's return to education in the U.S. *Empir. Econ.* 26:87–113
Carlier G, Chernozhukov V, Galichon A. 2016. Vector quantile regression: an optimal transport approach. *Ann. Stat.* 44:1165–92
Chamberlain G. 1984. Panel data. In *Handbook of Econometrics*, Vol. 2, ed. Z Griliches, MD Intriligator, pp. 1247–318. Amsterdam: Elsevier
Chamberlain G. 1994. Quantile regression, censoring, and the structure of wages. In *Advances in Econometrics*, ed. C Sims, pp. 171–208. Cambridge, UK: Cambridge Univ. Press
Chaudhuri P. 1991. Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Stat.* 19:760–77
Chen X. 2007. Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics*, Vol. 6, ed. JJ Heckman, EE Leamer, pp. 5549–632. Amsterdam: Elsevier
Chen X, Koenker R, Xiao Z. 2009. Copula-based nonlinear quantile autoregression. *Econom. J.* 12:S50–67
Chernozhukov V, Fernández-Val I, Galichon A. 2010. Quantile and probability curves without crossing. *Econometrica* 78:1093–125
Chernozhukov V, Hansen C. 2005a. The effects of 401(k) participation on the wealth distribution: an instrumental quantile regression analysis. *Rev. Econ. Stat.* 73:735–51
Chernozhukov V, Hansen C. 2005b. An IV model of quantile treatment effects. *Econometrica* 73:245–62
Chernozhukov V, Hansen C. 2006. Instrumental quantile regression inference for structural and treatment effect models. *J. Econom.* 132:491–525
Chernozhukov V, Hansen C. 2008. Instrumental variable quantile regression: a robust inference approach. *J. Econom.* 142:379–98
Chernozhukov V, Hansen C. 2013. Quantile models with endogeneity. *Annu. Rev. Econ.* 5:57–81
Chesher A. 2003. Identification in nonseparable models. *Econometrica* 71:1405–41
Chesher A. 2005. Nonparametric identification under discrete variation. *Econometrica* 73:1525–50
Chowdhury J, Chowdhury P. 2017. Nonparametric quantile regression for Banach-valued response. In *Handbook of Quantile Regression*, ed. R Koenker, V Chernozhukov, X He, L Peng. Boca Raton, FL: CRC Press. In press
Cole TJ, Green P. 1992. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat. Med.* 11:1305–19
Cox DR. 1984. Interaction. *Int. Stat. Rev.* 52:1–24
Diaz I. 2016. Efficient estimation of quantiles in missing data models. arXiv:1512.08110 [stat.ME]
Donoho D, Chen S, Saunders M. 1998. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* 20:33–61
Durbin J. 1954. Errors in variables. *Rev. Int. Stat. Inst.* 22:23–32
Edgeworth F. 1888a. A mathematical theory of banking. *J. R. Stat. Soc.* 51:113–27
Edgeworth F. 1888b. On a new method of reducing observations relating to several quantities. *Philos. Mag.* 25:184–91
Efron B. 1967. The two sample problem with censored data. *Proc. Berkeley Symp. Math. Statist. Prob., 5th, Berkeley, CA*, Vol. 4, pp. 831–53. New York: Prentice-Hall
Firpo S. 2007. Efficient semiparametric estimation of quantile treatment effects. *Econometrica* 75:259–76
Fitzenberger B, Wilke R, Zhang X. 2009. Implementing Box-Cox quantile regression. *Econom. Rev.* 29:158–81
Fox M, Rubin H. 1964. Admissibility of quantile estimates of a single location parameter. *Ann. Math. Stat.* 35:1019–30
Freedman DA. 2008. On regression adjustments in experiments with several treatments. *Ann. Appl. Stat.* 2:176–96
Galichon A. 2016. *Optimal Transport Methods in Economics*. Princeton, NJ: Princeton Univ. Press
Galvao AF. 2011. Quantile regression for dynamic panel data with fixed effects. *J. Econom.* 164:142–57
Gronau R. 1974. Wage comparisons—a selectivity bias. *J. Polit. Econ.* 82:1119–43
Hagemann A. 2011. Robust spectral analysis. arXiv:1111.1965 [math.ST]
Hagemann A. 2017. Cluster-robust bootstrap inference in quantile regression models. *J. Am. Stat. Assoc.* 112:446–56
Hallin M, Lu Z, Paindaveine D, Šiman M. 2015. Local bilinear multiple-output quantile/depth regression. *Bernoulli* 21:1435–66

Hallin M, Paindaveine D, Šiman M. 2010. Multivariate quantiles and multiple-output regression quantiles: from l_1 optimization to halfspace depth. *Ann. Stat.* 38:635–69
He X. 2017. Resampling methods. In *Handbook of Quantile Regression*, ed. R Koenker, V Chernozhukov, X He, L Peng. Boca Raton, FL: CRC Press. In press
He X, Liang H. 2000. Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Stat. Sin.* 10:129–40
Heckman JJ. 1974. Shadow prices, market wages, and labor supply. *Econometrica* 42:679–94
Heckman JJ. 1979. Sample selection bias as a specification error. *Econometrica* 47:153–61
Heckman JJ, Smith J, Clements N. 1997. Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. *Rev. Econ. Stud.* 64:487–535
Hotelling H. 1939. Tubes and spheres in *n*-spaces, and a class of statistical problems. *Am. J. Math.* 61:440–60
Imbens G, Angrist J. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62:467–75
Kadane JB, Seidenfeld T. 1996. Statistical issues in the analysis of data gathered in the new designs. In *Bayesian Methods and Ethics in a Clinical Trial Design*, ed. JB Kadane, pp. 115–25. Hoboken, NJ: Wiley
Kato K. 2012. Estimation in functional linear quantile regression. *Ann. Stat.* 40:3108–36
Kato K, Galvao AF, Montes-Rojas GV. 2012. Asymptotics for panel quantile regression models with individual effects. *J. Econom.* 170:76–91
Khmaladze EV. 1981. Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* 26:240–57
Kley T, Volgushev S, Dette H, Hallin M. 2016. Quantile spectral processes: asymptotic analysis and inference. *Bernoulli* 22:1770–807
Knight K, Bassett G. 2007. *Second order improvements of sample quantiles using subsamples*. Work. Pap., Dep. Stat., Univ. Toronto, Toronto, Can.
Koenker R. 2004. Quantile regression for longitudinal data. *J. Multivar. Anal.* 91:74–89
Koenker R. 2005. *Quantile Regression*. Cambridge, UK: Cambridge Univ. Press
Koenker R. 2011. Additive models for quantile regression: model selection and confidence bandaids. *Braz. J. Probab. Stat.* 25:239–62
Koenker R. 2016. Quantreg: quantile regression. *R Package Software*, *Version 5.27*. **https://cran.r-project.org**
Koenker R. 2017. Computational methods for quantile regression. In *Handbook of Quantile Regression*, ed. R Koenker, V Chernozhukov, X He, L Peng. Boca Raton, FL: CRC Press. In press
Koenker R, Bassett G. 1978. Regression quantiles. *Econometrica* 46:33–50
Koenker R, d'Orey V. 1987. Computing regression quantiles. *Appl. Stat.* 36:383–93
Koenker R, Geling O. 2001. Reappraising medfly longevity: a quantile regression survival analysis. *J. Am. Stat. Assoc.* 96:458–68
Koenker R, Mizera I. 2004. Penalized triograms: total variation regularization for bivariate smoothing. *J. R. Stat. Soc. B* 66:145–63
Koenker R, Ng P, Portnoy S. 1994. Quantile smoothing splines. *Biometrika* 81:673–80
Koenker R, Park B. 1996. An interior point algorithm for nonlinear quantile regression. *J. Econom.* 71:265–83
Koenker R, Xiao Z. 2002. Inference on the quantile regression process. *Econometrica* 70:1583–612
Koenker R, Xiao Z. 2006. Quantile autoregression, with discussion and rejoinder. *J. Am. Stat. Assoc.* 101:980–1006
Kong L, Mizera I. 2012. Quantile tomography: using quantiles with multivariate data. *Stat. Sin.* 22:1589–610
Lamarche C. 2010. Robust penalized quantile regression estimation for panel data. *J. Econom.* 157:396–408
Lee S. 2003. Efficient semiparametric estimation of a partially linear quantile regression model. *Econom. Theory* 19:1–31
Lee YK, Mammen E, Park BU. 2010. Backfitting and smooth backfitting for additive quantile models. *Ann. Stat.* 38:2857–83
Lehmann E. 1974. *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day
Li R, Peng L. 2017. Survival analysis with competing risks and semi-competing risks data. In *Handbook of Quantile Regression*, ed. R Koenker, V Chernozhukov, X He, L Peng. Boca Raton, FL: CRC Press. In press

Li TH. 2008. Laplace periodogram for time series analysis. *J. Am. Stat. Assoc.* 103:757–68
Li TH. 2012. Quantile periodograms. *J. Am. Stat. Assoc.* 107:765–76
Lipsitz S, Fitzmaurice G, Molenberghs G, Zhao L. 1997. Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with human immunodeficiency virus. *Appl. Stat.* 46:463–76
Ma L, Koenker R. 2006. Quantile regression methods for recursive structural equation models. *J. Econom.* 134:471–506
Machado J, Mata J. 2000. Box-Cox quantile regression and the distribution of firm sizes. *J. Appl. Econom.* 15:253–74
Manski C. 1975. Maximum score estimation of the stochastic utility model of choice. *J. Econom.* 3:205–28
Manski C. 1993. The selection problem in econometrics and statistics. In *Handbook of Statistics*, Vol. 11, ed. GS Maddala, CR Rao, H Vinod, pp. 73–84. Amsterdam: Elsevier
Manski CF. 2004. Statistical treatment rules for heterogeneous populations. *Econometrica* 72:1221–46
Matzkin RL. 2015. Estimation of nonparametric models with simultaneity. *Econometrica* 83:1–66
Meyer M, Woodroofe M. 2000. On the degrees of freedom in shape-restricted regression. *Ann. Stat.* 28:1083–104
Mu Y, He X. 2007. Power transformation toward a linear regression quantile. *J. Am. Stat. Assoc.* 102:269–79
Neyman J, Scott EL. 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16:1–32
Parikh N, Boyd S. 2014. Proximal algorithms. *Found. Trends Optim.* 1:123–231
Peng L. 2017. Quantile regression for survival analysis. In *Handbook of Quantile Regression*, ed. R Koenker, V Chernozhukov, X He, L Peng. Boca Raton, FL: CRC Press. In press
Peng L, Huang Y. 2008. Survival analysis with quantile regression models. *J. Am. Stat. Assoc.* 103:637–49
Portnoy S. 2003. Censored quantile regression. *J. Am. Stat. Assoc.* 98:1001–12
Portnoy S, Koenker R. 1997. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Stat. Sci.* 12:279–300
Powell JL. 1986. Censored regression quantiles. *J. Econom.* 32:143–55
R Development Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna: R Found. Stat. Comput.
Ramsay J, Silverman B. 1997. *Functional Data Analysis*. Berlin: Springer
Rosenblatt M. 1952. Remarks on a multivariate transformation. *Ann. Math. Stat.* 23:470–72
Schennach SM. 2008. Quantile regression with mismeasured covariates. *Econom. Theory* 24:1010–43
Schmeidler D. 1986. Integral representation without additivity. *Proc. Am. Math. Soc.* 97:255–61
Stone CJ. 1994. The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Stat.* 22:118–71
Strotz R, Wold H. 1960. A triptych on causal systems. *Econometrica* 28:417–63
Student. 1931. The Lanarkshire milk experiment. *Biometrika* 23:398–406
Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* 58:267–88
Wald A. 1940. The fitting of straight lines if both variables are subject to error. *Ann. Math. Stat.* 11:284–300
Wang HJ, Li D, He X. 2012. Estimation of high conditional quantiles for heavy-tailed distributions. *J. Am. Stat. Assoc.* 107:1453–64
Wang HJ, Stefanski LA, Zhu Z. 2012. Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika* 99:405–21
Wang L, Zhou Y, Song R, Sherwood B. 2016. *Quantile optimal treatment regimes*. Work. Pap., Dep. Stat., Univ. Minn, Minneapolis, MN
Wei Y. 2008. An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *J. Am. Stat. Assoc.* 103:397–409
Wei Y. 2017. Quantile regression with measurement errors and missing data. In *Handbook of Quantile Regression*, ed. R Koenker, V Chernozhukov, X He, L Peng. Boca Raton, FL: CRC Press. In press
Wei Y, Carroll RJ. 2009. Quantile regression with measurement error. *J. Am. Stat. Assoc.* 104:1129–43
Wei Y, Pere A, Koenker R, He X. 2005. Quantile regression for reference growth charts. *Stat. Med.* 25:1369–82
Yang X, Narisetty N, He X. 2016. *A new approach to censored quantile regression estimation*. Work. Pap., Dep. Stat., Univ. Mich., Ann Arbor, MI
Ying Z, Sit T. 2017. Survival analysis: a quantile perspective. In *Handbook of Quantile Regression*, ed. R Koenker, V Chernozhukov, X He, L Peng. Boca Raton, FL: CRC Press. In press

Annual Review of Economics

Volume 9, 2017

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Economics* articles may be found at http://www.annualreviews.org/errata/economics