

# Linear quantile regression models for longitudinal experiments: an overview

Maria Francesca Marino

*University of Perugia*

Alessio Farcomeni

*Sapienza - University of Rome*

## Abstract

We provide an overview of linear quantile regression models for continuous responses repeatedly measured over time. We distinguish between marginal approaches, that explicitly model the data association structure, and conditional approaches, that consider individual-specific parameters to describe dependence among data and overdispersion. General estimation schemes are discussed and available software options are listed. We also mention methods to deal with non-ignorable missing values, with spatially dependent observations and nonparametric and semiparametric models. The paper is concluded by an overview of open issues in longitudinal quantile regression.

**Key Words:** quantile regression, longitudinal data, marginal models, conditional models, random effects, fixed effects, generalized estimating equations.

## 1 Introduction

Quantile regression has been introduced by Koenker and Bassett (1978) as an extension of standard mean regression to model the conditional quantiles of a continuous response and to provide a thorough overview of its distribution. It has become a very popular and consolidated approach in the statistical literature and is nowadays applied in a wide range of fields, including econometrics, finance, biomedicine, ecology. Comprehensive reviews can be found, among

others, in Koenker and Hallock (2001), Koenker (2005), Hao and Naiman (2007) and Yu *et al.* (2003). Specific examples are provided by Machado and Mata (2005) in econometrics, by Austin and Schull (2003) in epidemiology, by Cade *et al.* (2005) in ecology. The literature on quantile regression methods is now extremely vast. In this work we will review a specific area of application; in particular, we will focus on *linear* quantile regression models for longitudinal observations. In the last two decades, longitudinal study designs have raised considerable attention. Obtaining additional information from a unit already in the study is cheaper than adding a new one; also, longitudinal studies allow to monitor the evolution of individual trajectories over time. Weiss (2005) lists a number of other benefits of collecting longitudinal in place of cross-sectional data. Additional issues have to be faced when dealing with longitudinal studies, though. Observations from the same individual are naturally dependent and this has to be taken into consideration to avoid potential bias in parameter estimates; moreover, individuals may leave the study before the planned end, thus, presenting incomplete data records. In such a context, standard regression models can not be directly used, and need to be extended to avoid misleading inferential conclusions. We will not propose a unifying framework, but rather try to discuss the available options in a logically sound sequence, discussing advantages and limitations of each of them. We will approach both the model specification and the estimation of model parameters mainly from a statistical perspective. Quantile regression is indeed a popular methodology also in the econometric framework; however, in some cases, a clash does exist between the two contexts, with rationale, language and problems to be faced that can be quite different. We will also try to underline that longitudinal quantile regression has some similar (e.g., dependence) and some different challenges with respect to standard longitudinal regression. Quantile regression is more appropriate than mean regression in a number of situations. In some cases, the researcher's interest may not rely on the center of the distribution, but rather in its tails; furthermore, covariates may have different effects on different quantiles. In HIV research, for instance, the effect of the treatment is more important on the left tail of the CD4 count distribution, where individuals are at higher risk. Additionally, a treatment might be beneficial for patients at lower risk but detrimental (due to complications or side effects) for those in the left tail. Other examples include longitudinal fetal growth studies (e.g., Daniel-Spiegel *et al.* 2013), which usually focus on low/high quantiles of key anthropometric measurements and, more in general, growth curves (Wei *et al.*, 2006; Wei

and He, 2006). An example regarding monitoring physical activity in children is given in Geraci and Farcomeni (2015). Another crucial point is that outliers may be present in the observed data, and this poses questions on the reliability of mean regression estimates. In these cases, one could use robust regression approaches (see Heritier *et al.*, 2009; Farcomeni and Ventura, 2012; Farcomeni and Greco, 2015, for reviews) or focus on the *median* of the outcome distribution. Quantile regression, additionally, allows the user to avoid transformations of the outcome in many cases, making parameter estimates more readily interpretable. While in cross-sectional designs a satisfactory transformation to normality might be often found, in longitudinal studies the outcome distribution may change in shape (e.g., skewness) at each time occasion. Thus, a global transformation may make time-specific distributions far from Gaussian, and time-specific transformations may lead to parameter estimates that are quite hard to be interpreted. The rest of the paper is structured as follows: in the next section, we briefly review the essentials of cross-sectional linear quantile regression. In Section 3, we provide an overview of modelling specifications for longitudinal quantile regression, while estimation strategies are presented in Section 4. In Section 5, we discuss several extensions and open issues; available software is briefly listed in Section 6.

## 2 Quantile regression for independent observations

Let  $y_1, \dots, y_n$  be the realizations of  $n$  independent and identically distributed random variables,  $Y_1, \dots, Y_n$ , denoting  $n$  copies of a continuous outcome of interest. Let also  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  be a  $p$ -dimensional vector of explanatory variables, recorded for unit  $i = 1, \dots, n$ . In this section, we deal with a cross-sectional experiment based on independent data. A possible way to characterize the distribution of  $Y_i$  is in terms of *quantiles*. The  $\tau$ -th quantile  $Q_\tau(y)$  of a variable  $Y$  is conveniently defined as

$$Q_\tau(y) = \arg \min_c \sum_{i=1}^n \mathbb{E}[\rho_\tau(y_i - c)],$$

where  $\rho_\tau(u) = u [\tau - \mathbb{I}(u < 0)]$  is an asymmetric absolute loss function that assigns weights  $\tau$  and  $(\tau - 1)$  to positive and negative deviations, respectively,

and is usually referred to as the *quantile loss function*. Now, suppose we want to estimate a *conditional* quantile, that is, the quantile of  $Y_i$  conditional on a given covariate configuration  $\mathbf{x}_i$ . For a fixed  $\tau \in (0, 1)$ , the following quantile regression model can be specified

$$Q_\tau(y_i | \boldsymbol{\beta}, \mathbf{x}_i) = \mathbf{x}_i' \boldsymbol{\beta}_\tau, \quad (1)$$

where  $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$  denotes a vector of unknown, fixed, parameters summarizing the effects of  $\mathbf{x}_i$  on the  $\tau$ -th (conditional) response quantile. Expression (1) can be, alternatively, formulated through the linear model

$$y_i = \mathbf{x}_i' \boldsymbol{\beta}_\tau + \varepsilon_i, \quad (2)$$

where  $\varepsilon_i$  is a random error term. The assumption  $Q_\tau(\varepsilon_i | \boldsymbol{\beta}, \mathbf{x}_i) = 0$  is introduced to guarantee that the random errors are centred on the  $\tau$ -th quantile. Interpretation of  $\boldsymbol{\beta}_\tau$  is straightforward: the intercept term simply represents the baseline predicted quantile, while each slope can be interpreted as the rate of change of the  $\tau$ -th response quantile per unit change in the value of the corresponding covariate (the others being fixed). Estimation of  $\boldsymbol{\beta}_\tau$  in (1) or (2) proceeds by solving

$$\hat{\boldsymbol{\beta}}_\tau = \arg \min_{\boldsymbol{\beta}_\tau} \sum_{i: y_i \geq \mathbf{x}_i' \boldsymbol{\beta}_\tau} \tau |y_i - \mathbf{x}_i' \boldsymbol{\beta}_\tau| + \sum_{i: y_i < \mathbf{x}_i' \boldsymbol{\beta}_\tau} (1 - \tau) |y_i - \mathbf{x}_i' \boldsymbol{\beta}_\tau|. \quad (3)$$

As suggested by Koenker and Bassett (1978), optimal solutions can be derived by setting appropriate linear programming algorithms. The most common is a modified version of the Barrodale and Roberts algorithm for  $L_1$ -regression, as described by Koenker and d'Orey (1987, 1994). For large dimensional problems, the Frisch-Newton interior point method (possibly after preprocessing) is a better option, as illustrated by Portnoy and Koenker (1997). Newey and Powell (1987) show that asymmetric least square estimators have properties similar to those of the solutions obtained from (3). A natural link between minimization of the quantile loss function and maximum likelihood theory is given by the assumption that the error term in equation (2) follows an asymmetric Laplace distribution (ALD), see, among others, Koenker and Machado (1999) and Yu and Moyeed (2001). An ALD random variable has density

$$f_y(y | \mu, \sigma, \tau) = \frac{\tau(1 - \tau)}{\sigma} \exp \left\{ -\rho_\tau \left( \frac{y - \mu}{\sigma} \right) \right\}, \quad (4)$$

where  $\rho_\tau(u)$  is the quantile loss function we have previously defined and  $\mu, \sigma$  and  $\tau$  are the location, the scale and the skewness parameter, respectively. By assuming that  $y_1, \dots, y_n$  are independent realizations of random variables  $Y_i \sim \text{ALD}(\mu_i, \sigma, \tau)$ , where  $\mu_i = \mathbf{x}'_i \boldsymbol{\beta}_\tau, i = 1, \dots, n$ , the following likelihood function can be derived

$$L(\boldsymbol{\beta}, \sigma, \tau) = \left[ \frac{\tau(1-\tau)}{\sigma} \right]^n \exp \left\{ - \sum_{i=1}^n \rho_\tau \left( \frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}_\tau}{\sigma} \right) \right\}. \quad (5)$$

It is worth to notice that the assumption of ALD errors is only introduced as a convenient computational trick that allows to recast quantile regression optimization in a maximum likelihood framework. Maximizing the likelihood (5) is equivalent to minimize expression (3) with respect to  $\boldsymbol{\beta}_\tau$ . Furthermore, such a distributional assumption allows several extensions of the basic framework, including modelling dependent observations.

### 3 Quantile regression for longitudinal observations

When data are repeatedly collected on a sample of individuals across time, the independence assumption may no longer hold. Three different sources of variability can influence the dependence between observations from the same individual: between-individual variability (reflecting individual propensities to the event which are shared by all the repeated measures coming from the same unit), within-individual variability (that is serial correlation between measurements from the same individual taken at different time points) and random error. Diggle *et al.* (2002) and Fitzmaurice *et al.* (2004), among others, give a detailed discussion on the topic. If dependence is not properly taken into consideration, model parameter estimates may be severely biased. Two main approaches to deal with dependent observations can be distinguished, essentially with reference to the families of marginal and conditional models. In the first case, the association structure is explicitly specified together with the model for the response quantiles. Here, model parameters have a *population-averaged* interpretation as they describe changes in the response values between separated groups of individuals (in the study population), as distinguished by their covariates. In the second case, the response quantiles and the dependence between repeated measurements are jointly

specified. Individual-specific (fixed or random) parameters are introduced in the model specification; these are shared by all the responses coming from the same sample unit and allow to describe sources of unobserved heterogeneity that influence the dependence between longitudinal observations. In this context, model parameters have an *individual-specific* interpretation, that is they reflect variations in the individual response values over time associated with a change in that individual’s covariates (Fieberg *et al.*, 2009).

An approach which is in-between the marginal and the conditional approach is discussed in Powell (2010). The author considers individual-specific fixed effects for identification purposes only. Parameters associated to the observed covariates are, instead, defined as a function of the “total disturbance”, that is the individual-specific effect plus the random error, thus having a population-averaged interpretation.

We introduce marginal and conditional models in Section 3.1 and 3.2, respectively. Differences between the two approaches are discussed in Section 3.3.

### 3.1 Marginal models

Let  $Y_{it}$  denote a *continuous* longitudinal response recorded on  $i = 1, \dots, n$  individuals at time occasions  $t = 1, \dots, T_i$  and let  $\mathbf{x}_{it}$  be a  $p$ -dimensional vector of explanatory variables associated with the parameter vector  $\boldsymbol{\beta}_\tau$ . Marginal models are specified like cross-sectional ones, that is

$$Q_\tau(y_{it} \mid \boldsymbol{\beta}, \mathbf{x}_{it}) = \mathbf{x}'_{it}\boldsymbol{\beta}_\tau. \quad (6)$$

The resulting error terms  $\varepsilon_{it} = y_{it} - \mathbf{x}'_{it}\boldsymbol{\beta}_\tau$  must fulfil the following assumptions: first,  $\Pr(\varepsilon_{it} \leq 0 \mid \mathbf{x}_{it}) = \tau$ , secondly, the vector of error terms is made by independent components over different individuals but dependent components over repeated measurements on the same individual.

The model is completed by an assumption on the association structure for measures coming from the same individual. This clearly represents an approximation of the true underlying dependence between repeated measurements and is treated as a nuisance parameter. As it is well known since Wedderburn (1974), consistent estimates for the  $\boldsymbol{\beta}_\tau$  parameters can be obtained also if the covariance matrix is misspecified. Even an independence assumption, albeit obviously incorrect, allows to obtain consistent estimates. On the other hand, a good approximation of the true association substantially decreases the MSE of parameter estimates. When compared to mean

regression, the identification of a reasonable covariance structure in the quantile regression framework is more challenging. For example, if the random noise vector for the repeated measures from an individual follows an AR(1) structure, the corresponding noise vector for the quantile regression is no longer AR(1) (Leng and Zang, 2014). This makes any assumption difficult to be motivated; based on these considerations, He *et al.* (2003) recommend the working independence assumption.

Despite these difficulties, some proposals besides such an assumption can be found in the literature. Assuming  $\Pr(\varepsilon_{it} \leq 0, \varepsilon_{it'} \leq 0 \mid \mathbf{x}_{it}) = \delta, t \neq t' = 1, \dots, T_i$ , Fu and Wang (2012) define an exchangeable covariance matrix for  $\mathbf{S}_i = (S_{i1}, \dots, S_{iT_i})$ , with  $S_{it} = \tau - \mathbb{I}(y_{it} - \mathbf{x}'_{it}\boldsymbol{\beta}_\tau \leq 0)$ , as

$$\mathbf{V}_i = \tau(1 - \tau)[(1 - c)\mathbf{I}_{T_i} + c\mathbf{J}_{T_i}]. \quad (7)$$

In the expression above,  $c = \text{Cor}(S_{it}, S_{it'}) = (\delta - \tau^2)/(\tau - \tau^2), \forall t \neq t'$ , while  $\mathbf{I}_{T_i}$  and  $\mathbf{J}_{T_i}$  represent an identity matrix and a square matrix of ones, respectively (both of dimension  $T_i$ ). Simulation results in Fu and Wang (2012) show that assuming an exchangeable structure in place of independence results in a higher efficiency when strong within-subject correlation does exist, at the price of a larger bias in the parameter estimates. A general stationary autocorrelation structure has been proposed by Lu and Fan (2014); here, the covariance matrix of  $\mathbf{S}_i$  is given by

$$\mathbf{V}_i(\mathbf{r}) = \mathbf{A}_i^{1/2} \mathbf{C}_i(\mathbf{r}) \mathbf{A}_i^{1/2}, \quad (8)$$

where  $\mathbf{A}_i = \text{diag}\{\text{Var}(S_{i1}), \dots, \text{Var}(S_{iT_i})\}$  and  $\mathbf{C}_i(\mathbf{r})$  is an autocorrelation matrix indexed by the parameter vector  $\mathbf{r} = (r_1, \dots, r_{T_i-1})$  and defined as

$$\mathbf{C}_i(\mathbf{r}) = \begin{pmatrix} 1 & r_1 & \dots & r_{T_i-1} \\ r_1 & 1 & \dots & r_{T_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{T_i-1} & r_{T_i-2} & \dots & 1 \end{pmatrix}.$$

Estimates for the autocorrelation parameters  $r_{n-l}, l = 1, \dots, T_i - 1$  are obtained by exploiting the method of moments. A homogeneous unstructured covariance matrix may be also assumed; see He *et al.* (2003).

### 3.2 Conditional models

An alternative approach to account for the dependence between longitudinal observations is based on the inclusion in the linear predictor of sources

of unobserved heterogeneity. This heterogeneity comes either from omitted covariates or from a different effect of measured covariates on the response due to genetic, environmental, social and/or economic factors. Individual-specific parameters fruitfully describe these individual features. For a given  $\tau \in (0, 1)$ , a conditional quantile regression model is defined by

$$Q_\tau(y_{it} \mid b_i, \boldsymbol{\beta}, \mathbf{x}_{it}) = b_i + \mathbf{x}'_{it}\boldsymbol{\beta}_\tau, \quad (9)$$

which may be, equivalently, written as

$$y_{it} = b_i + \mathbf{x}'_{it}\boldsymbol{\beta}_\tau + \varepsilon_{it}. \quad (10)$$

In the equation above,  $\varepsilon_{it}$  is an error term whose  $\tau$ -th conditional quantile is identically null, that is  $Q_\tau(\varepsilon_{it} \mid b_i, \boldsymbol{\beta}, \mathbf{x}_{it}) = 0$ , while  $\boldsymbol{\beta}_\tau$  summarizes the relation between the covariates in  $\mathbf{x}$  and the  $\tau$ -th response quantile for an individual whose baseline level is equal to  $b_i$ . The dependence between observations from the same individual  $i = 1, \dots, n$  arises since they share the same  $b_i$ : conditional on this parameter, repeated measures are no longer dependent.

Two different approaches to (conditional) quantile regression may be distinguished, referring to *distribution free* and *likelihood based* methods (Geraci and Bottai, 2014). Within the former approach, *fixed* individual-specific intercepts are considered and treated as pure location shift parameters common to all conditional quantiles. This implies that the conditional distribution for each individual has the same shape, but different locations as long as the  $b_i$ 's are different. Fixed effect quantile regression for longitudinal data has been introduced by Koenker (2004) and it has been subsequently extended to allow for general endogenous covariates (Harding and Lamarche, 2009) and lagged responses (Galvao and Montes-Rojas, 2010; Galvao, 2011).

On the other hand, in the likelihood framework, individual-specific parameters  $b_i$  are assumed to be independent and identically distributed *random* variables; the corresponding distribution allows to explain differences in the response quantiles across individuals (Geraci and Bottai, 2007). General random parameters have been considered by Liu and Bottai (2009) and Geraci and Bottai (2014). Let  $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})$  represent a  $q$ -dimensional vector of individual random parameters, with density  $f_b(\cdot; \mathbf{D}_\tau)$ . Usually,  $\mathbf{D}_\tau$  is a  $\tau$ -dependent covariance matrix. A linear quantile mixed model is, therefore, defined by

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}_\tau + \mathbf{z}'_{it}\mathbf{b}_i + \varepsilon_{it}, \quad (11)$$

where  $\mathbf{z}_{it}$  denotes a subset of  $\mathbf{x}_{it}$  and, as before,  $Q_\tau(\varepsilon_{it} \mid \mathbf{b}_i, \boldsymbol{\beta}, \mathbf{x}_{it}, \mathbf{z}_{it}) = 0$ . As it has been pointed out by Geraci and Bottai (2014), the random structure above allows to account for between-individual heterogeneity associated with given explanatory variables and does not require orthogonality between the observed and the omitted covariates. We mention that, when the error terms in equation (11) are Gaussian random variables, a standard (mean) mixed model for longitudinal observations (Laird and Ware, 1982) is straightforwardly obtained.

When sources of unobserved heterogeneity are time-varying, the time-constant random parameters we have described so far may not properly recover the true value of the parameter estimates, especially on quantiles that are far from the median. The assumption that  $b_i$  is time fixed has been relaxed by Farcomeni (2012) through the specification of a latent Markov model for conditional quantiles. Along these lines, Marino *et al.* (2015) describe a mixed latent Markov model for quantiles where both time-constant and time-varying sources of unobserved heterogeneity are taken into consideration. In Section 3.2.1 alternative distributions for the random parameters are discussed.

### 3.2.1 The random parameter distribution

When unobserved heterogeneity is captured by individual-specific random parameters, the specification of the corresponding distribution requires some caution. Time-constant individual-specific parameters distributed according to a (zero-mean) Gaussian, a symmetric Laplace or a multivariate  $t$  random variable have been considered by Geraci and Bottai (2007), Liu and Bottai (2009), Geraci and Bottai (2014) and Farcomeni and Viviani (2015). The last two distributions represent robust alternatives to the more typical Gaussian assumption.

Farcomeni (2012) considers random intercepts that vary over time according to a homogeneous, first order, latent Markov (LM) chain. The resulting model represents a semi-parametric alternative to the models we have discussed so far; it allows to account for time-varying unobserved heterogeneity and avoids the potential bias deriving from a misspecification of the random parameter distribution. When adopting this approach, the model in equation (10) becomes

$$y_{it} = b_{it} + \mathbf{x}'_{it}\boldsymbol{\beta}_\tau + \epsilon_{it}, \quad (12)$$

where  $b_{it}$  varies according to the states of a (quantile-specific) latent Markov chain and  $Q_\tau(\varepsilon_{it} \mid b_{it}, \boldsymbol{\beta}, \mathbf{x}_{it})$ . Under the assumption of Gaussian random errors, equation (12) reduces to the standard latent Markov model for longitudinal data; see Bartolucci *et al.* (2013) for a detailed discussion on this class of models. By assuming that the hidden transition matrix is diagonal, the random intercepts are time-constant and a parsimonious (quantile) latent class model is obtained. This comes at the price of losing some goodness of fit, so that this assumption should tentatively be tested (Bartolucci, 2006). The choice of a specific random parameter distribution should be data driven; models based on different specifications might be fit and compared according to penalized likelihood criteria, like the AIC (Akaike, 1973) or the BIC (Schwarz, 1978). Few additional guidelines follow. First of all, since linear quantile mixed models are essentially a specific kind of standard mixed models, when we focus on the center of the distribution and the number of repeated measurements per unit is large, the modelling structure is rather robust with respect to random parameter misspecification (Rizopoulos *et al.*, 2008). Even with small  $\max_i T_i$ , a negligible bias is often observed and it is essentially related to covariate effects associated to random parameters (Neuhaus *et al.*, 2013). Serious bias might be expected when focusing on quantiles corresponding to low density regions of the conditional outcome distribution, as well as when the covariates associated with random slopes are misspecified. A related matter is the use of random slopes and similar polynomial models to describe time-varying unobserved heterogeneity: the latent Markov framework naturally adapts to the true underlying dynamic heterogeneity. It is therefore resistant to misspecification of the random parameter distribution (as the discrete support approximates a possibly continuous underlying distribution), and the time-dynamics do not need to be parametrically specified. Of course, there are limitations to these statements, and we point the reader to the discussion in Bartolucci *et al.* (2014) for more details.

### 3.3 Marginal vs Conditional Models

To better highlight differences between the two families of models, we conclude this section by contrasting conditional and marginal approaches. In the mean regression context, Zeger *et al.* (1988) state that marginal models describe the dependence among repeated observations; in this setting, the target population is considered as composed by *homogeneous* individuals

with dependent repeated measurements and, therefore, population quantiles can be directly modelled as a function of the observed covariates. Due to these features, marginal models are often referred to as *population-averaged* models.

On the other hand, conditional models try to describe the potential sources of this dependence. They include in the model specification individual-specific parameters that account for unobserved factors which, once known, would make longitudinal observations from the same individual independent. In this perspective, individual characteristics play a central role: the target population is considered as composed by *heterogeneous* individuals and the effect of covariates on individual quantiles is analysed.

The choice between marginal and conditional model is always context-specific and has to be related to the analysed outcome; see e.g. Lindsey and Lambert (1998), Neuhaus *et al.* (1991), Fieberg *et al.* (2009), Lee and Nelder (2004). The most important thing to keep in mind is that interpretation of parameter estimates is different. In the marginal formulation, parameters  $\beta_\tau$  describe the effect of covariates on the  $\tau$ -th population response quantile. On the other hand, in the conditional framework, regression coefficients have an individual-specific interpretation. In fact,  $\beta_{\tau j}$  represents the change in the  $\tau$ -th quantile of the outcome distribution (associated to a unit increment in the corresponding covariate) for a unit with individual-specific parameters equal to  $b_i$ .

Therefore, the applied researcher must choose the fitting strategy according to the research aims. The use of marginal models for repeated measurements is often discouraged (Lindsey and Lambert, 1998; Crouchley and Davies, 1999; Lee and Nelder, 2004) as predictions correspond to hypothetical individuals only. In almost all cases, therefore, one should fit conditional models. However, population averaged models may be useful in epidemiological or ecological studies, for instance, if the key questions involve a comparison between *disjoint* groups. When the central questions entail evaluation of changes for any given individual (e.g., what happens if a treatment instead of another is used on a individual, what happens after  $t$  time-units, etc.) conditional models are the only sensible choice.

It shall be noticed that conditional models can be specified in conjunction with the assumption that subject-specific parameters are fixed or random. Choosing between fixed and random effects is often a delicate matter. Here, we do not get into too much details about the former and point the interested reader to the econometric literature (see e.g. Canay, 2011 and references

therein).

We only mention some general considerations that may help decide which approach is more appropriate for a given problem. First, random intercept models rely on the independence assumption between sources of observed and unobserved heterogeneity, while such a assumption is not used in the fixed effect counterpart. Therefore, if one thinks that unobserved features are related to the observables ( $\mathbf{x}_{it}$ ), then conditional models with fixed effects should be preferred. However, when general random coefficients are considered, one can let any of the model parameters vary from individual to individual, not just the intercept. This allows to relax the aforementioned independence assumption and leads to a more general model formulation. On the other hand, in the context of conditional models with fixed effects, only individual-specific intercepts can be considered and, although one is able to control for unobserved characteristics, their effect can not be directly estimated. Secondly, time-constant covariates can not be accommodated in conditional models with fixed effects as there would be overlapping between the two components; such a constraint does not exist in the random parameter framework. Last, conditional models with fixed individual-specific parameters require complex identifiability constraints, and in general can not be estimated when the available number of repeated measures per individual is small. Furthermore, the consistency of parameter estimates is ensured only when  $T_i$  diverges. This is another strong limitation in many common applications, where  $T_i$  is generally small and only  $n$  is large. A consequence is that, in many practical situations in the area of applications of statistics, random parameter models are the only viable choice.

We conclude this discussion by mentioning few examples. A benchmark dataset in longitudinal quantile regression is the labor pain data (Davis, 1991), which involves evaluation of pain levels at different stages of labor. A psychological component intuitively makes treatment less effective when we look at high quantiles of the distribution. In this application, population averaged effects would be misleading as we are interested in how pain evolves for *each* woman.

Fabrizi *et al.* (2012) evaluate wage dynamics in the Italian labor market, where the differences, for instance, among qualifications might be evaluated through marginal models. A longitudinal study on child obesity is described by Fenske *et al.* (2013), and the effects of fast-food prices on adult obesity are explored in Han and Powell (2013). Obesity is simply studied by focusing on high quantiles. Bottai *et al.* (2010) discuss data related to depression

in adolescents. In all these cases, conditional quantile regression has been exploited to analyse the effect of observed covariates on the evolution of individual trajectories over time. These data can probably be tackled from different perspectives, with a more descriptive approach, based on marginal modelling, in order to highlight differences between separate groups (e.g., males vs females) and conditional modelling in order to provide “individual” estimates of risk factors.

## 4 Estimation

In this section, we discuss estimation of model parameters for both the marginal and the conditional formulations. Generally, it has to be noticed that the former are computationally much simpler than the latter, as the same (weighted) routines used for cross-sectional data can be used for parameter estimation. On the other hand, when a working covariance structure is assumed, the covariance matrix of the estimates needs to be adjusted by the usual sandwich formula (Huber, 1967; Liang and Zeger, 1986). For general expressions of the sandwich formula in longitudinal quantile regression see, for instance, Fu and Wang (2012) and Leng and Zang (2014).

### 4.1 Marginal models

Marginal models proposed by He *et al.* (2003), Chen *et al.* (2004) and Yin and Cai (2005) use generalized estimating equations (Liang and Zeger, 1986) based on an independence working model; these are simple and have desirable properties but could lead to loss of efficiency when strong correlation does exist. For simplicity, we focus on the estimation of single quantiles, but a generalization to simultaneous estimation of several quantiles (as in the next section) is quite straightforward. Obtaining estimates under an independence working model simply involves a system of estimating equations of the kind

$$\sum_{i=1}^n \sum_{t=1}^{T_i} \mathbf{x}_{it} (\tau - \mathbb{I}(y_{it} - \mathbf{x}'_{it} \boldsymbol{\beta}_\tau \leq 0)) = \mathbf{0}.$$

As it is clear, results derived from the expression above are equivalent to those obtained from the minimization of the quantile loss function presented in Section 2 for cross-sectional data. As remarked above, the only difference is the robustification of the covariance matrix for parameter estimates.

Obviously, when a given covariance matrix,  $\mathbf{V}_i$ , is used for the error terms, estimating equations above need to be properly extended. In the context of median regression models, Jung (1996) suggests a quasi likelihood-type approach (Wedderburn, 1974) based on the following set of *weighted* estimating equations

$$\sum_{i=1}^n \mathbf{X}'_i \mathbf{\Gamma}_i \mathbf{V}_i^{-1} (\tau - \mathbb{I}(y_i - \mathbf{X}_i \boldsymbol{\beta}_\tau \leq 0)) = \mathbf{0}. \quad (13)$$

Here, the matrix  $\mathbf{\Gamma}_i = \text{diag}(f_{i1}(0), \dots, f_{iT_i}(0))$ , with  $f_{it}$  being the probability density function of  $\varepsilon_{it}$ , is introduced to account for possible overdispersion in the error distribution. Clearly, Jung's estimator requires this matrix to be known. Only in the case of identically distributed errors such quantity can be ignored and the resulting estimating functions can be shown to be *optimal* if  $\mathbf{V}_i$  is correctly specified. Lipsitz *et al.* (1997) extend the previous approach by proposing a weighted GEE model for longitudinal data affected by non-informative drop-out (MAR data). Karlsson (2008) develops a quantile weighted version of regression models for non-linear longitudinal data. Tang and Leng (2011) suggest to derive parameter estimates via an empirical maximum likelihood approach, in which a working model for the conditional mean is specified to enhance efficiency. Fu and Wang (2012) split the weighted estimating equation (13) into two components, the within-group and between-group estimating equations. The resulting estimator minimizes the combination of these two functions in a way similar to that of the generalized method of moments (Hansen, 1982). The method suggested by Fu and Wang (2012) is robust to the choice of the error correlation structure and leads to parameters that have a higher efficiency than those derived under the independence assumption. A different approach is proposed by Leng and Zang (2014). To circumvent the problem of correctly specifying the error covariance matrix, the authors suggest to combine multiple sets of estimating equations as those defined in (13), that is

$$\frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \sum_{t=1}^{T_i} \mathbf{x}'_{it} \mathbf{\Gamma}_i \mathbf{M}_{i1}^{-1} (\tau - \mathbb{I}(y_{it} - \mathbf{x}'_{it} \boldsymbol{\beta}_\tau \leq 0)) \\ \vdots \\ \sum_{t=1}^{T_i} \mathbf{x}'_{it} \mathbf{\Gamma}_i \mathbf{M}_{iK}^{-1} (\tau - \mathbb{I}(y_{it} - \mathbf{x}'_{it} \boldsymbol{\beta}_\tau \leq 0)) \end{pmatrix} = \mathbf{0}, \quad (14)$$

where  $\mathbf{M}_{ik}$ ,  $k = 1, \dots, K$ , are known matrices (e.g the identity matrix, the matrix with 0 on the diagonal and 1 off-diagonal, etc...). The last two proposals, as well as that of Lu and Fan (2014), extend the induced smoothing

method (Brown and Wang, 2005) to quantile regression. The corresponding (smoothed) estimating equations can be solved via a standard Newton-Raphson algorithm. Within the context of marginal quantile regression for longitudinal data, asymptotic normality of parameter estimators is guaranteed and a closed form expression for the asymptotic covariance matrix is available. In most of the cases, resampling and perturbing methods are used to estimate these latter quantity (Yin and Cai, 2005).

## 4.2 Conditional models

Parameter estimation in the conditional quantile regression framework is closely related to the assumptions postulated for the individual-specific parameters. In Sections 4.2.1-4.2.2, we discuss estimation procedures for conditional models with fixed and random parameters, respectively.

### 4.2.1 Fixed effect models

Following the proposal by Koenker (2004), when *fixed* individual-specific parameters account for dependence between longitudinal data, conditional quantiles are estimated simultaneously by minimizing a weighted piecewise linear quantile loss function. Let  $\tau = (\tau_1, \dots, \tau_q)$  be the set of quantiles of interest; parameter estimates are obtained by solving

$$\arg \min_{\beta, b} \sum_{k=1}^q \sum_{i=1}^n \sum_{t=1}^{T_i} \omega_k \rho_{\tau_k} [y_{it} - b_i - \mathbf{x}'_{it} \boldsymbol{\beta}_{\tau_k}], \quad (15)$$

where  $\rho_{\tau_k}$ ,  $k = 1, \dots, q$ , is the quantile loss function introduced by Koenker and Bassett (1978) and the weights  $\omega_k$  are introduced to control the influence of the different quantiles on the estimation of individual parameters. The author shows that, under some regularity conditions, parameter estimates obtained by (15) are consistent and asymptotically Gaussian. However, when  $n$  is large relatively to  $T_i$ , a penalized approach is suggested. In this case, a  $\ell_1$  penalty term is considered to shrink individual-specific intercepts towards a common value and parameter estimates are computed via

$$\arg \min_{\beta, b} \sum_{k=1}^q \sum_{i=1}^n \sum_{t=1}^{T_i} \omega_k \rho_{\tau_k} [y_{it} - b_i - \mathbf{x}'_{it} \boldsymbol{\beta}_{\tau_k}] - \lambda \sum_{i=1}^n |b_i|. \quad (16)$$

Koenker (2004) shows that fixed parameter estimators obtained under this approach are asymptotically unbiased and Gaussian; an asymptotic approximation for the optimal value of the tuning parameter  $\lambda$  can be found in Lamarche (2010). Harding and Lamarche (2009), Galvao and Montes-Rojas (2010) and Galvao (2011) build on the proposal by Chernozhukov and Hansen (2008) to develop a two step procedure that exploit instrumental variables to derive parameter estimates. In the first step, a weighted quantile loss function is minimized with respect to model parameters, keeping individual-specific effects constant. In the second step, the latter are estimated by minimizing a weighted distance function defined on the basis of the instrumental variable coefficients. A simpler two step procedure is used by Canay (2011). A transformation of the data is first applied to eliminate individual-specific terms, as in the first-difference approach to mean regression; then, the remaining parameters are estimated. The resulting estimators are shown to be consistent and asymptotically Gaussian when both  $n$  and  $T_i$  go to infinity. Identifiability of conditional quantile regression models with fixed effects has been extensively discussed in the econometric literature. When the number of repeated measurements is small, fixed effects can not be directly estimated. In the context of pure location shift effects (that is, when  $b_i$  is common to all quantiles), Canay (2011) proves identifiability of model parameters under the assumption of independence between random errors and individual parameters, and the existence of first order moments. Under these conditions, model parameters are identified as long as at least two observations are available for each unit. Rosen (2012) focuses on the identification of model parameters outside the pure location shift context; here, identification of model parameters holds if one assumes support conditions and conditional independence between the random errors  $\varepsilon_{it}$ . We point the interested reader to these works and references therein.

#### 4.2.2 Random parameter models

When individual-specific parameters are considered as random with the aim at capturing unobserved heterogeneity and the conditional assumption of asymmetric Laplace errors holds, parameter estimates can be obtained by exploiting a maximum likelihood approach. In the case of time-constant random parameters  $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})$ , the marginal likelihood function is

defined by:

$$L(\boldsymbol{\beta}, \sigma, \mathbf{D}; \tau) = \prod_{i=1}^n \int \prod_{t=1}^{T_i} f_y(y_{it} | \mathbf{b}_i; \boldsymbol{\beta}, \sigma, \tau) f_b(\mathbf{b}_i; \mathbf{D}_\tau) d\mathbf{b}_i. \quad (17)$$

The integral in the expression above does not have a closed form solution and numerical integration methods are required. Geraci and Bottai (2007) and Liu and Bottai (2009) describe the use of a MCEM algorithm (Wei and Tanner, 1990) derived from the context of generalized linear mixed models (Booth and Hobert, 1999). Because of the computational inefficiency of such an approach, Geraci and Bottai (2014) suggest the use of Gaussian quadrature methods to directly maximize the likelihood in (17).

The marginal likelihood function for the case of time-varying random intercepts with a Markovian distribution, as in Farcomeni (2012), can be expressed as

$$L(\boldsymbol{\beta}, \sigma, \boldsymbol{\delta}, \mathbf{Q}; \tau) = \prod_{i=1}^n \sum_{\mathbf{b}_i} f_y(\mathbf{y}_i | \mathbf{b}_i; \boldsymbol{\beta}, \sigma, \tau) f_b(\mathbf{b}_i; \boldsymbol{\delta}_\tau, \mathbf{Q}_\tau), \quad (18)$$

where  $\mathbf{b}_i = (b_{i1}, \dots, b_{iT_i})$  is the vector of individual and time-dependent intercepts evolving according to the latent Markov chain described by  $\boldsymbol{\delta}_\tau$  and  $\mathbf{Q}_\tau$ ; these represent the  $\tau$ -dependent initial probability vector and transition probability matrix, respectively. Farcomeni (2012) adapts the Baum-Welch algorithm (Baum *et al.*, 1970) for parameter estimation. It shall be noticed that this algorithm is a particular specification of a general EM algorithm (Dempster *et al.*, 1977), so that the two names may be used interchangeably. Marino *et al.* (2015) adopt a nonparametric distribution for time-constant and time-varying random parameters exploiting a nonparametric maximum likelihood approach for parameter estimation. The description of a general EM algorithm in the context of quantile regression models follows. Starting from the definition of the complete data log-likelihood, the following two steps are repeatedly alternated until convergence.

- In the E-step, the expected value of the complete data log-likelihood, given the observed data and the current parameter estimates, is computed. Monte Carlo integration or quadrature methods are needed, while a closed form expression is available when a discrete distribution is assumed for the random parameters.
- In the M-step, the expected complete data log-likelihood is maximized with respect to model parameters. The parameters in the longitudi-

nal data model are usually updated through a block algorithm, that is, fixed and random parameters are updated separately. Under the latent Markov formulation (Farcomeni, 2012), closed form expressions for initial and transition probabilities can be derived; under the mixed latent Markov formulation (Marino *et al.*, 2015), closed form expression for the mixture component probabilities are available as well. Fixed parameters are updated by using an algorithm for cross-sectional quantile regression with an offset given by the (current) random predictor. Finally, the scale parameter of the ALD and of the random parameter covariance matrix (if present) are updated through moment matching.

In the conditional mixed model formulation, standard errors are usually provided by means of bootstrap routines; an excellent discussion about resampling in quantile regression is provided by Buchinsky (1995a). Given that for each bootstrap sample we must run an algorithm which relies on non-convex optimization, resampling is often cumbersome. For this reason, the number of replicates (number of bootstrap samples) may be tuned as in Andrews and Buchinsky (2000). An additional issue regards the fact that we are bootstrapping longitudinal data, which is often a delicate matter. We do not get into details here on a comparison between the methods available and simply mention that the usual strategy is to resample *individuals*, rather than measurements trying to preserve the dependence structure between repeated measures. See also Parzen *et al.* (1994) and Yin and Cai (2005). We conclude this section by mentioning Bayesian approaches to mixed quantile regression. Reich *et al.* (2010) discuss a quasi distribution-free Bayesian approach; the error distribution is assumed to arise from an infinite mixture of Gaussian densities subject to a stochastic constraint which enables inference on the quantile of interest. Luo and Lian (2012) considers ALD errors and propose to use a Metropolis-Hastings algorithm or a Gibbs sampling approach to derive parameter estimates.

## 5 Extensions and open issues

In this section, we briefly review some extensions and generalizations of linear longitudinal quantile regression models described in the previous sections and list some open issues. All paragraphs refer to longitudinal data analysis. We

report many results for the cross-sectional case only as a preamble to longitudinal data, or only when they can be directly extended to the longitudinal setting.

## 5.1 Missing values

Missing data are ubiquitous in all statistical applications and, especially, in longitudinal studies. Missing values may be ignorable (non-informative) or non-ignorable (informative). We refer the reader to Diggle and Kenward (1994), Little (1995) and Little and Rubin (2002) for a detailed treatment of the topic. In the ignorable case, standard estimation routines might be used simply discarding the missing values. Use of (multiple) imputation may possibly improve efficiency of the estimators, as suggested by Geraci (2013), who does not focus specifically on longitudinal data but rather on potentially clustered data. A situation where it is often difficult to assume missing values are ignorable is in longitudinal studies with monotone missing data patterns, that is, when there is a drop-out from the study. Drop-out may occur due to reasons that are linked to the unobserved values of the outcome; for instance, in HIV studies patients may die before the last scheduled follow-up, in wage studies individuals may lose their job or retire, and so on. There are very few approaches to quantile regression in the presence of informative drop-out. Lipsitz *et al.* (1997) and Yi and He (2009) discuss the extension of the estimating equation approach to handle missing data, in a MAR context; the observed values are weighted by the inverse of the probability of drop-out. In a Bayesian framework, Yuan and Yin (2010) model missingness as a binary time series sharing a random parameters with the quantile regression process, as in shared parameter models (see e.g. Wu and Carroll, 1988). A joint model for a right-censored time-to-event and the quantile of a continuous longitudinal response is proposed by Farcomeni and Viviani (2015). Marino *et al.* (2015) define a latent drop-out class approach to handle potentially informative drop-out in the context of linear quantile latent Markov models.

## 5.2 Spatial quantile regression

Spatial data arise when data are collected at different sites on a geographical territory. In general, data are dependent and this dependence may be expressed as a function of the distance between locations on a multidimensional (usually, two or three dimensional) space. Spatial data arise in various ar-

eas of research, including econometrics, epidemiology, environmental science. Spatial quantile regression methods deal with the peculiar spatial dependence structure, while modelling conditional quantiles of the outcome of interest. They can be thought of as a direct extension of longitudinal quantile regression methods, where dependence arises according to distance in a one dimensional space (i.e. the time). Hallin *et al.* (2009) introduce a nonparametric conditional spatial quantile regression model; asymptotic behaviours of the local linear quantile regression estimates are derived through asymptotic methods that are typically used when dealing with time series. Kostov (2009) uses a spatial quantile regression model in the context of hedonic models for agricultural land prices to ensure robustness against the possible presence of outliers in the data. Dependence between adjacent sites is modelled by considering lagged responses in the linear predictor and an instrumental quantile regression approach is exploited to compute parameter estimates. Reich *et al.* (2011) develop a spatial quantile model that incorporates spatial dependence through spatially varying regression parameters. These are expressed as a weighted sum of Bernstein basis polynomials, where the weights are constrained spatial Gaussian processes. Lum and Gelfand (2012) specify a spatial quantile regression model for all quantiles of a response variable. Starting from the asymmetric Laplace distribution, a process for quantile regression with spatially dependent errors is defined and exploited to model the dependence between observations coming from neighbouring sites. Finally, under a robust semi-parametric framework, Lu *et al.* (2014) propose a spatial quantile regression model with functional coefficients to deal with the curse of dimensionality when there are more than three covariates. For a general overview on quantile regression models for spatial data, we refer to the monograph by McMillen (2013).

### 5.3 Nonparametric models

The literature on nonparametric quantile regression models is very rich. We only briefly mention the approaches that are tailored (or that can be tailored) to longitudinal studies. Koenker *et al.* (1994) and Koenker and Mizera (2004) use a total variation regularization approach to estimate possible univariate and bivariate nonparametric terms; De Gooijer and Zerom (2003), Yu and Lu (2004), Horowitz and Lee (2005), Cai and Xu (2008) base their estimation on local polynomial fitting; Takeuchi *et al.* (2006) and Li *et al.* (2007) explore reproducing kernel Hilbert space (RKHS) norms for nonparametric

quantile estimation. Fenske *et al.* (2013) use a boosting algorithm to allow for data-driven determination of the amount of smoothness required for nonlinear effects and combine model selection with an automatic variable selection property. Bayesian approaches are available, including Yue and Rue (2011) who use integrated nested Laplace approximation and MCMC methods. Finally, Mu and Wei (2009) propose a varying parameter model for conditional quantiles in longitudinal studies. Censored data are considered by Wang and Fygenon (2009), while Wang *et al.* (2009) deal with partially linear varying parameter models.

## 5.4 Open issues

Longitudinal quantile regression is still a relatively recent field. There are many open issues and many of them would require an extension of methods that have already been established for cross-sectional quantile regression. First, some works deal with outcome transformations in order to move from non-linear to linear relationships when bounded outcomes have to be handled. These include Buchinsky (1995b), Mu and He (2007), Bottai *et al.* (2010), Geraci and Jones (2014); all these works are focused on cross-sectional models and, as outlined before, transformations are more delicate when there are repeated measurements. A further issue that has been almost completely overlooked in longitudinal quantile regression is the joint modelling of more than one outcome. Methods for independent data have been proposed in Hallin *et al.* (2010) and Hahn and Burgette (2012). In the longitudinal setting, this situation is easier to work with as additional random parameters can be used in conjunction with a conditional independence assumption. An example can be found for instance in Capogna *et al.* (2010). A similar issue is the simultaneous modelling of more than one quantile. A rare problem that arises when separately modelling response quantiles (as in conditional mixed quantile regression) is that predictions may cross, that is, for certain covariate combinations a low quantile might be predicted to be larger than a higher one. There is a lot of literature on simultaneous modelling for independent data (Koenker, 2005). The problem is also closely linked to density regression; see e.g. Dunson *et al.* (2007). In the longitudinal setting we are only aware of the Bayesian approach proposed by Todkar and Kadane (2011). A specific issue is related to extremal quantiles: while estimation far in the tails may be extremely useful in many cases, tails are often associated with low density regions. Important contributions to the cross-sectional

context include Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011). Traditional approaches break down as standard errors are inflated and parameter estimates (excluding the intercepts) shrunk towards zero. To tackle these issues, Wang *et al.* (2012) propose a two step approach: first quantile regression is focused on intermediate conditional quantiles yielding to reliable estimates, then parameters are extrapolated to the tails based on assumptions on tail behaviours; see also Li *et al.* (2010). To the best of our knowledge, the issue of dealing with extremal quantiles in longitudinal studies has not been considered so far.

A further issue is that of planning the sample size, with the only exception, to our knowledge, of Chernozhukov and Fernandez-Val (2005). We finally mention that many results and properties of quantile regression which are well known and established in the cross-sectional framework have not been assessed yet in the longitudinal one. Work has still to be done in this area to give more substantial theoretical support to procedures and methods discussed throughout.

## 6 Software

We briefly mention in this section the readily available software to fit longitudinal quantile regression models. We report only on software which can be directly used for conditional or marginal modelling (with sandwich estimators) in the longitudinal quantile regression framework.

Conditional models with Gaussian and AL random parameters can be fit using the R package `lqmm` (Geraci, 2014). Linear quantile mixed models with a latent Markov random intercept can be fit using the R function `lmqr` (Farcomeni, 2012). A non-optimized version is currently available at <http://www.afarcome.altervista.org/lmqr.r>, while a refined version will be soon included in the R package `LMest` (Bartolucci *et al.*, 2015). The `lmqr` function can be used both for latent Markov (using option `lm=T`) and latent class (`lm=F`) models. Conditional models with fixed individual-specific parameters can be fit using methods in package `rqpd`, available from R-Forge. In `rqpd`, one can obtain both the penalized longitudinal quantile regression by Koenker (2004) and the corresponding generalization based on correlated random effects (Abrevaya and Dahl, 2008). Finally, we mention the `qreg` procedure in `Stata`, which provides sandwich estimates of the covariance matrix.

## Acknowledgements

The authors are grateful to two referees for several suggestions.

## References

- J. ABREVAYA AND C. M. DAHL (2008). The effects of birth inputs on birthweight: evidence from the quantile estimation on panel data. *Journal of Business and Economic Statistics*, **26**, 379–397.
- H. AKAIKE (1973). Information theory and an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*, 267–281. Akademinai Kiado.
- D. W. K. ANDREWS AND M. BUCHINSKY (2000). A three-step method for choosing the number of bootstrap repetitions. *Econometrica*, **68**, 23–52.
- P. C. AUSTIN AND M. J. SCHULL (2003). Quantile regression: a statistical tool for out-of-hospital research. *Academic Emergency Medicine*, **10**, 789–797.
- F. BARTOLUCCI (2006). Likelihood inference for a class of latent Markov models under linear hypotheses on the transition probabilities. *Journal of the Royal Statistical Society, Series B*, **68**, 155–178.
- F. BARTOLUCCI, A. FARCOMENI, S. PANDOLFI, AND F. PENNONI (2015). LMest: an R package for latent Markov models for categorical longitudinal data. *arXiv:1501.04448*.
- F. BARTOLUCCI, A. FARCOMENI, AND F. PENNONI (2013). *Latent Markov Models for Longitudinal Data*. Chapman & Hall/CRC Press.
- F. BARTOLUCCI, A. FARCOMENI, AND F. PENNONI (2014). Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates (with discussion). *TEST*, **23**, 433–486.
- L. E. BAUM, T. PETRIE, G. SOULES, AND N. WEISS (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 164–171.

- J. G. BOOTH AND J. P. HOBERT (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B*, **61**, 265–285.
- M. BOTTAI, B. CAI, AND R. E. MCKEOWN (2010). Logistic quantile regression for bounded outcomes. *Statistics in Medicine*, **29**, 309–317.
- B. M. BROWN AND Y. G. WANG (2005). Standard errors and covariance matrices for smoothed rank estimators. *Biometrika*, **92**, **1**, 149–158.
- M. BUCHINSKY (1995a). Estimating the asymptotic covariance matrix for quantile regression models: a Monte Carlo study. *Journal of Econometrics*, **68**, 303–338.
- M. BUCHINSKY (1995b). Quantile regression, Box-Cox transformation model, and the US wage structure. *Journal of Econometrics*, **65**, 109–154.
- B.S. CADE, B.R. NOON, AND C.H. FLATHER (2005). Quantile regression reveals hidden bias and uncertainty in habitat models. *Ecology*, **86**, 786–800.
- Z. CAI AND X. XU (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, **103**, 1595–1608.
- I.A. CANAY (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal*, **14**, **3**, 368–386.
- G. CAPOGNA, M. CAMORCIA, S. STIRPARO, G. VALENTINI, A. GARASSINO, AND A. FARCOMENI (2010). Multidimensional evaluation of pain during early and late labor: a comparison of nulliparous and multiparous women. *International Journal of Obstetric Anesthesia*, **19**, 167–170.
- L. CHEN, L.J. WEI, AND M.I. PARZEN (2004). Quantile regression for correlated observations. In: *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, vol. 179, 51–70.
- V. CHERNOZHUKOV (2005). Extremal quantile regression. *Annals of Statistics*, **33**, 806–839.

- V. CHERNOZHUKOV AND I. FERNANDEZ-VAL (2005). Subsampling on quantile regression processes. *Indian Journal of Statistics*, **67**, 253–276.
- V. CHERNOZHUKOV AND I. FERNÁNDEZ-VAL (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Review of Economic Studies*, **78**, 559–589.
- V. CHERNOZHUKOV AND C. HANSEN (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, **142**, 1, 379–398.
- R. CROUCHLEY AND R. B. DAVIES (1999). A comparison of population average and random-effect models for the analysis of longitudinal count data with base-line information. *Journal of the Royal Statistical Society, Series A*, **162**, 331–347.
- E. DANIEL-SPIEGEL, E. WEINER, I. YAROM, E. DOVEH, P. FRIEDMAN, A. COHEN, AND E. SHALEV (2013). Establishment of fetal biometric charts using quantile regression analysis. *Journal of Ultrasound in Medicine*, **32**, 23–33.
- S. DAVIS (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in Medicine*, **10**, 1959–1980.
- J. DE GOOIJER AND D. ZEROM (2003). On additive conditional quantiles with high-dimensional covariates. *Journal of the American Statistical Association*, **98**, 135–146.
- A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B. Methodological*, 39, 1977. With discussion.
- P. J. DIGGLE AND M.G. KENWARD (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49–93.
- P. J. DIGGLE, P. J. HEAGERTY, K.Y. LIANG, AND S. L. ZEGER (2002). *Analysis of longitudinal data*, vol. 25 of *Oxford Statistical Science Series*. Oxford University Press, second edn.

- D. B. DUNSON, N. PILLAI, AND J.-H. PARK (2007). Bayesian density regression. *Journal of The Royal Statistical Society, Series B*, **69**, 163–183.
- E. FABRIZI, A. FARCOMENI, AND V. GATTA (2012). Modelling work history patterns in the Italian labour market. *Statistical Methods & Applications*, **21**, 227–247.
- A. FARCOMENI (2012). Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Statistics and Computing*, **22**, 141–152.
- A. FARCOMENI AND L. GRECO (2015). *Robust Methods for Data Reduction*. Chapman & Hall/CRC Press.
- A. FARCOMENI AND L. VENTURA (2012). An overview of robust methods in medical research. *Statistical Methods in Medical Research*, **21**, 111–133.
- A. FARCOMENI AND S. VIVIANI (2015). Longitudinal quantile regression in presence of informative drop-out through longitudinal-survival joint modeling. *Statistics in Medicine*, **34**, 1199–1213.
- N. FENSKE, L. FAHRMEIR, T. HOTHORN, P. RZEHAK, AND M. HÖHLE (2013). Boosting structured additive quantile regression for longitudinal childhood obesity data. *International Journal of Biostatistics*, **9**, 1–18.
- J. FIEBERG, R. H. RIEGER, M. C. ZICUS, AND J. S. SCHILDCROUT (2009). Regression modelling of correlated data in ecology: subject-specific and population averaged response patterns. *Journal of Applied Ecology*, **46**, 1018–1025.
- G. M. FITZMAURICE, N. M. LAIRD, AND J. H. WARE (2004). *Applied longitudinal analysis*. Wiley-Interscience, Hoboken, NJ.
- L. FU AND Y.-G. WANG (2012). Quantile regression for longitudinal data with a working correlation model. *Computational Statistics and Data Analysis*, **56**, 2526–2538.
- A. F. GALVAO (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, **164**, **1**, 142–157.

- A. F. GALVAO AND G. V. MONTES-ROJAS (2010). Penalized quantile regression for dynamic panel data. *Journal of Statistical Planning and Inference*, **140**, 3476–3497.
- M. GERACI (2013). Estimation of regression quantiles in complex surveys with data missing at random: An application to birthweight determinants. *Statistical Methods in Medical Research*, available online.
- M. GERACI (2014). Linear quantile mixed models: The lqmm package for Laplace quantile regression. *Journal of Statistical Software*, **57**, 1–29.
- M. GERACI AND M. BOTTAI (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, **8**, **1**, 140–54.
- M. GERACI AND M. BOTTAI (2014). Linear quantile mixed models. *Statistics and Computing*, **24**, 461–479.
- M. GERACI AND A. FARCOMENI (2015). Probabilistic principal component analysis to identify profiles of physical activity behaviours in the presence of nonignorable missing data. *Journal of the Royal Statistical Society, Series C*, available online.
- M. GERACI AND C. JONES (2014). Prediction of conditional quantiles on the half line and the unit interval using transformation models. In: S. CABRAS, T. DI BATTISTA, AND W. RACUGNO, eds., *Proceedings of the 47th Scientific Meeting of the Italian Statistical Society*. Cooperativa Universitaria Editrice Cagliariitana, Cagliari, Italy.
- P. R. HAHN AND L. F. BURGETTE (2012). The mesa distribution: an approximation likelihood for simultaneous nonlinear quantile regression. *Tech. rep.*, University of Chicago.
- M. HALLIN, Z. LU, AND K. YU (2009). Local linear spatial quantile regression. *Bernoulli*, **15**, 659–686.
- M. HALLIN, D. PAINDAVEINE, AND M. SIMAN (2010). Multivariate quantiles and multiple-output regression quantiles: from  $l_1$  optimization to half-space depth. *Annals of Statistics*, **38**, 635–669.

- E. HAN AND L. M. POWELL (2013). Fast food prices and adult body weight outcomes: evidence based on longitudinal quantile regression models. *Contemporary Economic Policy*, **31**, 528–536.
- L. P. HANSEN (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 1029–1054.
- L. HAO AND D. Q. NAIMAN (2007). *Quantile regression*. Sage.
- M. HARDING AND C. LAMARCHE (2009). A quantile regression approach for estimating panel data models using instrumental variables. *Economics Letters*, **104**, **3**, 133 – 135.
- X. HE, B. FU, AND W. K. FUNG (2003). Median regression of longitudinal data. *Statistics in Medicine*, **22**, 3655–3669.
- S. HERITIER, E. CANTONI, S. COPT, AND M.P. VICTORIA-FESER (2009). *Robust Methods in Biostatistics*. Wiley, New York.
- J. L. HOROWITZ AND S. LEE (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, **100**, 1238–1249.
- P. J. HUBER (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In: *Fifth Berkely Sympos. Math. Statist. Probab.*, 221–223.
- S. JUNG (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association*, **91**, **433**, 251–257.
- A. KARLSSON (2008). Nonlinear quantile regression estimation of longitudinal data. *Communications in Statistics-Simulation and Computation*, **37**, **1**, 114–131.
- R. KOENKER (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**, 74–89.
- R. KOENKER (2005). *Quantile regression*. Cambridge university press.
- R. KOENKER AND G. BASSETT, JR. (1978). Regression quantiles. *Econometrica*, **46**, **1**, 33–50.

- R. KOENKER AND V. D'OREY (1987). Computing regression quantiles. *Applied Statistics*, **36**, 383–393.
- R. KOENKER AND V. D'OREY (1994). A remark on algorithm AS229: Computing dual regression quantiles and regression rank scores. *Applied Statistics*, **43**, 410–414.
- R. KOENKER AND K. HALLOCK (2001). Quantile regression. *Journal of Economic Perspectives*, **15**, 143–156.
- R. KOENKER AND J. A. F. MACHADO (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, **94**, 448, 1296–1310.
- R. KOENKER AND I. MIZERA (2004). Penalized trigrams: total variation regularization for bivariate smoothing. *Journal of the Royal Statistical Society, Series B*, **66**, 145–163.
- R. KOENKER, P. NG, AND S. PORTNOY (1994). Quantile smoothing splines. *Biometrika*, **81**, 673–680.
- P. KOSTOV (2009). A spatial quantile regression hedonic model of agricultural land prices. *Spatial Economic Analysis*, **4**, 1, 53–72.
- N. M. LAIRD AND J. H. WARE (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 4, 963–974.
- C LAMARCHE (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, **157**, 396–408.
- Y. LEE AND J. A. NELDER (2004). Conditional and marginal models: another view. *Statistical Science*, **19**, 219–238.
- C. LENG AND W. ZANG (2014). Smoothing combined estimating equations in quantile regression for longitudinal data. *Statistics and Computing*, **24**, 123–136.
- D. LI, L. PENG, AND J. YANG (2010). Bias reduction for high quantiles. *Journal of Statistical Planning and Inference*, **140**, 2433–2441.

- Y. LI, Y. LIU, AND J. ZHU (2007). Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association*, **102**, 255–268.
- K. Y. LIANG AND S. L. ZEGER (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.
- J. K. LINDSEY AND P. LAMBERT (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine*, **17**, 447–469.
- S. R. LIPSITZ, G. M. FITZMAURICE, G. MOLENBERGHS, AND L. P. ZHAO (1997). Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Journal of the Royal Statistical Society, Series C*, **46**, 463–476.
- R. J. A. LITTLE (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, **90**, 1112–1121.
- R. J. A. LITTLE AND D. B. RUBIN (2002). *Statistical analysis with missing data*. Wiley.
- Y. LIU AND M. BOTTAI (2009). Mixed-Effects Models for Conditional Quantiles with Longitudinal Data. *The International Journal of Biostatistics*, **5**, 1, 1–24.
- X. LU AND Z. FAN (2014). Weighted quantile regression for longitudinal data. *Computational Statistics*, 1–24.
- Z. LU, Q. TANG, AND L. CHENG (2014). Estimating spatial quantile regression with functional coefficients: A robust semiparametric framework. *arXiv:1402.0958*.
- K. LUM AND A. E. GELFAND (2012). Spatial quantile multiple regression using the asymmetric Laplace process. *Bayesian Analysis*, **7**, 235–258.
- Y. LUO AND M. LIAN, H. AND TIAN (2012). Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, **82**, 11, 1635–1649.

- J.A.F. MACHADO AND J. MATA (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, **20**, 445–465.
- M. F. MARINO, N. TZAVIDIS, AND M. ALFÓ (2015). Quantile regression for longitudinal data: unobserved heterogeneity and informative missingness. *arXiv:1501.02157v2*.
- D. P. McMILLEN (2013). *Quantile regression for spatial data*. Springer.
- Y. M. MU AND X. M. HE (2007). Power transformation toward a linear regression quantile. *Journal of the American Statistical Association*, **102**, 269–279.
- Y. M. MU AND Y. WEI (2009). A dynamic quantile regression transformation model for longitudinal data. *Statistica Sinica*, **19**, 1137–1153.
- J. M. NEUHAUS, J. D. KALBFLEISCH, AND W. W. HAUCK (1991). A comparison of cluster-specific and population-averaged approaches for correlated binary data. *International Statistical Review*, **59**, 25–35.
- J. M. NEUHAUS, C. E. MCCULLOCH, AND R. BOYLAN (2013). Estimation of covariate effects in generalized linear mixed models with a misspecified distribution of random intercept and slopes. *Statistics in Medicine*, **32**, 14, 2419–2429.
- W. K. NEWEY AND J. L. POWELL (1987). Asymmetric least squares estimation and testing. *Econometrica*, **55**, 819–847.
- M.I. PARZEN, L.J. WEI, AND Z. YING (1994). A resampling method based on pivotal estimating functions. *Biometrika*, **81**, 341–350.
- S. PORTNOY AND R. KOENKER (1997). The Gaussian hare and the Laplacean tortoise: Computability of squared-error vs absolute error estimators (with discussion). *Statistical Science*, **12**, 279–300.
- D. POWELL (2010). Unconditional quantile treatment effects in the presence of covariates. *Tech. rep.*, RAND working paper.
- B. J. REICH, H. D. BONDELL, AND H. J. WANG (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, **11**, 337–352.

- B. J. REICH, M. FUENTES, AND D. B. DUNSON (2011). Bayesian spatial quantile regression. *Journal of the American Statistical Association*, **106**, 6–20.
- D. RIZOPOULOS, G. VERBEKE, AND G. MOLENBERGHS (2008). Shared parameter models under random effects misspecification. *Biometrika*, **95**, 1, 63–74.
- A.M. ROSEN (2012). Set identification via quantile restrictions in short panels. *Journal of Econometrics*, **166**, 1, 127–137.
- G. SCHWARZ (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 2, 461–464.
- I. TAKEUCHI, Q. LE, T. SEARS, AND A. SMOLA (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*, **7**, 1231–1264.
- C. Y. TANG AND C. LENG (2011). Empirical likelihood and quantile regression in longitudinal data analysis. *Biometrika*, **89**, 1001–1006.
- S. T. TODKAR AND J. B. KADANE (2011). Simultaneous linear quantile regression: a semiparametric Bayesian approach. *Bayesian analysis*, **6**, 1–22.
- H.J. WANG AND M. FYGENSON (2009). Inference for censored quantile regression models in longitudinal studies. *Annals of Statistics*, **37**, 756–781.
- H.J. WANG, D. LI, AND X. HE (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, **107**, 1453–1464.
- H.J. WANG, Z. ZHU, AND J. ZHOU (2009). Quantile regression in partially linear varying coefficient models. *Annals of Statistics*, **37**, 3841–3866.
- R.W.M. WEDDERBURN (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 3, 439–447.
- Y. WEI AND X. HE (2006). Conditional growth charts. *Annals of Statistics*, **34**, 2069–2097.

- Y. WEI, A. PERE, R. KOENKER, AND X. HE (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, **25**, 1369–1382.
- R. E. WEISS (2005). *Modeling longitudinal data*. Springer Texts in Statistics. Springer, New York.
- M. C. WU AND R. J. CARROLL (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, **44**, **1**, 175–188.
- G. Y. YI AND W. HE (2009). Median regression models for longitudinal data with dropouts. *Biometrics*, **65**, 618–625.
- G. YIN AND J. CAI (2005). Quantile regression models with multivariate failure time data. *Biometrics*, **61**, 151–161.
- K. YU AND Z. LU (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics*, **31**, 333–346.
- K. YU, Z. LU, AND J. STANDER (2003). Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society, Series D*, **52**, 331–350.
- K. YU AND R. A. MOYEED (2001). Bayesian quantile regression. *Statistics & Probability Letters*, **54**, **4**, 437–447.
- Y. YUAN AND G. YIN (2010). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, **66**, **1**, 105–114.
- Y. R. YUE AND H. RUE (2011). Bayesian inference for additive mixed quantile regression models. *Computational Statistics and Data Analysis*, **55**, 84–96.
- Greg CG Wei and Martin A Tanner. A monte carlo implementation of the em algorithm and the poor man’s data augmentation algorithms. *Journal of the American statistical Association*, 85:699–704, 1990.
- S.L. ZEGER, K.Y. LIANG, AND P.S. ALBERT (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 1049–1060.