Ad Astra 4, 2005
www.ad-astra.ro/journal

Ad Astra

# Quantile Regression: Overview and Selected Applications [1]

I. Sebastian Buhai
Tinbergen Institute and Erasmus University Rotterdam, Netherlands
buhai@tinbergen.nl

## 1 Introduction

This report provides a short and generally accessible overview of the technique of quantile regression, with focus on introducing the method and discussing some major applications, rather than exclusively devoting space to either a technical summary of the theory or to a complete survey of recent advances in implementation, plenty of specialized literature having achieved that by now. To the declared aim, two applications of quantile regression to survival analysis and respectively to recursive structural equations models, with supporting empirical implementations from the literature, will be selected for analysis after the basic quantile regression model and its critical features would have been briefly reviewed.

Quantile regression has lately received much attention, both from a theoretical and from an empirical viewpoint. Defined in the simplest way, quantile regression is a statistical procedure intended to estimate conditional quantile functions. In analogy with classical linear regression methods, based on minimizing sums of squared residuals and meant to estimate models for conditional mean functions, quantile regression methods are based on minimizing asymmetrically weighted absolute residuals and intended to estimate conditional median functions[2] and a full range of other conditional quantile functions. The basic motivation for using quantiles rather than simple mean regression is that the stochastic relationship between random variables can be portrayed much better and with much more accuracy. As extensively discussed in Koenker and Bassett(1978b), conventional least squares estimators may be seriously deficient in linear models constructed on some non-Gaussian settings, where quantile regression would provide more robust and consequently more efficient estimates. The price paid in terms of loss of efficiency (to the least squares estimators) at the normal distribution would moreover not outweigh the gain at the non-Gaussian spreads. Thus quantile regression methods complement and improve established means regression models.

[1]This overview was written in February 2004. Roger Koenker's lecture notes from the Netherlands Network of Economics Workshop in Groningen, December 2003, provided most of the references used in here.

[2]A particular instance of quantile regression, the median regression minimizes sums of absolute residuals and is estimated by the least absolute deviations (LAD) estimator, discussed also in standard econometrics books like Greene (1993) or Wooldridge (2002). The median regression has been analyzed already extensively in early papers such as Huber(1967) or Koenker and Bassett (1978a).

To give a flavour of the possible applications of quantile regression methods in economics, I will enumerate some of the subfields rich in empirical implementations. For an excellent and almost exhaustive treatment of the recent quantile regression applications I suggest however the collection of studies in Fitzenberger, Koenker and Machada, eds (2002) or the more concise presentation in Koenker and Hallock (2001). In economics there seems to be a rapidly expanding empirical quantile regression trend that tries to make a case for the value of "going beyond models for the conditional mean". Thus, quantile regression has been widely employed for instance within labour or educational economics to study wage determinants, discrimination effects, transition or duration data, trends in income inequality or effects of socioeconomic characteristics and policy variables on educational attainment. Quantile regression methods have also been used lately in micro-demand analysis and there even seems to be a growing literature using quantile regression in empirical finance and particularly, on value at risk. In general quantile regression proves to be extremely useful whenever one is interested in focusing on particular segments of the analyzed conditional distribution, or for instance on upper or lower quantile reference curves as a function of several covariates of interest, and this without having to impose any sort of strict parametric assumptions.

I will continue by presenting the basics and some straightforward properties of the quantile regression model in the second section of this report, where the fundamental framework, interpretation and robustness of the estimation, equivariance properties and the quantile treatment effect, next to computational aspects, will all be briefly examined[3]. The implementation of the technique in the selected applications of survival analysis and respectively recursive structural models, is left for the third part of this paper[4], while the last section sums up the report and concludes.

## 2 Fundamentals and Features of Quantile Regression

### 2.1 Basic Model and Interpretation

The quantile regression classic model has been introduced by Koenker and Bassett (1978b) as an extension from the notion of ordinary quantiles (also called "percentiles") in a location model, to a more general class of linear models in which the conditional quantiles have a linear form. To briefly recall the ordinary quantile, consider a real valued random variable $Y$ characterized by the following distribution function

$$F(y) = \Pr(Y \leq y) \tag{1}$$

Then for any $\tau \in (0,1)$, the $\tau$-th quantile of $Y$ is defined as follows:

$$Q(\tau) = \inf\{y : F(y) \geq \tau\} \tag{2}$$

[3]Inference and asymptotics in quantile regression are not reviewed in this report, given the sizeable literature that does this very well. I suggest Koenker and Bassett (1978) for the elementary form of the quantile regression asymptotic theory and recommend some survey studies for the alternative ways of inference: survey of rank based tests inference in Koenker (1997), survey of inference based on resampling methods in Buchinsky (1998) and a survey of general goodness of fit measures and related inference methods based on the whole quantile regression progress in Koenker and Machado (1999) , all this as probably sufficient background for a reasonable general perspective on inference in quantile regression.

[4]I will not discuss at all the recently emerging literature on nonparametric quantile regression models and their applications, although this is one of the most fascinating areas in quantile regressions nowadays. For reasons of space and scope, polynomial methods, quantile smoothing splines, penalized triograms and other techniques will thus be overlooked herein. The interested reader is referred to articles such as Chaudhuri (1991), Koenker, Ng and Portnoy (1994) or Koenker and Mizera (2003), and the references therein, for more information on these topics.

The median is then $Q(1/2)$, the first quartile $Q(1/4)$ and the first decile $Q(1/10)$. The quantile function provides a complete characterization of $Y$, just like the distribution function $F$. The quantiles can be written as solutions to the following optimization problem: for any $\tau \in (0,1)$, define the piecewise linear "check function"

$$\rho_\tau(u) = u(\tau - I(u < 0)) \tag{3}$$

where $I(.)$ is the usual indicator function. The solution to the minimization problem is then

$$\widehat{\alpha}(\tau) = \arg\min_{\xi \in \mathbb{R}} E[\rho_\tau(Y - \xi)] \tag{4}$$

The sample analogue of $Q(\tau)$ is based on a random sample $\{y_1,...,y_n\}$ of $Y$. The $\tau$-th quantile can then be identified, in the spirit of (4) above, as any solution to:

$$\widehat{\alpha}_\tau = \arg\min_{\xi \in \mathbb{R}} \sum_{i=1}^{n} \rho_\tau(y_i - \xi) \tag{5}$$

Let $x_i$, $i = \overline{1...n}$, a $K \times 1$ vector of regressors. We can then write the equivalent of expression (1) as:

$$F_{u_\tau}(\tau - x_i^{'}\beta_\tau | x_i) = \Pr(y_i \leq \tau | x_i) \tag{6}$$

which is essentially a different form derived from the more familiar:

$$y_i = x_i'\beta_\tau + u_{\tau_i} \tag{7}$$

where the distribution of the error term $u_{\tau_i}$ is left unspecified, the only constraint being the (usual) quantile restriction $Q_\tau(u_{\tau_i}|x_i) = 0$.

Using as analogy the estimation of conditional mean functions as in

$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^K} \sum_{i=1}^{n} (y_i - x_i'\beta)^2 \tag{8}$$

the linear conditional quantile function

$$Q_Y(\tau | X = x) = x_i'\beta_\tau \tag{9}$$

can be estimated by solving the equivalent of expression (8) for this case:

$$\widehat{\beta}_\tau = \arg\min_{\beta \in \mathbb{R}^K} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta) \tag{10}$$

We have not yet asked the question about the interpretation of quantile regression. A least squares estimator of the mean regression model would be concerned with the dependence of the conditional mean of $Y$ on the covariates $X$. The quantile regression estimator tackles this issue at each quantile of the conditional distribution, providing thus a more complete description of how the conditional distribution of $Y$ given $X = x$ depends on $x$. In other words, instead of assuming that covariates shift only the location or the scale of the conditional distribution, quantile regression looks at the potential effects on the shape of the distribution as well[5]. Let us look also at a more

[5]A straightforward example can be considered for illustration: the effect of an on-the-job training on the length of the participants' current unemployment spells. It might well be that the shortest spells will become longer as a result of the training program, while the very long spells could be dramatically reduced. In such a case the mean treatment effect might not capture any effect (if two opposite effects in different segments of the distribution completely average out) or it might indicate a distorted one (if one effect dominates the other and they are both significant), while clearly the shape of the unemployment durations would be significantly altered.

practical question: what interpretation does one attach to the quantiles' coefficients? As discussed for example in Buchinsky (1998), the partial derivative of the conditional quantile of $y$ (see (9) above) with respect to one of the regressors, say the $j$-th one, is to be read as the marginal change in the $\tau$-th quantile due to the marginal change in the $j$-th element of $x$. If further, as in the hypothesis in this section, $x$ has $K$ distinct variables, then this derivative would simply be the coefficient on the $j$-th variable, $\beta_j$. Caution is required however when interpreting this result: it certainly does not imply that a subject who happens to be in the $\tau$-th quantile of one conditional distribution would still find himself/herself there, had the corresponding value of his/her $x$ changed.

In the introduction to this report I made a reference to the robustness of quantile regression relative to the case of the mean regression estimates. Since robustness to distributional assumptions is in general a crucial consideration throughout statistics, it is compulsory to say a few more words about it, in the context of the quantile regression. There is indeed very good news about the robustness interpretation, in that the estimates and the inference process have an inherent distribution-free character given that quantile estimation is influenced only by the local behavior of the conditional distribution of the response near the specified quantile. The signs of the residuals are the only thing that matters in the determination of the estimates and thus outliers in the values of the response variables influence the fit in so far as their being above or below the fitted hyperplane, but how far below or above is really irrelevant. There is however more than meets the eye and not everything is so positive: while we have just seen that the quantile regression estimates are inherently robust to contamination of the response observations, they can be quite sensitive to contamination of the design observations.

### 2.2 Monotone Transformations Equivariance and Quantile Treatment Effect

Quantile regression has a very important property that sharply distinguishes it from a linear mean regression. This is the property of equivariance to any monotone transformations[6], see e.g. Koenker and Geiling (2001), for a thorough treatment. In few words, considering any monotone function $h(.)$, we will have the following expression holding

$$Q_{h(Y)|X}(\tau|x) = h(Q_{Y|X}(\tau|x)) \tag{11}$$

This follows easily (given the monotonicity of $h$) using the next intermediary step:

$$\Pr(T < t|x) = \Pr(h(T) < h(t)|x) \tag{12}$$

With the equivariance transformation property one can essentially decouple the potentially conflicting objectives of transformations of the response variable, which is obviously not possible when estimating transformation models for conditional mean relationships, where we would have the generally true

$$E(h(Y)|X) \neq h(E(Y|X) \tag{13}$$

The equivariance property has a particularly useful application in settings which deal with censorship of the observed response variable, see for instance Powell (1986)

[6]Particular cases of quantile regression equivariance under monotone transformations are the scale equivariance, regression shift equivariance and reparameterization of design equivariance. The interesed reader can consult e.g. Buchinsky(1998) for more discussion on these specific equivariance properties of the quantile regressions.

in the case of fixed censoring. Powell (1986) made the crucial observation that linear conditional quantile models can solve the fixed censoring issue by a simple nonlinear modification of the response function. The matters are slightly more complicated in the case of random censoring but advances have been registered also in here, e.g. Honore, Powell and Khan (2000). Buchinsky (1998) provides a recent concise survey in the topic of quantile regression applied to solve censorship of the observed response variable..

It is maintained that the simplest formulation of quantile regression is the two-sample treatment-control model. Lehmann (1974) was the first to propose a general model of treatment response instead of the classical experimental design in which the treatment induces a simple location shift of the response distribution. So suppose that the treatment adds by hypothesis an amount $\Delta(x)$ while the response of the untreated subject would be $x$. Then, paraphrasing Lehmann, the distribution $G$ of the treatment responses is that of the random variable $X+\Delta(X)$ where $X$ is distributed according to $F$. Special cases of this simple quantile treatment effect model include the location shift model for which the treatment effect is a constant, $\Delta(X) = \Delta_0$, and the scale shift model where the treatment addition is a first degree polynomial in $X$: $\Delta(X) = \Delta_0 X$.

Following Doksum (1974), if $\Delta(x)$ is defined as the "horizontal distance" between $F$ and $G$ at $x$, so that

$$F(x) = G(x+\Delta(x)) \tag{14}$$

then one can uniquely define $\Delta(x)$ and express it as follows:

$$\Delta(x) = G^{-1}(F(x)) - x \tag{15}$$

Or this is exactly the quantile treatment effect[7]; by performing the variable change $\tau = F(x)$, we can write (15) as follows

$$\delta_\tau = \Delta(F^{-1}(\tau)) = G^{-1}(\tau) - F^{-1}(\tau) \tag{16}$$

Subsequently, in the 2-sample setting that we discuss here, we can estimate $\delta(\tau)$ easily:

$$\widehat{\delta}_\tau = \widehat{G}_n^{-1}(\tau) - \widehat{F}_m^{-1}(\tau) \tag{17}$$

with $G_n$ and $F_m$ denoting the empirical distribution functions of the treatment and control observations, based on $n$ and $m$ observations respectively. Now, if we think of the quantile regression model for the binary treatment problem, we will have

$$Q_{Y_i}(\tau|D_i) = \alpha_\tau + \delta_\tau D_i \tag{18}$$

with $D_i$ denoting the treatment indicator, $D_i = 1$ indicating treatment and $D_i = 0$ indicating control. The quantile treatment effect can then be estimated directly by solving the following optimization problem:

$$(\widehat{\alpha}_\tau, \widehat{\delta}_\tau)' = \arg\min_{(\alpha,\delta)\in\mathbb{R}^2} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha - \delta D_i) \tag{19}$$

[7] This expression in (15) is also the essence of the QQ-plot, which has has quite some background as graphical diagnostic device. For instance, if $F$ and $G$ are the same, then $G^{-1}(F(x)$ will lie along the 45 degree line in the QQ plot. Quantile regression is thus a way of extending the traditional two-sample QQ plot and related methods to general regression settings with continuous covariates.

The solution yields

$$\begin{cases} \widehat{\alpha}_\tau = F_m^{-1}(\tau), \text{ corresponding to the control sample} \\ \widehat{\delta}_\tau = G_n^{-1}(\tau) - F_m^{-1}(\tau), \text{ as claimed above in (17)} \end{cases} \tag{20}$$

I will examine more closely a particular example where quantile treatment effects are related to duration analysis, mainly looking at the paper by Ma and Koenker (2003) in the last section of this report.

### 2.3 Computational Aspects

It would not be too much of an advantage applying the quantile regression technique if its computation were too cumbersome. Fortunately, this is not the case since quantile regression has a convenient linear programming (LP) representation. This fact has important consequences from both theoretical and practical standpoints. I will not make a target from summarizing the bulk of the literature on this topic, but rather follow Buchinsky (1998), who discussed the computational issues in quantile regression concisely and from a non-technical perspective.

Using expressions (7) and (10) in the presentation of the basics of the quantile regression model above, we can write $y_i$ as a function of only positive elements and then translate it into matrix notation so that we emphasize the primal formulation as LP. One can thus write successively:

$$y_i = \sum_{j=1}^{K} x_{ij}\beta_{\tau_j} + u_{\tau_j} = \sum_{j=1}^{K} x_{ij}(\beta^1_{\tau_j} - \beta^2_{\tau_j}) + (\varepsilon_{\tau_i} - \nu_{\tau_i}) \tag{21}$$

with $\beta^1_{\tau_j}$, $\beta^2_{\tau_j}$, $\varepsilon_{\tau_i}$ and $\nu_{\tau_i}$, non-negative ($j = \overline{1,K}$, $i = \overline{1,n}$). The matrix notation for the primal LP problem is then

$$\min_z c'z \text{ st. } Az = y,\, z \geq 0 \tag{22}$$

where $A = (X,\, -X,\, I_n,\, -I_n)$, $z = (\beta^{1'},\, \beta^{2'},\, u',\, v')'$, $c = (0',\, 0',\, \tau l',\, (1-\tau)l')'$; further $I_n$ is the n dimensional identity matrix, $0'$ is a $K \times 1$ vector of zeros and $l$ is an $n \times 1$ vector of ones. The dual side of the LP is easy to expose now, after having obtained (22):

$$\max_w w'y \text{ st. } w'y \leq c' \tag{23}$$

The duality theorem implies that solutions exist for both formulations if $X$ is a full rank matrix. Further, the equilibrium theorem of LP guarantees the optimality of this solution. For a more technical but still accessible perspective on LP in general, the interested reader could consult the extensive discussion on numerical methods in Judd (1999).

As far as the computation algorithms for quantile regression are concerned, in the 1940's it was recognized that the median regression could be formulated as an LP program, and the simplex method has been since the most employed method to solve it. The most popular algorithm remains even today the one by Barrodale and Roberts (1973, 1974), algorithm which has been implemented in most statistical software packages. The mechanism is the following: at each step there is a trial of $p$ initial observations whose exact fit may constitute a solution. The Barrodale-Roberts algorithm computes next the directional derivative in each of the $2p$ directions resulting from removal of one of the current basic observations and takes a positive or a negative step; the algorithm stops when none of these directional derivative is negative, after having always chosen the direction of the most negative, steepest descent, and having gone in that direction until the objective function ceased to decrease. In essence what the algorithm does is find the solution to a weighted quantile problem by making use of a modified simplex strategy. The method performs quite well on reasonably low numbers of observations, where it does achieve speeds comparable to the corresponding least squares solutions. The dilemma appears however for larger number of observations (say $n$ in the order of $100,000$ and more): the simplex tool becomes considerably slow. There are some recent developments of interior point methods for LP, highly effective for such large problems; one of these can be consulted for instance in Portnoy and Koenker (1997). Portnoy and Koenker basically show that a combination of interior point methods and effective preprocessing can render large scale quantile regression computation competitive with least squares problems of the same size.

One more observation to make with regards to the computational aspects of quantile regression is that, as a great advantage of the LP representation, an entire range of solutions can be efficiently computed by purely parametric estimation. Hence, at any solution $\widehat{\beta}_{\tau_0}$, there is an interval of quantiles $\tau$ over which this solution remains optimal; since it is straightforward to compute the endpoints of this interval, one can solve iteratively for the entire sample path $\widehat{\beta}_{\tau}$ with one simplex pivot at each of these endpoints.

# 3 Applications of Quantile Regression

## 3.1 Quantile Regression for Duration Models

There are many potential econometric applications of quantile regression to survival analysis, as this area have proven to be quite productive for the growth of semi-parametric methods. In this sense Chaudhuri, Doksum and Samarov (1997) have argued that quantile regression provides a unifying approach for transformation models more generally, including also a variety of duration models such as proportional hazards, proportional odds, accelerated failure time, etc. A few studies that implement the technique for survival data analysis are for instance Koenker and Geiling (2001), who describe a sort of large scale application in experimental demography and make a link to the Lehmann (1974) quantile treatment effect; then we have studies like Horowitz and Neumann (1987) or Fitzenberger (1997), who implement quantile regression in the analysis of duration of employment spells;.there would probably be many others, I shall not claim that I attempted to exhaust the list. In the remaining of this subsection I will revisit a study in the category of those dealing with unemployment spells, choosing a recent paper by Koenker and Bilias (2001).

Koenker and Bilias argue in their paper that quantile regression can play a constructive role in the analysis of duration spells, offering a more flexible and more complete approach than the existing conventional method. The authors illustrate this by performing a re-examination of the data from the by now famous Pennsylvania Re-employment Bonus Experiments conducted between 1988-89 and intended to test the efficacy of cash bonuses paid for early re-employment in shortening the length of insured unemployment spells. Drawing on previous work in Koenker and Geiling (2001), Koenker and Bilias (2001) stress a general formulation of the experimental treatment effect introduced in Lehmann (1974) or Doksum (1974) that I also overviewed in the previous section. A link between the transformation models investigated under the equivariance properties above and survival analysis is made, which will constitute the essence of the methodology employed in their article.

Doksum and Gasko (1990) start by putting forward a link between duration analysis and the general transformation model

$$h(T_i) = x_i'\beta + u_i \tag{24}$$

Many survival models in parametric or semiparametric models can actually be expressed in the form above: what we have in (24) is translated in some monotone transformation of a survival time $T_i$ represented as a linear predictor plus an *iid* error. No matter whether we consider a Cox proportional hazard model or a Weibull survival model, they can all be written in a transformation model form as above (see Koenker and Bilias (2001) for the extended discussion). The common item in these models is that we assume the error to be *iid*, which in other words means that for some suitable $h(.)$ one can express the transformed survival times $h(T)$ as a pure location shift in the covariates $x$. Or this clearly imposes some very drastic constraints on the relationship between the covariates and the survival distribution. In response to this and as an alternative to the location shift model, Koenker and Bilias (2001) propose a family of linear-in-parameters quantile regression models for the transformed survival time $h(T)$,

$$Q_{h(T)}(\tau|x) = x'\beta_\tau \tag{25}$$

In this model, potentially all the parameters of the $p$-vector $\beta_\tau$ depend now on the specified quantile, $\tau$. In particular, allowing for the slope coefficients of $\beta_\tau$ to depend on $\tau$, we can introduce various forms of heterogeneity in the conditional distribution of $h(T)$ on the covariates.

Before I present the analysis of the experiment and the conclusions drawn by Koenker and Bilias (2001), some specific insights in the data and the experiment per se, are necessary. While I assume that the reader is mode or less familiar with the US Bonus experiments (Meyer (1995) for instance provides some excellent reviews of and some general conclusions about the re-employment bonus experiments in, New Jersey, Illinois and Pennsylvania), I will recall the experimental design by presenting a summary table and some essential remarks:

| *Group* | *Bonus Amount* | *Qualification Period* |
|---|---|---|
| Controls | 0 | 0 |
| Treatment1 | Low$\simeq$ \$500 | Short=6 weeks |
| Treatment2 | Low | Long=12 weeks |
| Treatment3 | High$\simeq$ \$1000 | Short |
| Treatment4 | High | Long |
| Treatment5 | Declining | Long |
| Treatment6 | High | Long |

To start with, all claimants who became unemployed and registered for unemployment benefits were randomly assigned in either the control group or one of the 6 treatment groups. The groups differ between them on more coordinates:

i) on the type of bonus paid, a ”low” amount of about 3 times the weekly unemployment insurance (UI) benefit and a ”high” amount of about 6 times the UI; the ”declining” bonus declined from the ”high” level to zero, over a 12 week period;

ii) on the length of the qualification period: a ”short” period of 6 weeks and a ”long” period of 12 weeks.

As for the rationale behind the Pennsylvania experiment (and the other related experiments in Illinois and New Jersey), the questions that were intended to be answered were two: firstly, could policy relevant treatments yield detectable cost savings to existing UI benefit programs, and secondly, how sensitive are program costs to various elements of the treatment design?

A classical ”careful” (the quotation is from Koenker and Bilias (2001)) framework for the analysis of this kind of duration data is performed in Meyer (1996) . There a Cox proportional hazard analysis of the similar bonus experiments in Illinois is done, with Meyer handling the qualification period by the introduction of a time-varying covariates that would permit a discrete jump in the treatment effect at the end of the qualification period. Koenker and Bilias (2001) debate however the plausibility of such a jump, although they do stress that accommodation of time-varying covariate effects such as in the Cox model is an important challenge in extending the applicability of quantile regression to survival data analysis.

The central idea in designing the basic model of Koenker and Bilias (2001) boils down to assuming that the logarithm of the duration, using weeks as time unit, of subjects' spells on UI benefits has linear conditional quantile function of the form

$$Q_{\log(T)}(\tau|x) = x'\beta(\tau) \tag{26}$$

Koenker and Bilias (2001) motivate the choice of the log transformation by the desire to linearize the parametric specification and for the sake of the interpretation. One can see clearly the role of the transformation in the quantile regression setting, with (25) above implying

$$Q_T(\tau|x) = h^{-1}(x'\beta(\tau)) \tag{27}$$

Inter alia, the model includes the following effects: indicators for the 5 treatment groups, with treatments 4 and 6 pooled (see table above); indicators for female, black and Hispanic respondents; number of dependents, with 2 indicating two or more dependents; indicators for whether the respondent was young (age less than 35) or old (older than 54). The results from the estimation of the model can be visualized in plots denoting one coordinate of the vector-valued function, $\widehat{\beta}(\tau)$, viewed as a function of $\tau \in [\alpha, 1-\alpha]$. In Koenker and Bilias (2001) $\alpha$ is chosen to be .20, effectively neglecting thus the proportion of the sample that are immediately reemployed in week one and those whose employment exceeds the insured limit (26 weeks).

Now, before actually discussing the estimates, one should understand how to interpret them in the light of the survival analysis. Basically, the easiest case would be that we are in the pure location shift model in which, looking back at the general form (24), the expression would correspondent to the classical accelerated failure time model

$$\log T_i = x_i'\beta + u_i, \text{ with } u_i \text{ iid from some distribution } F \tag{28}$$

Empirically this would correspond to the coefficients $\widehat{\beta}_\tau$ oscillating around a constant value and thus alluding to the fact that the shift due to a change in the covariates is constant over the entire observed range of the distribution. Another possible outcome would be the linear location-scale model

$$\log T_i = x_i'\beta + (x_i'\gamma)u_i, \text{ with } u_i \text{ taken again iid} \tag{29}$$

The difference with (28) is that now the covariates are allowed to influence both the scale and the location of the conditional distribution of durations. Observationally this would be equivalent to the slope coefficients $\widehat{\beta}_\tau$ looking just like the intercept coefficient up to a location and a scale shift.

So what is actually observed by doing the estimation of this model? Basically treatments 1,2,3 and 5 are estimated to be only marginally significant, achieving a rather modest reduction of duration only in the center of the distribution of about 10%. On the contrary, the combined treatments 4 and 6 (offering a high bonus and long qualification period) induced a 15% reduction in median duration, with a much stronger statistical significance than seen in the other treatments. The authors also noted that no treatment effect is observed in the tails, implying that the treatments had no effects in changing the probability of immediate reemployment (week one) or in affecting the probability of durations beyond the insured period (beyond week 26). Furthermore, with regards to the covariates considered, women seem to be 5 to 15% slower than men in completing durations, Blacks and Hispanics are much quicker reemployed than whites, the young tend to get reemployed earlier than the middle aged and much earlier than the old. These would be basically the main estimation results from the model of Koenker and Bilias (2001).

The more important question is of course why we had to go through such a procedure like quantile regression to do all this, or in other words what did quantile regression analysis bring more when compared with a more conventional type of survival analysis? As we have seen in the discussion above, the bonus effect gradually reduced durations from a null effect in the lower tail to a maximum reduction of 10% (for treatments 1-4 and 5) and 15% (for the pooled treatments 4 and 6) at the median and then it gradually returned again to a null effect in the upper tail. Or this is clearly not a support for the location shift paradigm from (28), which would be estimated by conventional approaches, since this would require that the treatments exert a constant percentage change in all durations. Actually the finding is in perfect agreement with the timing imposed by the qualification period of the experiment.

The study by Koenker and Bilias (2001) seems to actually recommend decision makers that, since the effect of the bonus experiment is considerably attenuated away from the median and essentially null in both the upper and the lower tails of the distribution, the extrapolated effect to the whole eligible population would render a modest net saving to the UI system. However, the plus is to be searched for really in the methodology employed than in the possible policy outcome conclusion[8]. It is more than clear that providing a way to focus on particular regions of the conditional duration distribution, quantile regression proves to be a more flexible approach than traditional survival analysis methods.

### 3.2 Recursive Structural Equation Models

Before examining the contribution of quantile regression to the structural models, let us first define some essential notions that will be used throughout this exposition. Chesher (2003) elegantly defines a ”structure” as being:

1. a system of equations delivering unique values of observable outcomes given values of covariates and latent variates, and,
2. a conditional probability distribution of latent variates given covariates

[8] Meyer (1995) for instance does give some very persuasive arguments for implementing the bonus system on a larger scale resorting to incentive effects on eligibility.

Chesher further deals in his paper with the identification issue in structural models: each structure implies a conditional distribution of outcomes given covariates and implicitly the same distribution may be generated by different structures; in this case a structural feature that takes different values than its value in the data generating structure cannot be identified. This can be equivalently written as a feature of a structure being identified when among any set of observationally equivalent admissible structures, the value of the structural feature does not vary. The restrictions defining admissible structures constitute what we call a "model".

I will not delve further in the study of Chesher (2003) but instead consider a more pragmatic perspective of estimation of the type of structural models in Chesher(2003): namely, I will shortly review the paper by Ma and Koenker (2003). This article discusses two classes of quantile regression methods for the recursive structural equation models of Chesher (2003), one based directly on the identification strategy of Chesher, the second being a control variate approach. An empirical application of the methods to the study of the effect of class size on the performance of students is also contained in Ma and Koenker (2003).

There is some literature on the estimation of the structural equation model starting in the 80's. Amemiya (1982) was essentially the first study to consider this topic, showing the consistency and asymptotic normality of a class of two-stage median regressor estimators. More recent work has extended the conditional median problem: for instance Abadie, Angrist and Imbens (2002) have considered quantile regression methods for estimating endogenous treatment effects focusing on the binary treatment case; Chesher (2003) has expanded considerably the scope of quantile regression methods for structural econometric models by considering triangular stochastic structures and conditioning recursively so that the structural effects are identified and characterized. These latter models of Chesher (2003) were used then in Ma and Koenker (2003) for estimation and inference.

Consider the nonlinear recursive model:

$$Y_{i1} = \varphi_1(Y_{i2}, x_i, \nu_{i1}, \nu_{i2}) \tag{30}$$

$$Y_{i2} = \varphi_2(z_i, x_i, \nu_{i2}) \tag{31}$$

where we assume that $\nu_{ij}$ is iid with $\nu_{ij} \sim F_j$ $(j = 1,2)$. The pairs $(\nu_{i1}, \nu_{i2})$ are also independent of $(z_i, x_i^T)^T$. Other assumptions are that $\varphi_1$ is assumed strictly monotonic in $\nu_1$ and differentiable with respect to $Y_{i2}$ and $x$, while $\varphi_2$ is strictly monotonic in $\nu_2$ and differentiable in $z$ and $x$. Having all these conditions fulfilled one can write the conditional quantile functions:

$$\begin{cases} Q_1(\tau_1|Q_2(\tau_2|x,z),x) = \varphi_1(Q_2(\tau_2|x,z),x,F_1^{-1}(\tau_1),F_2^{-1}(\tau_2)) \\ Q_2(\tau_2|x,z) = \varphi_2(z,x,F_2^{-1}(\tau_2)) \end{cases} \tag{32}$$

In order to proceed with the estimation, the model is described further by expressions (30) and respectively (31), but this time it is explicitly assumed that the functions $\varphi_1$ and $\varphi_2$ are known up to some finite dimensional parameter vectors $\alpha$ and $\beta$, respectively. Under these conditions we will have an inverse function for $\varphi_2$ w.r.t $\nu_2$ and so if we denote this inverse function with $\widetilde{\varphi}_2$ we have

$$\nu_{i2} = \widetilde{\varphi}_2(Y_{i2}, z_i, x_i; \beta) \tag{33}$$

and it follows immediately that

$$Y_{i1} = \varphi_1(Y_{i2}, x_i, v_{i1}, \widetilde{\varphi}_2(Y_{i2}, z_i, x_i; \beta); \alpha) \tag{34}$$

Ma and Koenker (2003) write the conditional functions of $Y_1$ and $Y_2$ as

$$\begin{gathered} Q_1(\tau_1|Y_{i2}, x_i, z_i) = h_1(Y_{i2}, x_i, z_i; \theta) \\ Q_2(\tau_2|z_i, x_i) = h_2(z_i, x_i; \beta) \end{gathered} \tag{35}$$

In the light of this formulation, fixing $\tau_1$ and $\tau_2$ implies that $\theta(\tau_1)$ and $\beta(\tau_2)$ can be estimated by solving the following optimization problems:

$$\widehat{\theta}(\tau_1) = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} \sigma_{i1}\rho_{\tau 1}(Y_{i1} - h_1(Y_{i2}, x_i, z_i; \theta)) \tag{36}$$

and respectively

$$\widehat{\beta}(\tau_2) = \arg\min_{\beta\in B} \sum_{i=1}^{n} \sigma_{i2}\rho_{\tau 2}(Y_{i2} - h_2(x_i, z_i; \beta)) \tag{37}$$

with the weights $\sigma_{ij}$ being strictly positive.

One objective of the study in Ma and Koenker (2003) is the estimation of the weighted average quantile treatment effect implied by Chesher (2003). To this extent the authors consider first the class of weighted average derivative estimators:

$$\widehat{\pi}_1(\tau_1, \tau_2) = \sum_{i=1}^{n} w_i \left[ \nabla_y \widehat{h}_1(\tau_1|y, x_i, z_i, \widehat{\theta}) + \frac{\nabla_z \widehat{h}_1(\tau_1|y, x_i, z_i, \widehat{\theta})}{\nabla_z \widehat{h}_2(\tau_2|x_i, z_i, \widehat{\beta})} \right] \tag{38}$$

evaluated at $y = h_2(\tau_2|x_i, z_i, \beta)$. For the structural effect of $x$ the form is of course similar to the one above in (38):

$$\widehat{\pi}_2(\tau_1, \tau_2) = \sum_{i=1}^{n} w_i \left[ \nabla_x \widehat{h}_1(\tau_1|y, x_i, z_i, \widehat{\theta}) - \frac{\nabla_z \widehat{h}_1(\tau_1|y, x_i, z_i, \widehat{\theta})}{\nabla_z \widehat{h}_2(\tau_2|x_i, z_i, \widehat{\beta})} \nabla_x \widehat{h}_2(\tau_2|x_i, z_i, \widehat{\beta}) \right] \tag{39}$$

The weights are by hypothesis positive and summing up to 1. A convenient choice exploited in Ma and Koenker (2003) is $w_i \equiv n^{-1}$.

The estimator obtained

$$\widehat{\pi}_n(\tau_1, \tau_2) = (\widehat{\pi}_1(\tau_1, \tau_2), \widehat{\pi}_2^T(\tau_1, \tau_2))^T \tag{40}$$

is based on Chesher's (2003) identification strategy. According to Ma and Koenker, its advantage rests in the fact that this estimator takes a rather skeptical attitude toward the original model and is thereby based on a rather loosely restricted form of the two conditional quantile functions.

Ma and Koenker (2003) derive also a control variate estimator for the structural quantile treatment effect. In order to apply this approach, the conditional $\tau_2$ quantile function of $Y_2$ has to be estimated so that we can recover an estimate of $v_2(\tau_2) = v_2 - F_2^{-1}(\tau_2)$. In this respect, in analogy to (35) above, we denote

$$\begin{gathered} Q_1(\tau_1|Y_{i2}, x_i, v_{i2}(\tau_2)) = g_1(Y_{i2}, x_i, v_{i2}(\tau_2); \alpha(\tau_1, \tau_2)) \\ Q_2(\tau_2|z_i, x_i) = g_2(z_i, x_i; \beta(\tau_2)) \end{gathered} \tag{41}$$

Solving similarly to (37) for

$$\widehat{\beta}(\tau_2) = \arg\min_{\beta \in B} \sum_{i=1}^{n} \sigma_{i2} \rho_{\tau 2}(Y_{i2} - g_2(z_i, x_i; \beta)) \tag{42}$$

the conditions on $\varphi_2$ ensure, as previously for the other estimator, that one can invert to obtain

$$v_2 = \widetilde{\varphi}_2(Y_2, z, x, \beta) \tag{43}$$

and thus, successively:

$$F_2^{-1}(\tau_2) = \widetilde{\varphi}_2(g_2(z, x; \beta), z, x; \beta) \tag{44}$$

$$\widehat{v}_{i2}(\tau_2) = \widetilde{\varphi}_2(Y_{i2}, z_i, x_i; \widehat{\beta}) - \widetilde{\varphi}_2(g_2(z_i, x_i; \widehat{\beta}), z_i, x_i; \widehat{\beta}) \tag{45}$$

Having (45) we can estimate $g_1(.)$ and then solve for $\widehat{\alpha}(\tau_1, \tau_2)$ once $F_1^{-1}(\tau_1)$ is stored into the new parameter vector $a$:

$$g_1(Y_{i2}, x_i, \widehat{v}_{i2}(\tau_2); a) = \varphi_1[Y_{i2}, x_i, F_1^{-1}(\tau_1), \widehat{v}_{i2}(\tau_2); \alpha] \tag{46}$$

$$\widehat{\alpha}(\tau_1, \tau_2) = \arg\min_{a \in A} \sigma_{i1} \rho_{\tau 1}[Y_{i1} - g_1(Y_{i2}, x_i, \widehat{v}_{i2}(\tau_2); a)] \tag{47}$$

As a comparison between the two estimation frameworks now, we see that this control variate method is valid regardless of the dimension of $z_i$, meaning that as long as the model is identifiable $\widehat{v}_{i2}(\tau_2)$ has information on all of the available instruments, as seen from the derivations in expressions (41)-(45), and does not necessarily need a single instrumental variable $z$ available, without over-identification, as we need for the derivations in (35)-(39) above. However in the first estimator $z_i$ was introduced directly and did not need a parsimonious construction like in the second case.

Ma and Koenker (2003) also describe an extension of their results to a system of $m$ structural equations and to this extent use again Chesher (2003) who proved that there are no real impediments to extension of the recursive structural model to more than 2 equations. While I will skip the review of the rather technical asymptotics treatment, which would take up considerable space, I shall continue by shortly describing the empirical implementation in Ma and Koenker (2003).

The authors reconsider an application by Levin (2001) investigating the effect of class size on student performance in Dutch primary schools. The data used is the first wave of the PRIMA cohort study, which contains detailed information on Dutch primary school students in grades 2,4,6, and 8 and information on the associated teacher and school characteristics for 1994/1995. This comprehensive survey of primary education in the Netherlands can therefore enable the testing of a series of very interesting links between scholastic achievements (in this context the subjects were tested with regards to intelligence, reading abilities, the Dutch language and mathematics) of the pupils, their characteristics, their socioeconomic background as well as class and school level features. The authors use about 450 schools from about the 700 in total in the survey, considering only grades 4,6, and 8, which are pooled together in the analysis. Since it is not in our purpose and it would take up too much space, I shall ignore here otherwise extremely important considerations on the model specifications, particularly concerns on the exogeneity of the class size and by implication, on the causality mechanisms at play. The approach to the modelling is based on a conventional linear structural equation model:

$$y = \alpha_0 + X_i\alpha_1 + X_c\alpha_2 + X_s\alpha_3 + Y\delta + u \tag{48}$$

$$Y = \beta_0 + X_i\beta_1 + X_c\beta_2 + X_s\beta_3 + Z\gamma + U \tag{49}$$

where math or language test scores are denoted by $y_i$ for student $i$ in class $c$ and school $s$; $X_i$ are pupil $i$'s individual characteristics such as gender, IQ, socioeconomic status, peer effects; $X_c$ are class $c$'s characteristics including teacher's experience; $X_s$ are school $s$'s characteristics including school denomination (public/non-public); $Y$ is the covariate for class size and $Z$ denotes the instrument for class size; $u$ and $U$ designate random components. $Z$ is a very interesting IV representing the reported weighted school enrollment (WSE) to the Dutch Ministry of Education[9] with weights determined by the socioeconomic status of the students. It seems clear that the WSE should be closely related to the actual class size but that it should not have any direct relationship with the students' performances, conditional on characteristics.

After some specification search which is inherent in this sort of modelling, Ma and Koenker (2003) select a final model where the class size is allowed to influence both the location and the scale of the student performance distribution. That means, that with reference to the random components $u$ and $U$ from (48) and respectively (49), the following apply:

$$u_i = (\lambda\nu_{i2} + \nu_{i1})(Y_i\xi + 1) \tag{50}$$

$$U_i = \nu_{i2} \tag{51}$$

with $\nu_1$ and $\nu_2$ independent of one another and iid over individuals. Both estimation frameworks discussed above are used. Ma and Koenker (2003) observe that when the model is correctly specified both methods yield consistent estimators, as expected from the theoretical analysis, with the control variate estimate being more efficient.

In terms of structural estimation upshots, both estimation methods perform similarly, which sort of provides a first support for the model specification. The findings are briefly that, as far as class size effects on language scores are concerned, for weaker students the larger classes are better, for students at the median class size effects are not significant and for better students smaller classes appear to be marginally profitable. As for the class size effect on math scores, for weaker students, contrary to the language situation, smaller classes are better, while for the average and better students the class size effect is not significant.

The important conclusion is that the class size has no significant influence on median performance in either language or math, result which is consistent with previous literature investigating the same relationships by employing conditional mean estimation. The difference with this previous literature (very much the biggest part of the existing literature actually) is that seemingly one should be cautious when interpreting findings of insignificant mean effects since they can arise from averaging incorrectly

[9]The Dutch Ministry of Education imposed a new funding allocation rule during the time period of the first PRIMA experiment wave: each primary school had to report this WSE variable and on its basis the schools were allocated funding to each school, the funding determining how many teachers the school could hire. The WSE variable has been used as an IV in Levin (2001) and has been taken over to Ma and Koenker (2003) from there.

opposite significant benefits for different locations on the distribution conditional on covariates: from the case of the language performance in this context for instance, significant benefits from reductions in class size for good students and significant benefits from increases in class size for poorer students[10].

## 4 Summary and Conclusions

This report presented a general overview of the quantile regression method, consisting of a non-technical introduction to the basic model and its crucial features and of a short review of two major applications. We have seen that quantile regression offers an extension of univariate quantile estimation to estimation of conditional quantile functions and that it complements the established mean regression methods by adding more flexibility in the estimation and more robustness particularly in non-Gaussian distribution settings. The covariate effects are allowed to influence location, scale and shape of the response distribution unlike conventional techniques which usually investigated location-shift paradigms. Furthermore, by focusing on local parts of the conditional distribution, quantile regression methods offer a useful deconstruction of conditional mean models. We have also learned that quantile regression is a powerful tool applied to duration analysis in general and that it provides excellent assistance to elegantly solve structural equation models, these being however just two of the successful recent applications of quantile regressions, far too many as a whole to be discussed in one paper. Despite the fact that the bulk of the specialized literature can still be viewed as paying tribute to the traditional mean regression approach, quantile regression appears to convince more and more. Indeed, quantile regression promises to be a challenging but fascinating research field from a theoretical point of view and at the same time, a technique ever growing and reaching out for more and more applications, from an empirical perspective. In fewer words and probably most concisely, quoting from Koenker and Hallock (2001), "quantile regression is gradually developing into a comprehensive strategy for completing the regression picture".

## References

Abadie, A., J. Angrist & G. Imbens (2002), 'Instrumental variables estimates of the effects of subsidized training on the quantiles of trainee earnings', *Econometrica* **70**, 91–117.

Amemiya, T. (1982), 'Two stage least absolute deviations estimators', *Econometrica* **50**, 689–711.

Barrodale, I. & F. Roberts (1973), 'An improved algorithm for discrete linear approximation', *SIAM Journal of Numerical Analysis* **10**, 839–48.

Barrodale, I. & F. Roberts (1974), 'Solution of an overdetermined system of equations in the l1 norm', *Communications ACM* **17**, 319–320.

[10]Ma and Koenker (2003) stress several times in their paper that of course it is not changes in class size per se that produce academic gains or losses, but their combination with other instructional practices and institutional settings. Nonetheless and as a sequel to what has been just stated, the crucial point here is that structural methods based on quantile regression may be able to constructively contribute to the policy debate around these issues, by providing a more nuanced view of the apparently heterogeneous effects of the class size.

Buchinsky, M. (1998), 'Recent advances in quantile regression models: A practical guide for empirical research', *Journal of Human Resources* **33**, 88–126.

Chaudhuri, P. (1991), 'Global nonparametric estimation of conditional quantile functions and their derivatives', *Journal of Multivariate Analysis* **39**, 246–269.

Chaudhuri, P., K. Doksum & A. Samarov (1997), 'On average derivative quantile regression', *Annals of Statistics* pp. 715–44.

Chesher, A. (2003), 'Identification in nonseparable models', *Econometrica* **71**, 1405–42.

Doksum, K. (1974), 'Empirical probability plots and statistical inference for nonlinear models in the two sample case', *Annals of Statistics* **2**, 267–277.

Doksum, K. & M. Gasko (1990), 'On a correspondance between models in binary regression and survival analysis', *International Statistical Review* **58**, 243–52.

Fitzenberger, B. (1997), *A Guide to Censored Quantile Regressions*, pp. 235–87.

Fitzenberger, B., R. Koenker & J.A.F. Machado (2002), 'Economic applications of quantile regression (studies in empirical economics)'.

Greene, W.H. (1993), *Econometric Analysis, 5th Edition*, Pearson Education, Inc., New Jersey.

Honore, B., J. Powell & S. Khan (2000), Quantile regression under random censoring. preprint.

Horowitz, J.L. & G.R. Neumann (1987), 'Semiparametric estimation of employment duration models', *Econometric Reviews* **6**, 5–40.

Huber, P. (1967), The behavior of maximum likelihood estimates under nonstandard conditions, *in* 'Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics', Berkeley: University of California Press.

Judd, K.L. (1999), *Numerical Methods in Economics*, MIT press.

Koenker, R. (1997), *Rank Tests for Linear Models*, North-Holland.

Koenker, R. & G. Bassett (1978a), 'The asymptotic distribution of the least absolute error estimator', *Journal of the American Statistical Association* **73**, 618–622.

Koenker, R. & G. Bassett (1978b), 'Regression quantiles', *Econometrica* **46**, 33–50.

Koenker, R. & I. Mizera (2003), Penalized triograms: Total variation regularization for bivariate smoothing. Journal of Royal Statistical Society, forthcoming.

Koenker, R. & J. Machado (1999), 'Goodness of fit and related inference processes for quantile regression', *Journal of the American Statistical Association* **94**, 1296–1310.

Koenker, R. & K. Hallock (2001), 'Quantile regression', *Journal of Economic Perspectives* **15**, 143–156.

Koenker, R. & O. Geiling (2001), 'Reappraising medfly longetivity: A quantile regression approach', *Journal of American Statistic Association* **96**, 458–468.

Koenker, R., P. Ng & S. Portnoy (1994), 'Quantile smoothing splines', *Biometrika* **81**, 673–680.

Koenker, R. & Y. Bilias (2001), 'Quantile regression for duration data: A reappraisal of the pennsylvania reemployment bonus experiments', *Empirical Economics* **26**, 199–220.

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

Levin, J. (2001), 'For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement', *Empirical Economics* **26**, 221–246.

Ma, L. & R. Koenker (2003), Quantile regression methods for recursive structural models, Technical report.

Meyer, B. (1995), 'Lessons from the u.s. unemployment insurance experiments', *Journal of Economic Literature* **33**, 91–131.

Meyer, B. (1996), 'What have we learned from the illinois reemployment bonus experiment', *Journal of Labor Economics* **14**, 26–51.

Portnoy, S. & R. Koenker (1997), 'The gaussian hare and the laplacian tortoise: Computability of squarred-error vs. absolute-error estimators, with discussion', *Statistical Science* **12**, 279–300.

Powell, J.L. (1986), 'Censored regression quantiles', *Journal of Econometrics* **32**, 143–55.

Wooldridge, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge, Massachusetts.