Functional Data Analysis of Generalized Quantile Regressions

Mengmeng Guo*
Lhan Zhou**
Jianhua Z. Huang**
Wolfgang Karl Härdle***

* Southwestern University of Finance and Economics, China
** Texas A&M University, U.S.A.
*** Humboldt-Universität zu Berlin, Germany

BERLIN

ECONOMIC RISK

SFB 649

Functional Data Analysis of Generalized Quantile Regressions

Mengmeng Guo [*], Lan Zhou [†], Jianhua Z. Huang [‡], Wolfgang Karl Härdle [§]

November 7, 2012

Abstract

Generalized quantile regressions, including the conditional quantiles and expectiles as special cases, are useful alternatives to the conditional means for characterizing a conditional distribution, especially when the interest lies in the tails. We develop a functional data analysis approach to jointly estimate a family of generalized quantile regressions. Our approach assumes that the generalized quantile regressions share some common features that can be summarized by a small number of principal component functions. The principal component functions are modeled as splines and are estimated by minimizing a penalized asymmetric loss measure. An iterative least asymmetrically weighted squares algorithm is developed for computation. While separate estimation of individual generalized quantile regressions usually suffers from large variability due to lack of sufficient data, by borrowing strength across data sets, our joint estimation approach significantly improves the estimation efficiency, which is demonstrated in a simulation study. The proposed method is applied to data from 150 weather stations in China to obtain the generalized quantile curves of the volatility of the temperature at these stations. These curves are needed to adjust temperature risk factors so that gaussianity is achieved. The normal distribution of temperature variations is vital for pricing weather derivatives with tools from mathematical finance.

KEY WORDS: Asymmetric loss function; Common structure; Functional data analysis; Generalized quantile curve; Iteratively reweighted least squares; Penalization.

JEL Classification: C13; C23; C38; Q54;

[*]Assistant Professor at Research Institute of Economics and Management of Southwestern University of Finance and Economics, China, Email: gmm0701@gmail.com.

[†]Assistant Professor at Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A., Email: lzhou@stat.tamu.edu.

[‡]Professor at Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A., Email: jianhua@stat.tamu.edu.

[§]Professor at Humboldt-Universität zu Berlin and Director of C.A.S.E - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099, Berlin, Germany, Email: haerdle@wiwi.hu-berlin.de.

1 Introduction

Conventional regression analysis is concerned about the conditional mean of a response given explanatory variables and focuses on the center of the conditional distribution. When the interest lies in the tails of the conditional distribution, the quantile regression (Koenker and Bassett, 1978), expectile regression (Newey and Powell, 1987), and M-quantiles (Breckling and Chambers, 1988) become useful tools. We refer to these tools broadly as generalized quantile regressions. The generalized quantile regression has found applications in many areas, such as financial markets, demographic studies, and weather analysis, especially for the statistical analysis of extreme events. Taylor (2008) applied generalized quantiles to calculate Value at Risk (VaR) and expected shortfall (ES) for financial risk management. Generalized quantiles were used by Schnabel and Eilers (2009a) to study the relationship between GDP and population, and by Härdle and Song (2010) to study the correlation of the wage and the level of education.

The specific application that motivates our work is the statistical modeling of weather risk (Anastasiadou and López-Cabrera, 2012). Extreme fluctuation of weather often causes great losses to industries that are weather related, such as the tourism and energy industries, which are temperature-dependent, and the agriculture industry, which are both temperature- and rainfall-dependent. To hedge the weather risk, an important financial instrument is the weather derivative. Statistical modeling and forecasting of the weather using historical data plays a crucial role in pricing weather derivatives (Odening et al., 2008). Guo and Härdle (2012) estimated the generalized quantiles of the volatility of temperatures as a function of time at a particular weather station and used them to identify temperature risk drivers. One problem with the generalized quantile regression is the high variability of the estimate at extreme quantile levels due to insufficiency of data at the tails of the distribution. The goal of this paper is to improve the estimation efficiency by borrowing strength across multiple data sets.

We consider the scenario that there is a need to estimate a collection of generalized quantile regressions each coming from a different data set. Our motivating example, detailed in Section 5, is concerned about estimating the generalized quantile of the distribution of the volatility of temperature as a function of time separately at multiple weather stations. Taking a functional data analysis approach (FDA; Ramsay and Silverman, 2005), we assume that the generalized quantile regressions under consideration share some common features that in turn can be summarized

by a small number of principal component functions. By pooling data sets together to estimate the principal component functions, we obtain more efficient estimates of individual generalized quantile regressions. In a related work but for a different context from ours, Cardot et al. (2005) considered a quantile regression where functional covariates are used to explain a scalar response variable.

More precisely, we assume that each generalized quantile regression (function) in the collection can be written as the summation of an overall mean function and a linear combination of several principal component functions. We model the mean and functional principal component functions as spline functions and use a roughness penalty to regularize the spline fit. Our estimation method makes use of the fact that the generalized quantile is the minimizer of an expected asymmetric loss function. By minimizing the corresponding empirical loss over spline coefficients, we obtain estimates of the mean and the principal component functions and consequently the generalized quantile regressions. We develop an iterative least asymmetrically weighted squares algorithm for computation. Our algorithm can be seen as an extension of previous functional PCA algorithms of James et al. (2000) and Zhou et al. (2008) for sparse functional data. As a result we obtain, PC funtions that tell us main sources of variations. The PC scores indicate us intra subject variation of the random curves and therefore allow us to price derivatives according to the locations of risk factors.

The rest of this paper is organized as follows. Section 2 reviews the formulation of generalized quantile regressions, their connections to asymmetric loss functions, and their estimation using penalized splines. Section 3 presents our FDA approach for estimating a collection of generalized quantile regressions; both the FDA model construction and the computational algorithm are discussed in detail. In Section 4, we use a simulation study to investigate the performance of our FDA-based joint estimation approach and compare it with the separate estimation approach. In Section 5, our method is applied to data from 150 weather stations in China to understand the risk drivers of the volatility of the temperature at these stations. Section 6 concludes the paper. The Appendix contains some technical details. The complete algorithm can be found on www.quantlet.org.

2 Generalized Quantile Regressions

Any random variable Y can be characterized by its cdf $F_Y(y) = P(Y \leq y)$, or equivalently, by its quantile function (qf)

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}, \qquad 0 < \tau < 1.$$

The τ-th quantile $Q_Y(\tau)$ minimizes the expected loss,

$$Q(\tau) = \arg\min_y \mathsf{E}\{\rho_\tau(Y - y)\}, \tag{1}$$

for the asymmetric loss function $\rho_\tau(Y - y)$ with

$$\rho_\tau(u) = u\{\tau - \mathbf{I}(u < 0)\}. \tag{2}$$

When Y is associated with a vector of covariates X, one is interested in studying the conditional (or regression) quantile $Q_{Y|X}(\tau|x) = F_{Y|X=x}^{-1}(\tau)$ as a function of x. Assuming linear dependence on covariates, the τ-th theoretical quantile regression is $Q_{Y|X}(\tau|x) = x^\top \beta^*$, where

$$\beta^* = \arg\min_\beta \mathsf{E}\{\rho_\tau(Y - X^\top \beta)|X = x\}. \tag{3}$$

Koenker and Bassett (1978) used this fact to define a minimum contrast estimator of regression quantiles. Since the loss function used in (1) and (3) can be interpreted as asymmetrically weighted absolute errors, it is natural to consider the asymmetrically weighted squared errors or other asymmetrically weighted loss functions. The expectile regressions of Newey and Powell (1987) are the solutions of the optimization problem (3) with the loss function corresponding to

$$\rho_\tau(u) = u^2|\tau - \mathbf{I}(u < 0)|.$$

More general asymmetric loss functions have been considered by Breckling and Chambers (1988) to define their M-quantiles which include quantiles and expectiles as special cases.

We now restrict our attention to a univariate covariate but consider the more flexible nonparametric estimation. For fixed τ, the τ-th *generalized quantile regression* is defined as

$$l_\tau(x) = \arg\min_\theta \mathsf{E}\{\rho_\tau(Y - \theta)|X = x\}, \tag{4}$$

4

where $\rho_\tau(Y - y)$ is an asymmetric loss function. Because it is a univariate function, $l_\tau(x)$ is also referred to as the τ-th generalized quantile curve. In this paper we focus on the quantile and expectile regressions, corresponding to

$$\rho_\tau(u) = |u|^\alpha |\tau - \mathbf{I}(u < 0)| \tag{5}$$

with $\alpha = 1, 2$, respectively, although with slight modifications our methodology is generally applicable for any $\alpha > 0$. According to Jones (1994), the expectiles can be interpreted as quantiles, not of the distribution $F(y|x)$ itself, but of a distribution related to $F(y|x)$. Specifically, write $H(y|x)$ for the conditional partial moment $\int_{-\infty}^y u\,F(du|x)$, and denote

$$G(y|x) = \frac{H(y|x) - yF(y|x)}{2\{H(y|x) - yF(y|x)\} + \{y - \mu(x)\}},$$

where $\mu(x) = H(\infty|x) = \int_{-\infty}^\infty u\,F(du|x)$ is the conditional mean function. The τ-th expectile of the conditional distribution $\mathcal{L}(Y|X = x)$ is the quantile of $G(y|x)$, that is, $l_\tau(x) = G^{-1}(\tau|x)$. When they are well-defined, both the conditional quantile and expectile characterize the conditional distribution, and there is a one-to-one mapping between them (Yao and Tong, 1996). Quantiles are intuitive, but expectiles are easier to compute and more efficient to estimate (Schnabel and Eilers, 2009b).

To estimate the generalized quantile regressions, assume we have paired data (X_i, Y_i), $i = 1, \ldots, n$, an i.i.d. sample from the joint distribution of (X, Y). It follows from (4) that the generalized quantile regression $l_\tau(\cdot)$ minimizes the unconditional expected loss,

$$l_\tau(\cdot) = \arg\min_{f \in \mathcal{F}} \mathsf{E}[\rho_\tau\{Y - f(X)\}], \tag{6}$$

where $\mathcal{F}$ is the collection of functions such that the expectation is well-defined. Using the method of penalized splines (Eilers and Marx, 1996; Ruppert et al., 2003), we represent $f(x) = b(x)^\top \gamma$, where $b(x) = \{b_1(x), \ldots, b_q(x)\}^\top$ is a vector of B-spline basis functions and γ is a q-vector of coefficients, and minimize the penalized average empirical loss,

$$\widehat{l}_\tau(\cdot) = \arg\min_{f(\cdot) = b(\cdot)^\top \gamma} \sum_{i=1}^n \rho_\tau\{Y_i - f(X_i)\} + \lambda\gamma^\top \Omega\gamma, \tag{7}$$

where Ω is a penalty matrix and λ is the penalty parameter. The penalty term is introduced to penalize the roughness of the fitted generalized quantile function $\widehat{l}_\tau(\cdot)$. When X_i's are evenly spaced, the penalty matrix Ω can be chosen such that $\gamma^\top \Omega\gamma = \sum_i (\gamma_{i+1} - 2\gamma_i + \gamma_{i-1})^2$ is the squared second difference penalty. In this case, $\Omega = D^\top D$ and D is the second-differential matrix such that $D\gamma$ creates the

vector of second differences $\gamma_{i+1} - 2\gamma_i + \gamma_{i-1}$. In general, the penalty matrix Ω can be chosen to be $\int \ddot{b}(x)\ddot{b}(x)^\top \, dx$ such that $\gamma^\top \Omega \gamma = \int \{\ddot{b}(x)^\top \gamma\}^2 \, dx$, where $\ddot{b}(x) = \{\ddot{b}_1(x), \ldots, \ddot{b}_q(x)\}^\top$ denotes the vector of second derivatives of the basis functions. The minimizing objective function in (7) can be viewed as the penalized negative log likelihood for the signal-plus-noise model

$$Y_i = l_\tau(X_i) + \varepsilon_i = b(X_i)^\top \gamma + \varepsilon_i, \tag{8}$$

where ε_i follows a distribution with a density proportional to $\exp\{-\rho_r(u)\}$, which corresponds respectively to the asymmetric Laplace distribution or the asymmetric Gaussian distribution for $\alpha = 1$ and $\alpha = 2$ (Koenker and Machado, 1999). Since these distributions are rather implausible for real-world data, their likelihood is better interpreted as a quasi-likelihood.

For expectiles ($\alpha = 2$ in the definition of loss function), Schnabel and Eilers (2009b) developed an iterative least asymmetrically weighted squares (LAWS) algorithm to solve the minimization problem (7), by extending an idea of Newey and Powell (1987). They rewrote the objective function in (7) as

$$\sum_{i=1}^{n} w_i(\tau)\{Y_i - b(X_i)^\top \gamma\}^2 + \lambda \gamma^\top \Omega \gamma, \tag{9}$$

where

$$w_i(\tau) = \begin{cases} \tau & \text{if } Y_i > b(X_i)^\top \gamma, \\ 1 - \tau & \text{if } Y_i \leq b(X_i)^\top \gamma. \end{cases} \tag{10}$$

For fixed weights $w_i(\tau)$'s, the minimizing $\widehat{\gamma}$ has a closed-form expression

$$\widehat{\gamma} = (B^\top W B + \lambda \Omega)^{-1} B^\top W Y, \tag{11}$$

where B is a matrix whose i-th row is $b(X_i)^\top$, W is the diagonal matrix whose ith diagonal entry is $w_i(\tau)$, and $Y = (Y_1, \ldots, Y_n)^\top$. Note that the weights $w_i(\tau)$'s depend on the spline coefficient vector γ. The LAWS algorithm iterates until convergence between computing (11) and updating W using (10) with γ being its current value obtained from (11).

With a slight modification, the LAWS algorithm can also be used to calculate the penalized spline estimator of conditional quantile functions, which correspond to $\alpha = 1$ in the asymmetric loss function. The weights for calculating the expectiles given in (10) need to be replaced by

$$w_i(\tau) = \begin{cases} \dfrac{\tau}{|Y_i - b(X_i)^\top \gamma| + \delta} & \text{if } Y_i > b(X_i)^\top \gamma, \\ \dfrac{1 - \tau}{|Y_i - b(X_i)^\top \gamma| + \delta} & \text{if } Y_i \leq b(X_i)^\top \gamma, \end{cases} \tag{12}$$

where $\delta > 0$ is a small constant used to avoid numerical problems when $Y_i - b(X_i)^\top \gamma$ is close to zero. In this case, the LAWS algorithm can be interpreted as a variant of Majorization-Minimization (MM) algorithm and the convergence of the LAWS algorithm then follows from the general convergence theory of the MM algorithm; see Hunter and Lange (2000).

One advantage of expectiles is that they can always be calculated no matter how low or high of the generalized quantile level τ, while the empirical quantiles can be undefined at extreme tails of the data distribution. It is also known that estimation of expectiles is usually more efficient than that of quantiles since it makes more effective use of data (Schnabel and Eilers, 2009b). However, when τ is close to 0 or 1, both estimation of expectiles and quantiles exhibits high variability, because of sparsity of data in the tails of the distribution. In the next section, we will present a method for better quantile and expectile estimation when there is a need to estimate a collection of generalized quantile regressions and, if these regressions share some common features. We use functional data analysis techniques to improve the estimation efficiency by borrowing strength across data sets.

3 Functional data analysis for a collection of regression quantiles

3.1 Approach

When we are interested in a collection of generalized quantile curves, denoted as $l_i(t)$, $i = 1, \ldots, N$, we may treat them as functional data. (To emphasize the one-dimensional nature of the covariate, from now on we change notation for the covariate from x to t.) Suppose $l_i(t)$'s are independent realizations of a stochastic process $l(t)$ defined on a compact interval $\mathcal{T}$ with the mean function $\mathsf{E}\{l(t)\} = \mu(t)$ and the covariance kernel $K(s,t) = \mathsf{Cov}\{l(s), l(t)\}, s, t \in \mathcal{T}$. If $\int_{\mathcal{T}} K(t,t)dt < \infty$, then Mercer's Lemma states that there exists an orthonormal sequence of eigenfunctions (ψ_j) and a non-increasing and non-negative sequence of eigenvalues (κ_j) such that

$$(K\psi_j)(s) \overset{\text{def}}{=} \int_{\mathcal{T}} K(s,t)\psi_j(t)dt = \kappa_j \psi_j(s),$$

$$K(s,t) = \sum_{j=1}^{\infty} \kappa_j \psi_j(s)\psi_j(t),$$

and

$$\sum_{j=1}^{\infty} \kappa_j = \int_{\mathcal{T}} K(t,t)dt.$$

Moreover, we have the following Karhunen-Loève expansion

$$l(t) = \mu(t) + \sum_{j=1}^{\infty} \sqrt{\kappa_j}\xi_j\psi_j(t), \tag{13}$$

where $\xi_j \overset{\text{def}}{=} \frac{1}{\sqrt{\kappa_j}} \int l(t)\psi_j(s)ds$, $\mathsf{E}(\xi_j) = 0$, $\mathsf{E}(\xi_j\xi_k) = \delta_{j,k}$, $j,k \in \mathbb{N}$, and $\delta_{j,k}$ is the Kronecker delta.

Usually statistical estimation demands a parsimonious model for estimation efficiency and thus the terms associated with small eigenvalues in (13) can be neglected. As a result for $i = 1,\cdots,n$ observations of l and therefore ξ_j, we obtain the following reduced-rank model:

$$l_i(t) = \mu(t) + \sum_{k=1}^{K} f_k(t)\alpha_{ik} = \mu(t) + f(t)^{\top}\alpha_i, \tag{14}$$

where $f(t) = \{f_1(t),\cdots,f_K(t)\}^{\top}$ and K is a fixed integer. In practice, K can be chosen by cross validation (CV). As in (13) and (James et al., 2000; Zhou et al., 2008), μ is the mean function, f_k the k-th principal component function (PC) and $\alpha_i = (\alpha_{i1},\cdots,\alpha_{iK})^{\top}$ the vector of PC scores for the i-th curve. α_{ij} corresponds to $\kappa_{ij}\xi_{ij}$ in (13). Since the approximations (14) share the same mean function and the same set of principal components for the collection of generalized quantile curves, it enables to borrow information across data sets to improve estimation efficiency.

Accepting the parameterizations in (14), estimation of the generalized quantile curves is reduced to the estimation of the mean and principal components functions. Using the method of penalized splines again, we represent these functions in the form of basis expansions

$$\begin{aligned} \mu(t) &= b(t)^{\top}\theta_{\mu}, \\ f(t)^{\top} &= b(t)^{\top}\Theta_f, \end{aligned} \tag{15}$$

where $b(t) = \{b_1(t),\cdots,b_q(t)\}^{\top}$ is a q-vector of B-splines, θ_{μ} is a q-vector and $\Theta_f = \{\theta_{f,1},\cdots,\theta_{f,K}\}^{\top}$ is a $q \times K$ matrix of spline coefficients. The B-splines are normalized so that

$$\int b(t)b(t)^{\top}dt = \mathbf{I}_q.$$

Thus the estimation problem is further reduced to the estimation of spline coefficients. For identifiability, we impose the following restriction

$$\Theta_f^\top \Theta_f = \mathbf{I}_K.$$

The above two equations imply the usual orthogonality requirements of the principal component curves:

$$\int f(t)f(t)^\top dt = \Theta_f^\top \int b(t)b(t)^\top dt \Theta_f = \mathbf{I}_K.$$

Denote the observations as $\{Y_{ij}\}$ with $i = 1, \cdots, N$, $j = 1, \cdots, T_i$. Combining (14) and (15) yields the following representation

$$l_{ij} \stackrel{\text{def}}{=} l_i(t_{ij}) = b(t_{ij})^\top \theta_\mu + b(t_{ij})^\top \Theta_f \alpha_i. \tag{16}$$

Here, the scores α_i's are treated as fixed effects instead of random effects for convenience in applying the asymmetric loss minimization, see the last paragraph of this section for more information. For identifiability, we assume that $\sum_{i=1}^N \alpha_{ik} = 0$, $1 \leq k \leq K$, and $\sum_{i=1}^N \alpha_{i1}^2 > \cdots > \sum_{i=1}^N \alpha_{iK}^2$. The empirical loss function for generalized quantile estimation is

$$S = \sum_{i=1}^N \sum_{j=1}^{T_i} \rho_\tau \{Y_{ij} - b(t_{ij})^\top \theta_\mu - b(t_{ij})^\top \Theta_f \alpha_i\}, \tag{17}$$

where $\rho_\tau(u)$ is the asymmetric loss function defined in (5). To ensure the smoothness of the estimates of the mean curve and the principal components curves, we use a moderate number of knots and apply a roughness penalty to regularize the fitted curves. The squared second derivative penalties for the mean and principal components curves are given by

$$M_\mu = \theta_\mu^\top \int \ddot{b}(t)\ddot{b}(t)^\top \, dt \, \theta_\mu = \theta_\mu^\top \Omega \, \theta_\mu,$$

$$M_f = \sum_{k=1}^K \theta_{f,k}^\top \int \ddot{b}(t)\ddot{b}(t)^\top \, dt \, \theta_{f,k} = \sum_{k=1}^K \theta_{f,k}^\top \Omega \, \theta_{f,k}.$$

The penalized empirical loss function is then

$$S^* = S + \lambda_\mu M_\mu + \lambda_f M_f, \tag{18}$$

where λ_μ and λ_f are nonnegative penalty parameters. Note that we use the same penalty parameter for all principal components curves for the sake of simplicity,

9

similar strategy has been used in Zhou et al. (2008). We propose to minimize the penalized loss (18) to estimate the parameters θ_μ, Θ_f, and α_i's. The choice of the penalty parameters will be discussed later in the paper.

Define the vector $L_i = \{l_{i1}, \cdots, l_{iT_i}\}^\top$ and the matrix $B_i = \{b(t_{i1}), \cdots, b(t_{iT_i})\}^\top$. The representation (16) can be written in matrix form

$$L_i = B_i\theta_\mu + B_i\Theta_f\alpha_i \tag{19}$$

Writing $Y_i = (Y_{i1}, \ldots, Y_{iT_i})^\top$, the data have the following signal-plus-noise representation

$$Y_i = L_i + \varepsilon_i = B_i\theta_\mu + B_i\Theta_f\alpha_i + \varepsilon_i \tag{20}$$

where ε_i is the random error vector whose components follow some asymmetric distribution as in (8), corresponding to the asymmetric loss minimization for the generalized quantile regression. Equation (20) has also been used in Zhou et al. (2008) for a random effects model of functional principal components, where both α_i and ε_i are multivariate normally distributed. Since the signal-plus-noise model (20) for generalized quantile regression is not a plausible data generating model but rather an equivalent representation of the asymmetric loss minimization, the EM-algorithm used in Zhou et al. (2008) can not be simply extended and justified in the current context.

3.2 Algorithm

This subsection develops an iterative penalized least asymmetrically weighted squares (PLAWS) algorithm for minimizing the penalized loss function defined in (18), by defining weights in a similar manner as in (10) and (12).

We fix the quantile level $\tau \in (0, 1)$. To estimate the expectile curves, for $i = 1, \cdots, N$ and $j = 1, \cdots, T_i$, define the weights

$$w_{ij} = \begin{cases} \tau & \text{if } Y_{ij} > l_{ij}, \\ 1 - \tau & \text{if } Y_{ij} \leq l_{ij}. \end{cases} \tag{21}$$

where $l_{ij} = b(t_{ij})^\top\theta_\mu - b(t_{ij})^\top\Theta_f\alpha_i$ is a function of the parameters. To estimate the quantile curves, define the weights

$$w_{ij} = \begin{cases} \dfrac{\tau}{|Y_{ij} - l_{ij}| + \delta} & \text{if } Y_{ij} > l_{ij}, \\ \dfrac{1 - \tau}{|Y_{ij} - l_{ij}| + \delta} & \text{if } Y_{ij} \leq l_{ij}, \end{cases} \tag{22}$$

10

where l_{ij} is defined as in (21) and δ is a small positive constant. Using these weights, the asymmetric loss function in (17) can be written as the following asymmetrically weighted sum of squares

$$S = \sum_{i=1}^{N} \sum_{j=1}^{T_i} w_{ij} \{Y_{ij} - b(t_{ij})^\top \theta_\mu - b(t_{ij})^\top \Theta_f \alpha_i\}^2, \tag{23}$$

and the penalized loss function (18) becomes the following penalized weighted sum of squares criterion

$$S^* = \sum_{i=1}^{N} (Y_i - B_i \theta_\mu - B_i \Theta_f \alpha_i)^\top W_i (Y_i - B_i \theta_\mu - B_i \Theta_f \alpha_i)$$
$$+ \lambda_\mu \theta_\mu^\top \Omega\, \theta_\mu + \lambda_f \sum_{k=1}^{K} \theta_{f,k} \Omega\, \theta_{f,k}, \tag{24}$$

where $W_i = \text{diag}\{w_{i1}, \ldots, w_{iT_i}\}$. Since the weights depend on the parameters, the PLAWS algorithm iterates until convergence between minimizing (24) and updating the weights using (21) and (22).

To minimize (24) for fixed weights, we alternate minimization with respect to θ_μ, Θ_f, and α_i. Such minimizations have closed-form solutions

$$\widehat{\theta}_\mu = \left\{ \sum_{i=1}^{N} B_i^\top W_i B_i + \lambda_\mu \Omega \right\}^{-1} \left\{ \sum_{i=1}^{N} B_i^\top W_i (Y_i - B_i \widehat{\Theta}_f \widehat{\alpha}_i) \right\}, \tag{25}$$

$$\widehat{\theta}_{f,l} = \left\{ \sum_{i=1}^{N} \widehat{\alpha}_{il}^2 B_i^\top W_i B_i + \lambda_f \Omega \right\}^{-1} \left\{ \sum_{i=1}^{N} \widehat{\alpha}_{il} B_i^\top W_i (Y_i - B_i \widehat{\theta}_\mu - B_i Q_{il}) \right\},$$

$$\widehat{\alpha}_i = (\widehat{\Theta}_f^\top B_i^\top W_i B_i \widehat{\Theta}_f)^{-1} \left\{ \widehat{\Theta}_f^\top B_i^\top W_i (Y_i - B_i \widehat{\theta}_\mu) \right\},$$

where

$$Q_{il} = \sum_{k \neq l} \widehat{\theta}_{f,k} \widehat{\alpha}_{ik},$$

and $i = 1, \cdots, N$, $k, l = 1, \cdots, K$, $\widehat{\theta}_{f,k}$ is the k-th column of $\widehat{\Theta}_f$.

A summary of the complete algorithm is presented in Appendix A.1. A procedure for obtaining initial values is given in Appendix A.2. After we get the parameter estimates from the PLAWS algorithm, we can estimate the individual quantile curves by plugging the parameter estimates into (14) and (15).

3.3 Choice of Auxiliary Parameters

In this paper, for simplicity, we use equally spaced knots for the B-splines. The choice of the number of knots to be used is not critical, as long as it is moderately large, since the smoothness of the fitted curves is mainly controlled by the roughness penalty term. For typical sparse functional datasets, 10-20 knots are often sufficient, see Zhou et al. (2008). The optimal choice of the penalty parameter for the single curve estimation used in initialization follows the method in Schnabel and Eilers (2009b). There are several well developed methods for choosing the auxiliary parameters in the FDA framework, such as, AIC, BIC and cross-validation (CV), Ramsay and Silverman (2005). In this paper, all the auxiliary parameters, such as the number of principal components/factors to be included, and the penalty parameters λ_μ and λ_f, will be chosen via the 5-fold cross-validation by minimizing the cross-validated asymmetric loss function. Explicitly, the 5-fold cross-validation can be written as:

$$CV(K, \lambda_\mu, \lambda_f) = \frac{1}{5} \sum_{i=N-(m-1)\times5}^{N-m\times5} \sum_{j=1}^{T_i} \widehat{w}_{ij} |Y_{ij} - \widehat{l}_{ij}|^2$$

where $m = 1, 2, \cdots, [N/5]$, $\widehat{w}_{ij} = w_{ij}(Y_{ij} - \widehat{l}_{ij})$ defined in (10) and (12)

$$\widehat{l}_{ij} = b(t_{ij})^\top \widehat{\theta}_\mu + b(t_{ij})^\top \widehat{\Theta}_f \widehat{\alpha}_i$$

.

4 Simulation

We conducted a simulation study to illustrate the proposed FDA approach in estimating a collection of generalized quantile curves. For each case, we considered N data sets. For the i-th data set, the data were generated from the model

$$Y_{ij} = \mu(t_j) + f_1(t_j)\alpha_{1i} + f_2(t_j)\alpha_{2i} + \varepsilon_{ij}, \qquad j = 1, \ldots, T, \tag{26}$$

where t_j's are sampling points equidistant in $[0, 1]$ with $t_j = j/T$, the mean curve $\mu(t) = 1 + t + \exp\{-(t - 0.6)^2/0.05\}$, the principal component curves are $f_1(t) = \sqrt{2}\sin(2\pi t)$ and $f_2(t) = \sqrt{2}\cos(2\pi t)$, and $\varepsilon_{ij} = \varepsilon_i(t)$ are independent errors. The scores α_{1i} and α_{2i} were generated independently from $N(0, 36)$ and $N(0, 9)$ distribution, respectively. The errors ε_{ij} were generated from either (1) $N(0, 0.5)$, (2)

$N(0, \mu(t) \times 0.5)$ or (3) $t(5)$ distributions. The τ-th quantile or expectile curve for the i-th data set is

$$l_i(t) = \mu(t) + f_1(t)\alpha_{1i} + f_2(t)\alpha_{2i} + c_\tau,$$

where c_τ represents the corresponding τ-th theoretical quantile or expectile of $\varepsilon_i(t)$. We considered two setups of sample sizes: (1) $N = 20$ data sets with $T = 100$ observation points for each set and (2) $N = 40$ data sets with $T = 150$ observation points for each set. The code for simulation may be found in www.quantlet.de.

We ran the simulation 200 times for each setup. We applied both the proposed FDA method and the separate estimation method to each simulation to estimate the 95% expectile and quantile curves. For simplicity, we assume there are two principal components, i.e. $K = 2$, and the penalty parameters are chosen by 5-fold cross validation. We calculated the integrated squared error for estimating individual generalized quantile curves. These errors were then averaged over data sets to obtain the mean integrated squared errors (MSEs). The summary statistics (mean and SD) of the MSEs are reported in Table 1, where the same quantities for the separate estimation approach are also reported. We observe that the proposed FDA method outperforms the separate estimation approach in all scenarios by producing smaller MSEs. We also observe that, for each setup, the MSEs for estimating the expectiles are smaller than those for estimating the quantiles; this is consistent to our earlier discussion that expectile estimates are less variable than quantile estimates. Moreover, comparing results for Scenario 1 and Scenario 3, we see that the MSEs are bigger when the distribution has fatter tails.

Figure 1 shows the estimated results of the 95% expectile curves by the proposed FDA approach with the error terms normally distributed. One observes that the mean of the estimated mean curves fit the respective true ones well, and the confidence intervals cover the real ones. We also notice that the results from the large data set fit slightly better to that from the small sample size, especially for the two functional principal component curves, since their confidence intervals become smaller as the sample size increases. However, the estimated mean curve, due to bias, is slightly underestimated at some points. Further, we observe that the large dataset gives us narrower confidence interval that the small one. Figure 2 shows the estimated mean curves and principal component curves of 95% quantile curves when the error terms are still normally distributed. The quantile curves perform similarly to that of the expectile curves. While, the results from the mean curves are slightly better than that of expectile curves, to say, the confidence intervals actually cover

13

Scenario	Sample Size	Expectile curves		Quantile curves	
		FDA	Separate	FDA	Separate
1	$N = 20, T = 100$	0.0815	0.1407	0.1733	0.2539
		(0.0296)	(0.0149)	(0.0283)	(0.0227)
	$N = 40, T = 150$	0.0189	0.0709	0.0723	0.1875
		(0.0025)	(0.0052)	(0.1205)	(0.0127)
2	$N = 20, T = 100$	0.1436	0.3188	0.2769	0.8039
		(0.0248)	(0.0339)	(0.1061)	(0.0860)
	$N = 40, T = 150$	0.0931	0.2751	0.1785	0.0.6029
		(0.0106)	(0.0188)	(0.0813)	(0.0503)
3	$N = 20, T = 100$	0.2859	0.5194	0.4490	1.2227
		(0.0525)	(0.1284)	(0.2867)	(0.2290)
	$N = 40, T = 150$	0.1531	0.4087	0.2340	0.8683
		(0.0212)	(0.0707)	(0.1259)	(0.1085)

Table 1: The summary statistics (mean and SD) of the MSEs for estimating 95% generalized quantile curves by the FDA approach and the separate estimation approach. Scenario 1 with $\varepsilon_{ij} \sim N(0, 0.5)$, Scenario 2 with $\varepsilon_{ij} \sim N(0, \mu(t) \times 0.5)$ and Scenario 3 with $\varepsilon_{ij} \sim t(5)$.

the real mean curves well. In general, one can say that though the performance of the proposed FDA method for quantile regression is comparable to the result of expectile curves.

5 Application

Temperature derivatives are financial instruments that provide protection and investment opportunities contributed on weather events. Understanding of the risk factors of temperature is crucial to the pricing of temperature derivatives. In this section, we apply the proposed FDA method to study the variation of temperature at 150 weather stations in China using daily average temperature data of year 2010 obtained from the Chinese Meteorological Administration, which was obtained from Research Data Center (RDC) in Humboldt University at Berlin. The locations of the weather stations are shown in Figure 3.

The temperature record has a clear seasonable pattern — low in winter and high in summer — and also displays strong autocorrelation. We studied the volatility of the temperature using the residuals after de-trending and removing the autoregressive effect as well as the seasonal effect. The standard procedure (for pricing) is
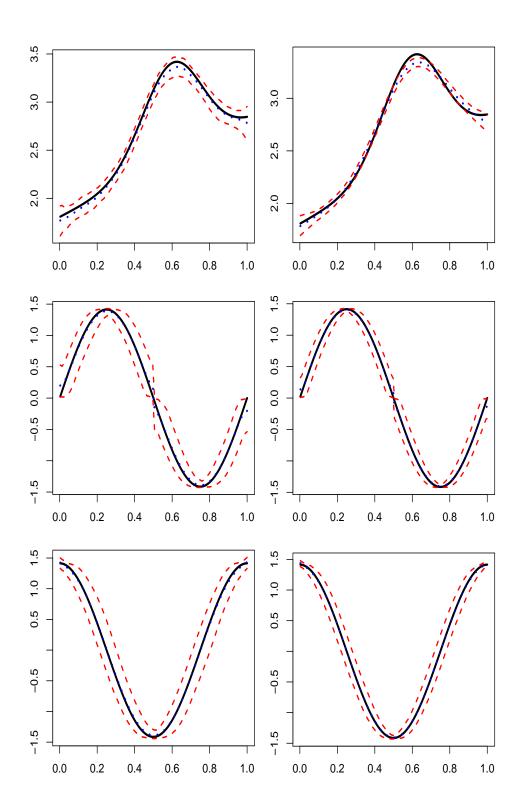
Figure 1: The estimated μ (blue dotted), the real μ (black solid) and the $5\% - 95\%$ pointwise confidence intervals (red dashed), Upper Panel; the estimated first principal component f_1, Middle Panel; the estimated second principal component f_2, Bottom Panel; for 95% expectile curves when the error term is normally distributed with mean 0 and variance 0.5. The sample size are respectively $N = 20, M = 100$ (Left) and $N = 40, M = 150$ (Right).
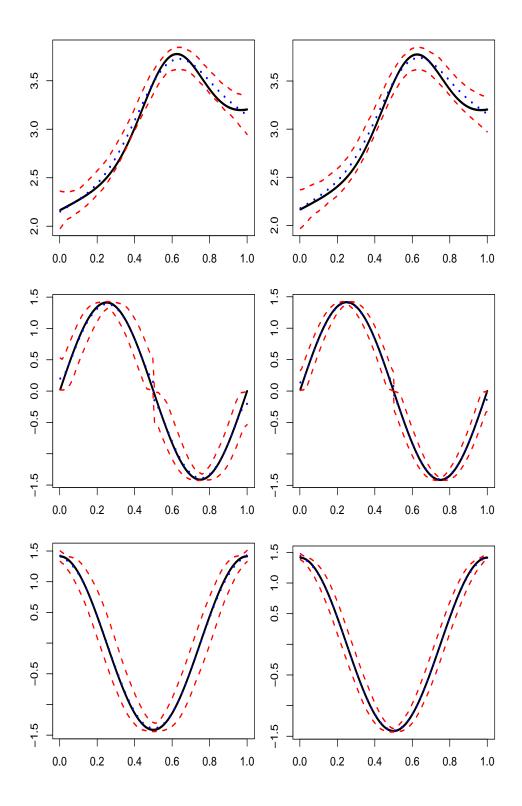
15

Figure 2: The estimated μ (dotted blue), the real μ (solid black) and the $5\% -$ 95% pointwise confidence intervals (red dashed), Upper Panel; the estimated first principal component f_1, Middle Panel; the estimated second principal component f_2, Bottom Panel; for 95% quantile curves with error term normally distributed with mean 0 and variance 0.5. The sample size is $N = 20, M = 100$ (Left) and $N = 40, M = 150$ (Right).
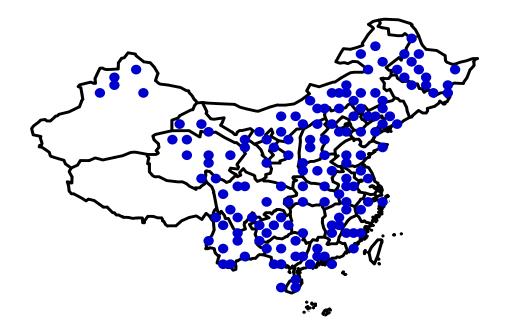
Figure 3: 150 Weather Stations in China

well-documented in the literature (Campbell and Diebold, 2005; Härdle and López-Cabrera, 2011). Let T_{it} denote the average temperature on day t for city (station) i. The standard model described e.g. in Benth et al. (2007) is:

$$
\begin{aligned}
T_{it} &= X_{it} + \Lambda_{it}, \\
\Lambda_{it} &= a_i + b_i t + \sum_{m=1}^{M} c_{im} \cos\left\{\frac{2\pi(t - d_{im})}{m \cdot 365}\right\}, \\
X_{it} &= \sum_{j=1}^{p_i} \beta_{ij} X_{i,t-j} + \varepsilon_{it}.
\end{aligned}
\tag{27}
$$

The seasonal effect Λ_{it} is captured by a small number of Fourier terms, and autocorrelation by an autoregressive AR structure. Our interest is the collection of expectile curves of different percentages at each station i which characterize the distribution of ε_{it} as a function of t. We fit model (27) to the temperature data and obtained estimated residuals ε_{it}. In principle, the distribution function of the volatility can be deduced from the generalized quantile curves. Further, the distribution function is crucial to pricing the weather derivatives, more description can be found in Härdle and López-Cabrera (2011). We applied our FDA method to these residuals to estimate the 5%, 25%, 75%, and 95% expectile curves for each weather station. In each application of our method, penalty parameters were selected using cross-validation.

Evaluating the empirical variance of the estimated PC scores suggest that,

| | Expectile levels | | | |
PC index	5%	25%	75%	95%
1	0.3833	0.0596	0.0659	0.4421
2	0.0665	0.0131	0.0194	0.1102
3	0.0471	0.0077	0.0158	0.0746
4	0.0415	0.0074	0.0123	0.0657
5	0.0306	0.0072	0.0056	0.0455
6	0.0262	0.0051	0.0050	0.0226

Table 2: The empirical variances of PC scores for the Chinese temperature data.

for all four expectile levels, the first principal component is a dominating factor in explaining the variability among the weather stations; see Table 2. Figure 4 shows the estimated first principal component functions $f_1(t)$ for four expectile levels. These PC functions have the following interpretation: A positive score on the first PC of the 5% and 25% expectiles implies that the corresponding distribution has a lighter than average left tail, while a positive score on the first PC of the 75% and 95% expectiles implies that the corresponding distribution has a heavier than average right tail. The U shape of the PC functions suggest that the effect is stronger in winter than in summer.

Figure 5 shows the estimated PC scores α_{1i} for the first principal components at four expectile levels. To help the interpretation, the values of the scores are shown as colored dots at the locations of the stations on the map of China. For expectiles at lower levels, i.e., 5% and 25% levels, the weather stations in northern China tend to have positive PC scores, while those in the south are opposite; for expectiles at higher levels, i.e., 75% and 95% levels, the weather stations in northern China tend to have negative PC scores, while those in the south are opposite. According to the interpretation of the first principal components given earlier, these results suggest that the temperature distribution has heavier left and right tails (and is so more spread out) in southern China than that in the north, and this phenomenon is more pronounced in winter than in summer. Therefore, one can say that it has more potential to buy weather derivatives to hedge the corresponding risk in the south of China, especially to hedge the temperature risk in winter. We may understand the result as that in winter, the north part of China already has the perfect heating system, therefore even the big changes in temperature cannot influence the residents, the energy companies or other related industries. While, in the south, the weather related sectors, such as the crops and energy companies are more sensitive to the
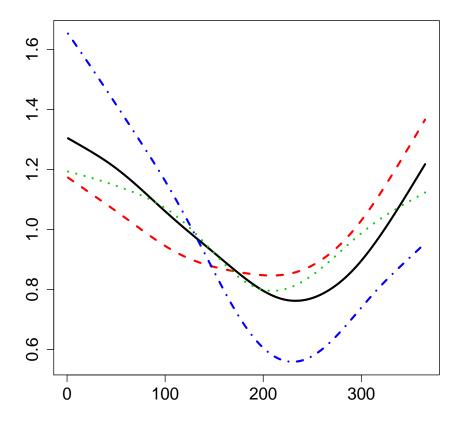
Eigenfunctions

Figure 4: The estimated first principal component for the 5% (black solid), 25% (red dashed), 75% (green dotted), 95% (blue dash-dotted) expectiles curves of the volatility of the temperature of China in 2010 with the data from 150 weather stations.

variation of temperature. Extreme cold weather in southern China even may kill people. Thus, weather derivatives are necessary tools to hedge temperature risk and avoid the corresponding loss, especially in the Southern China.

6   Conclusion

This paper develops an approach for jointly estimating a family of generalized quantile curves. By applying ideas from functional data analysis, we can borrow strength across populations. The simulation study demonstrates the proposed FDA approach is more efficient than separate estimation. Our method also provides principal component functions for the generalized quantile curves, which is useful for describing the major source of variation among these curves. The application to temperature data yielded scores that gave message into the distribution of tail events of
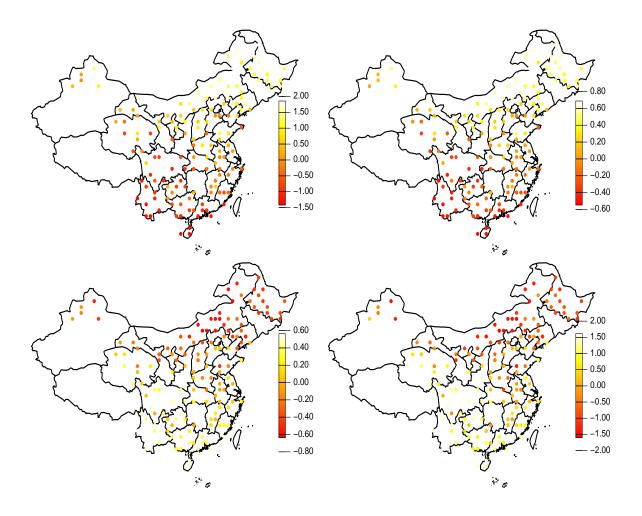
Figure 5: The estimated first principal component scores α_1 for the 5%, 25%, 75% and 95% expectile curves of the temperature distribution.

temperature in China.

A Appendix

A.1 The complete PLAWS algorithm

We give the complete algorithm in this appendix. The parameters that appear on the right hand side of the equations are all fixed at the values from the last iteration.

a. Initialization the algorithm using the procedure described in Appendix A.2.

b. Update $\widehat{\theta}_\mu$ using

$$\widehat{\theta}_\mu = \left\{ \sum_{i=1}^N B_i^\top \widehat{W}_i B_i + \lambda_\mu \Omega \right\}^{-1} \left\{ \sum_{i=1}^N B_i^\top \widehat{W}_i (Y_i - B_i \widehat{\Theta}_f \widehat{\alpha}_i) \right\}.$$

c. For $l = 1, \cdots, K$, update the l-th column of $\widehat{\Theta}_f$ using

$$\widehat{\theta}_{f,l} = \left\{ \sum_{i=1}^N \widehat{\alpha}_{il}^2 B_i^\top \widehat{W}_i B_i + \lambda_f \Omega \right\}^{-1} \left\{ \sum_{i=1}^N \widehat{\alpha}_{il} B_i^\top \widehat{W}_i (Y_i - B_i \widehat{\theta}_\mu - B_i Q_{il}) \right\},$$

where $\widehat{\theta}_{f,k}$ is the k-th column of $\widehat{\Theta}_f$, and

$$Q_{il} = \sum_{k \neq l} \widehat{\theta}_{f,k} \widehat{\alpha}_{ik}, \quad i = 1, \cdots, N.$$

d. Use the QR decomposition to orthonormalize the columns of $\widehat{\Theta}_f$.

e. Update $(\widehat{\alpha}_1, \ldots, \widehat{\alpha}_N)$ using

$$\widehat{\alpha}_i = (\widehat{\Theta}_f^\top B_i^\top W_i B_i \widehat{\Theta}_f)^{-1} \left\{ \widehat{\Theta}_f^\top B_i^\top W_i (Y_i - B_i \widehat{\theta}_\mu) \right\},$$

and then center $\widehat{\alpha}_i$ such that $\sum_i^N \widehat{\alpha}_i = 0$.

f. Update the weights, defined in (21) for expectiles and (22) for quantiles.

g. Iterate Steps b-f until convergence is reached.

A.2 Initial values for the PLAWS algorithm

The initialization of the PLAWS algorithm is:

a. Estimate N expectile/quantile curves $\widehat{l}_i(t)$ separately by applying the single curve estimation algorithm described in Section 2.

b. Set $\widehat{L}_i = \{\widehat{l}(t_{i1}), \ldots, \widehat{l}(t_{iT_i})\}^\top$. Run the linear regression

$$\widehat{L}_i = B_i \theta_\mu + \varepsilon_i, \qquad i = 1, \ldots, N, \tag{28}$$

to get the initials of $\widehat{\theta}_\mu$ as follows

$$\widehat{\theta}_\mu^{(0)} = \left(\sum_{i=1}^N B_i^\top B_i \right)^{-1} \left(\sum_{i=1}^N B_i^\top \widehat{L}_i \right).$$

c. Calculate the residuals of the regression (28), denoted as $\widetilde{L}_i = \widehat{L}_i - B_i \widehat{\theta}_\mu^{(0)}$. For each i, run the following linear regression

$$\widetilde{L}_i = B_i \Gamma_i + \varepsilon_i.$$

The solution, denoted as $\widehat{\Gamma}_i^{(0)}$, is used in later steps for finding initials of Θ_f and α_i. Set $\widehat{\Gamma}_0 = (\widehat{\Gamma}_1^{(0)}, \cdots, \widehat{\Gamma}_N^{(0)})$.

d. Calculate the singular value decomposition of $\widehat{\Gamma}^{(0)\top}$:

$$\widehat{\Gamma}^{(0)\top} = UDV^\top.$$

The initial of Θ_f is chosen as $\widehat{\Theta}_f^{(0)} = V_k D_k$ where V_k consists of the first K columns of V and D_k is the corresponding $K \times K$ block of D.

e. Run the following regression

$$\widehat{\Gamma}_i^{(0)} = \widehat{\Theta}_f \widehat{\alpha}_i + \varepsilon_i \tag{29}$$

to get the initials of $\widehat{\alpha}_i$; use a ridge penalty if the regression is singular. Center $\widehat{\alpha}_i$'s such that $\sum_{i=1}^N \widehat{\alpha}_i = 0$.

Acknowledgment

References

Anastasiadou, Z. and López-Cabrera, B. (2012). Statistical modelling of temperature risk. *SFB discussing paper 2012-029, submitted to Statistical Science.*

Benth, F., Benth, J., and Koekebakker, S. (2007). Putting a price on temperature. *Scandinavian Journal of Statistics,* 34(4):746–767.

Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika,* 74(4):761–772.

Campbell, S. and Diebold, F. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association,* 469:6–16.

Cardot, H., Crambes, C., and Sarda, P. (2005). Quantile regression when the covariates are funtions. *Nonparametric Statistics,* 17(7):841–856.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Journal of the American Statistical Association,* 11:89–121.

Guo, M. and Härdle, W. (2012). Simultaneous confidence bands for expectile functions. *AStA Advances in Statistical Analysis,* 96:517–542.

Härdle, W. and López-Cabrera, B. (2011). Implied market price of weather risk. *Applied Mathematical Finance,* 5:1–37.

Härdle, W. and Song, S. (2010). Confidence bands in quantile regression. *Econometric Theory,* 26(4):1180–1200.

Hunter, D. and Lange, K. (2000). Quantile regression via an MM algorithm. *Journal of Computational and Graphical Statistics,* 9(1):60–77.

James, G., Hastie, T., and Sugar, C. (2000). Principle component models for sparse functinal data. *Biometrika,* 87:587–602.

Jones, M. (1994). Expectiles and M-quantiles are quantiles. *Statistics and Probability Letters,* 20:149–153.

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

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

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847.

Odening, M., Berg, E., and Turvey, C. (2008). Management of climate risk in agriculture. *Special Issue of the Agricultural Finance Review*, 68(1):83–97.

Ramsay, J. and Silverman, B. (2005). *Functional data analysis, 2nd ed.* New York: Springer.

Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric Regression*. Cambridge University Press.

Schnabel, S. and Eilers, P. (2009a). An analysis of life expectancy and economic production using expectile frontier zones. *Demographic Research*, 21:109–134.

Schnabel, S. and Eilers, P. (2009b). Optimal expectile smoothing. *Computational Statistics and Data Analysis*, 53:4168–4177.

Taylor, J. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6:231–252.

Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: a nonparametric approach. *Journal of Nonparametric Statistics*, 6(2-3):273–292.

Zhou, L., Huang, J., and Carroll, R. (2008). Joint modelling of paired sparse functional data using principle components. *Biometrika*, 95(3):601–619.

SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001   "Functional Data Analysis of Generalized Quantile Regressions" by Mengmeng Guo, Lhan Zhou, Jianhua Z. Huang and Wolfgang Karl Härdle, January 2013.