

# CONDITIONAL QUANTILE ANALYSIS WHEN COVARIATES ARE FUNCTIONS, WITH APPLICATION TO GROWTH DATA

Kehui Chen<sup>1,2</sup> and Hans-Georg Müller<sup>1,3</sup>

Revised Version, May 2011

**SUMMARY.** Motivated by the conditional growth charts problem, we develop a method for conditional quantile analysis when predictors take values in a functional space. The proposed method aims at estimating conditional distribution functions under a generalized functional regression framework. This approach facilitates balancing of model flexibility and the curse of dimensionality for the infinite-dimensional functional predictors. Its good performance in comparison with other methods, both for sparsely and densely observed functional covariates, is demonstrated through theory as well as in simulations and an application to growth curves, where the proposed method can, for example, be used to assess the entire growth pattern of a child by relating it to the predicted quantiles of adult height.

**KEY WORDS:** Functional Data Analysis, Quantile Analysis, Conditional Distribution Function, Conditional Reference Chart, Growth Data Analysis, Uniform Weak Convergence.

---

<sup>1</sup> Department of Statistics, University of California, Davis, USA

<sup>2</sup> Corresponding author, e-mail: [kehchen@ucdavis.edu](mailto:kehchen@ucdavis.edu); address: Department of Statistics, UC Davis, One Shields Avenue, Davis, CA 95616, USA.

<sup>3</sup> Research supported in part by NSF grant DMS08-06199.

## 1. Introduction

Quantile estimation is of interest in a variety of settings. For instance, growth charts, also known as reference centile charts, are widely used to screen growth status. Conventional growth charts are based on marginal percentile curves combined with a transformation to normality (Cole, 1988). Figure 1 shows an example of cross-sectional growth charts as well as the longitudinal growth curves of two girls. Evaluation of children on the basis of their observed growth path leads to much better understanding of their current auxological status than simply comparing their height with a reference chart at a specific age. As a consequence, several authors, including Cole (1994), Royston (1995), Thompson and Fatti (1997), have considered conditional reference charts based on longitudinal measurements and advocated the importance of adequately reflecting the longitudinal aspect when quantifying growth status. However, these methods usually assume joint normality and a certain type of parametric structure of the longitudinal measurements and usually also require regular spacing of the measurement times; an example is the LMS-AR model considered in Cole (1994). Wei and He (2006) proposed a more general semiparametric model based on  $p$  prior measurements for the growth path when constructing conditional quantiles for current height status. Their proposed model waives the normality assumption and to some extent allows for varying time spacing of the measurements, by incorporating the time spacing into the coefficients through a linear structure.

It is not obvious how to compare a child with prior measurements  $(Y_1, \dots, Y_m)$  with the growth status of a reference group, especially as the measurements often have variable time spacing and their timing may be subject-specific. Addressing these issues, we suggest constructing conditional quantiles and adopting a functional data analysis perspective, by viewing the entire growth history of a child as a functional but latent covariate, for which the only information available is a series of irregularly timed and potentially noisy measurements. For growth measurements, the noise is due to direct measurement error and also diurnal and seasonal variations in height. Taking the potentially irregular nature of the timing of the available measurements and noise into account, we aim at recovering the underlying growth curves, constructing conditional reference charts for current status and quantifying and assessing current

growth status through the predicted conditional quantiles of a future defined growth status, such as adult height.

The smooth functional nature of growth curves has been explored in various previous statistical analyses, including Gasser *et al.* (1984, 1985), Gasser *et al.* (1991), Kneip and Gasser (1992), Gasser and Kneip (1995) and Sheehy *et al.* (1999). For individual trajectories measured at a dense grid of regularly spaced time points, there exists an extensive literature on recovering the underlying functions. Older approaches have focused on parametric modeling (Marubini *et al.*, 1971; Ellozy, 1978), while modern approaches tend to be more flexible, allowing for more variable shapes of the growth trajectories and their derivatives, often by employing nonparametric methods. Such flexible approaches for the analysis of irregular, noisy or sparsely sampled longitudinal data have for example been developed by Shi *et al.* (1996) and Rice and Wu (2001), based on modeling with B-splines that include random effects, and James *et al.* (2000) and Yao *et al.* (2005), based on modeling with functional principal components. Overviews of the rapidly evolving interface between longitudinal and functional methodology are provided in Rice (2004), Zhao *et al.* (2004) and Müller (2005).

Mainstream approaches to estimate conditional quantiles when covariates are vectors include (1) the estimating equation approach, introduced by Koenker and Bassett (1978), which has spawned a large literature and more recently has been modified and extended to more general cases (Koenker *et al.*, 1994; Li *et al.*, 2007), and (2) obtaining quantiles by inverting a conditional distribution function. Regarding the latter, which is the approach we adopt here, various nonparametric methods for estimating conditional distribution functions and then quantiles with low-dimensional covariates have been proposed over the years (Fan *et al.*, 1996; Yu and Jones, 1998; Hall *et al.*, 1999; Cai, 2002; Hall and Müller, 2003).

For the case of functional covariates, Cardot *et al.* (2005) extended the estimating equation approach to a linear functional quantile regression model, while Ferraty *et al.* (2005) proposed to extend the double kernel method for estimating a conditional distribution function to functional covariates. Due to the nature of the estimating equations, these methods assume structural constraints in the quantiles, which we abandon here by proposing to estimate the conditional

distribution function under a functional generalized regression framework. The proposed approach facilitates balancing of model flexibility and of the so called curse of dimensionality, which affects direct “nonparametric” approaches. Although this paper is inspired and illustrated by the growth charts problem, our proposed methods and theoretical results pertain to a general setting and can be used in a variety of situations where the need arises to estimate conditional quantiles when covariates are functions.

The paper is organized as follows. In section 2, we introduce the proposed model and an extended time-varying model framework. Methods for estimation are discussed in section 3. In section 4, we highlight some features of the proposed method using simulated data. This is followed by an analysis of the Berkeley Growth Data in section 5. A description of asymptotic consistency properties is in section 6 and further discussion in section 7.

## 2. Modeling Conditional Distributions With Functional Covariates

The covariates we consider are random trajectories  $X$ , for which one records an associated response  $Y$ , a one-dimensional random variable. Predictor processes  $X$  are assumed to be square integrable on a domain  $\mathcal{T} = [0, T]$ , with mean function  $EX(s) = \mu(s)$  and a continuous covariance function  $\text{cov}(X(s), X(t)) = G(s, t)$ . The covariance function  $G(s, t)$  has an orthogonal expansion  $G(s, t) = \sum_k \lambda_k \phi_k(s) \phi_k(t)$  with nonincreasing eigenvalues  $\lambda_k$  and functions  $\phi_k$ , which correspond to the eigenfunctions of the associated autocovariance operator  $A_G$ , defined by  $A_G(f)(t) = \int_{\mathcal{T}} G(s, t) f(s) ds$ , for any  $f$  in  $L^2$ . Then one may represent random processes  $X$  by a Karhunen-Loève expansion  $X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t)$ , where  $\xi_k = \int_{\mathcal{T}} (X(t) - \mu(t)) \phi_k(t) dt$  are the functional principal components (FPCs). The FPCs form a sequence of uncorrelated random variables with mean zero and decreasing variances, which correspond to the eigenvalues, i.e.  $E(\xi_k^2) = \lambda_k$ , and serve as random coefficients in functional statistical models.

The conditional distribution of  $Y$  given  $X$ , denoted as  $F(y | X) = P(Y \leq y | X) = E(I(Y \leq y) | X)$ , where  $I$  is the indicator function, may be viewed as the regression of the indicator  $I(Y \leq y)$  on the functional predictor  $X$ . Then, a reasonable approach is to connect the functional quantile problem with the functional generalized linear regression model with

known link function  $g$  and functional predictors  $X$ . Such models have been widely studied in the literature (James, 2002; Cardot and Sarda, 2005; Müller and Stadtmüller, 2005; Leng and Müller, 2006; Escabias *et al.*, 2007).

More precisely, for a fixed  $y$ , the indicator  $I(Y \leq y)$ , given the covariate function  $X$ , has a binomial distribution with mean parameter  $F(y | X)$ , leading to a functional generalized linear regression model with link function  $g$ ,

$$F(y | X) = E(I(Y \leq y) | X) = g^{-1}\{\beta_0(y) + \int X^c(t)\beta(y, t)dt\}, \quad (1)$$

where  $X^c(t) = X(t) - \mu(t)$ , and  $g$  is a monotone link function, for example, the logit link, with  $g^{-1}(z) = \exp(z)/(1 + \exp(z))$ .

The coefficient function  $\beta(y, t)$ , for fixed  $y$ , is assumed to be square integrable on  $\mathcal{T}$  and can then be represented in the orthogonal eigenbasis,  $\beta(y, t) = \sum_{k=1}^{\infty} \beta_k(y)\phi_k(t)$ , where  $\int \phi_j(t)\phi_k(t)dt = \delta_{jk}$ ,  $\delta_{jk} = 1$  for  $j = k$  and  $\delta_{jk} = 0$  for  $j \neq k$ . Then the model in eq. (1) can be equivalently written as

$$F(y | X) = E(I(Y \leq y) | X) = g^{-1}(\beta_0(y) + \sum_{k=1}^{\infty} \beta_k(y)\xi_k), \quad (2)$$

where the  $\xi_k$  are the FPCs of  $X$ , and the  $\beta_k(\cdot)$ ,  $k = 1, 2, \dots$ , are assumed to be smooth functions of  $y$ .

The desired conditional quantiles  $Q(\alpha)$ ,  $0 < \alpha < 1$ , of  $Y$  may be obtained by inverting the conditional distribution function  $F(y | X)$ , defining

$$Q(\alpha) = \inf\{y : F(y | X) \geq \alpha\}, \text{ for } 0 < \alpha < 1. \quad (3)$$

Model (2) can be easily extended to a time-varying model, where the right endpoint of  $\mathcal{T}$  is increasing (adapting an approach described in Müller and Zhang, 2005). For time-varying conditional quantiles that depend on the domain on which predictors have been observed, we consider the extended model

$$F(y | X_{[0,s]}) = E(I(Y \leq y) | X_{[0,s]}) = g^{-1}\{\beta_0(y, s) + \sum_{k=1}^{\infty} \beta_k(y, s)\xi_k\}, \quad (4)$$

where  $y$  denotes response levels at a time  $\tau \geq s$  and  $X_{[0,s]}$  denotes the predictor curve on the interval  $[0, s]$ . A motivating example is provided by selecting  $y$  as adult height and  $s$  as an age at which the status of a child is to be assessed, using the available growth data from birth to age  $s$ . Then the entire growth history of a child up to age  $s$  is used for predicting the quantile of adult height for this child. This device allows to associate an overall growth status assessment that is based on the predicted adult height, with the specific growth history of the child on currently available domains from birth to age  $s$ , for varying ages  $s$  of the child.

### 3. Estimating the Model Components

To address the difficulty caused by the infinite dimensionality of the predictors, we approximate model (2) with a series of models, for which the number of predictors is truncated at  $p = p_n$  included components, and the dimension  $p_n$  increases asymptotically, as the sample size  $n \rightarrow \infty$ ; compare with Müller and Stadtmüller (2005).

A heuristic motivation for this truncation strategy is as follows: Setting

$$\eta_p(y) = \beta_0(y) + \sum_{j=1}^p \beta_j(y)\xi_j, \quad R_p(y) = \sum_{j=p+1}^{\infty} \beta_j(y)\xi_j, \quad Z(y) = I(Y \leq y), \quad X_p = \mu + \sum_{k=1}^p \xi_k \phi_k,$$

model (2) can be written as  $F(y | X) = E(Z(y) | X) = g^{-1}(\eta_p(y) + R_p(y))$ . If the FPC's are independent, as is the case for a Gaussian random process,  $R_p(y)$  is independent of  $X_p$ . With distribution function  $F_{R_p(y)}$  of  $R_p(y)$ ,

$$E(Z(y) | X_p) = E[g^{-1}(\eta_p(y) + R_p(y)) | X_p] = \int g^{-1}(\eta_p(y) + s) dF_{R_p(y)}(s) = g_{p,y}^{-1}(\eta_p(y)),$$

for suitable functions  $g_{p,y}(\cdot)$ , and  $\text{var}(Z(y) | X_p) = g_{p,y}^{-1}(\eta_p(y))(1 - g_{p,y}^{-1}(\eta_p(y)))$ . Therefore the conditional distribution of  $Z(y)$ , conditioning on  $X_p$ , is a binomial distribution with parameter  $F(y | X_p) = g_{p,y}^{-1}(\eta_p(y))$ , motivating the consideration of an increasing series of truncated binomial models. The approximation error of these truncated models is directly tied to the eigenvalues  $\lambda_j = \text{var}(\xi_j)$  and vanishes asymptotically as  $p \rightarrow \infty$  (see proof of the Theorem in Section 6).

A feature that complicates theoretical analysis is that the link functions  $g_{p,y}$  must be considered as unknown, as they depend on  $\eta_p$  and the distribution of  $R_p(y)$ , which are both unknown.

In actual estimation, we are confined to fit the model using the fixed known link function  $g$ , instead of using the actual and changing link function  $g_{p,y}$ , therefore fitting

$$\hat{F}(y | X) = g^{-1}(\hat{\beta}_0(y) + \sum_{k=1}^p \hat{\beta}_k(y)\xi_k). \quad (5)$$

Examining the discrepancy between link functions  $g$  and  $g_{p,y}$  carefully, it can be proved that the discrepancy between these functions vanishes asymptotically under some mild conditions (see proof of the Theorem). This leads to the uniform consistency of our estimate  $\hat{F}(y | X)$  under regularity conditions, and as a consequence, one obtains a consistent estimate  $\hat{Q}(\alpha)$  of  $Q(\alpha)$  from (3),

$$\hat{Q}(\alpha) = \inf\{y : \hat{F}(y | X) \geq \alpha\}, \text{ for } 0 < \alpha < 1. \quad (6)$$

In practice, the choice of the number of components  $p$  to be included in eq. (5) can be determined by leave-one-curve-out cross-validation or by the fraction of variance explained by the first  $p$  components. One can also adopt AIC or BIC type criteria, see Yao *et al.* (2005) or Müller and Stadtmüller (2005) for further details.

To estimate the coefficients in a truncated model, an initial step is to obtain the first  $p$  FPCs  $\{\xi_{ik}, k = 1, \dots, p\}$ , for each predictor function  $X_i$ . The predictor trajectories  $X_i$  are usually recorded at a grid of time points, which could differ from subject to subject, and the measurements might be contaminated by measurement errors and other aberrations. For example, in auxology it is well known that children and also adults are taller in the morning than in the evening. To model functional data that are contaminated with measurement errors, denoting the observation of the random trajectory  $X_i$  at time  $t_{il}$  by  $U_{il}$ , the corresponding error by  $\varepsilon_{il}$  and the number of observations made for the  $i$ th subject by  $L_i$ , we may represent the observed data as follows:

$$U_{il} = X_i(t_{il}) + \varepsilon_{il} = \mu(t_{il}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t_{il}) + \varepsilon_{il}, \quad 1 \leq i \leq n, 1 \leq l \leq L_i, \quad (7)$$

where  $X_i$  is the smooth underlying random trajectory for the  $i$ th subject and the errors  $\varepsilon_{il}$  are assumed to be i.i.d. with  $E\varepsilon_{il} = 0$ ,  $E\varepsilon_{il}^2 = \sigma^2$ , and to be independent of the FPCs.

As mean, covariance and eigenfunctions are assumed to be smooth, we propose using local linear smoothing to obtain the estimated mean function  $\hat{\mu}(t)$ , and a two-dimensional weighted

least squares smoother, omitting the diagonal elements, for the estimated covariance surface  $\hat{G}(s, t)$  (for further details see Yao *et al.*, 2005). Smoothing bandwidths are chosen by cross-validation or generalized cross-validation. Estimated eigenfunctions and eigenvalues are then the solutions of the eigenequations,

$$\int G(s, t)\phi_k(s)ds = \lambda_k\phi_k(t),$$

where  $\hat{G}$  is substituted for  $G$  and the eigenfunctions are subject to the constraints  $\int \phi_k(s)^2 ds = 1$  and  $\int \phi_j(s)\phi_k(s)ds = 0$ , for  $j < k$ . These solutions are obtained by discretizing the estimated covariance surface and then using the corresponding matrix procedures; FPCs then are determined by  $\hat{\xi}_{ik} = \int (X_i(t) - \hat{\mu}(t))\hat{\phi}_k(t)dt$  for densely observed data or through conditional expectation (BLUP) in the case of sparse irregular observations. More details can be found in Yao *et al.* (2005); these procedures are available in the PACE package, which is written in Matlab and available at <http://anson.ucdavis.edu/~mueller/data/pace.html>. We specifically note that the code for our proposed method is available in PACE version 2.14.

With the first  $p$  FPCs, estimates of the coefficients  $\hat{\beta}_0, \dots, \hat{\beta}_p$  are easily obtained through the usual estimating equation for the generalized linear model, leading immediately to the estimate of the distribution function  $\hat{F}(y | X)$  in eq. (5). Generally, we will be interested in conditional distribution functions and quantiles over a certain range of values, which we denote as  $[q_1, q_2]$ . We obtain coefficient functions  $\hat{\beta}_k(y)$ ,  $0 \leq k \leq p$ , by fitting the model for a dense grid of  $y \in [q_1, q_2]$ . An optional step is to additionally smooth the coefficient functions  $\hat{\beta}_k(y)$ , or, more directly, the estimates  $\hat{F}(y|X)$ , across  $y$ , which may lead to smoother and better behaved estimates. In our implementations, we use the second type of smoothing, using kernel smoothers with approximately 5 data points in each smoothing window.

We note that the range  $[q_1, q_2]$  needs to be large enough to obtain desired quantiles for small or large  $\alpha$ . In practice, the choice of  $q_1$  and  $q_2$  depends on the observed range of the responses

Y. For a subject with trajectory  $X_i$ , the estimate for  $F(y | X_i)$  then becomes

$$\hat{F}(y | X_i) = \begin{cases} 0 & \text{for } y \leq q_1 \\ g^{-1}(\hat{\beta}_0(y) + \sum_{k=1}^p \hat{\beta}_k(y)\hat{\xi}_{ik}) & \text{for } y \in (q_1, q_2) \\ 1 & \text{for } y \geq q_2 \end{cases} \quad (8)$$

#### 4. Simulation Results

To illustrate our methods for conditional quantile estimation with functional covariates, pairs of  $(X_i, Y_i)$ ,  $X_i \in L^2[0, 10]$ ,  $Y_i \in \mathbb{R}$ ,  $i = 1, \dots, 400$ , were generated, and of these, 300 randomly selected pairs were used for training the model and the remaining 100 pairs for testing. We compute and evaluate the proposed approach  $\hat{Q}_i(\alpha)$  for the test data, and its performance was compared with the following two methods: (a) a naive approach, obtained by using unconditional empirical quantiles, denoted by  $Q^U(\alpha)$ ; and (b) the quantile linear regression model with functional covariates as proposed in Cardot *et al.* (2005).

The method of Cardot *et al.* (2005) is an extension of the well known quantile linear regression approach for vector covariates  $X$ , based on the estimating equation  $E(l_\alpha(Y - \beta_0 - \beta_1 X))$ , where  $l_\alpha(u) = |u| + (2\alpha - 1)u$ . For a functional predictor  $X(t)$ , Cardot *et al.* (2005) proposed to replace  $\beta_0 + \beta_1 X$  by  $c + \int \Psi_\alpha(t)X(t)dt$ , where the coefficient function  $\Psi_\alpha(t)$  is expanded in the B-spline basis. The corresponding coefficient vector  $\theta$  is estimated via the following penalized estimating equation,

$$\min_{c \in \mathbb{R}, \theta \in \mathbb{R}^{k+q}} \left\{ \frac{1}{n} \sum_{i=1}^n l_\alpha(Y_i - c - \int B_{k,q}^T(t)\theta X_i(t)dt) + \rho \| (B_{k,q}^T \theta)^{(m)} \|^2 \right\}. \quad (9)$$

The predicted quantiles, denoted by  $\hat{Q}_i^C(\alpha)$ , are obtained by substituting estimates of the coefficients in the equation  $Q_i^C(\alpha) = c + \int \Psi_\alpha(t)X_i(t)dt$ .

Predictor trajectories  $X_i$  were generated from a  $L^2$  random process with mean function  $\mu(t) = t + \sin(t)$ ,  $0 \leq t \leq 10$ , and a covariance function derived from  $K$  eigenfunctions,  $\phi_i(t) = \cos((i+1)\pi t/10)/\sqrt{5}$  for odd values of  $i$  and  $\phi_i(t) = \sin(i\pi t/10)/\sqrt{5}$  for even values of  $i$ ,  $0 \leq t \leq 10$ , for  $K = 2, 4, 6, 8, 10$ . The corresponding non-zero eigenvalues were chosen as the first  $K$  numbers from the sequence  $L = \{4^2, 3^2, 2.75^2, 2.25^2, 1.75^2, 1.25^2, 1, 0.75^2, 0.5^2, 0.25^2\}$ ,

i.e.  $\lambda_k = L(k)$ ,  $k = 1, \dots, K$ , and  $\lambda_k = 0$ , for  $k > K$ . The underlying FPCs in eq. (7) were generated as  $\xi_{ik} \stackrel{iid}{\sim} N(0, \lambda_k)$ , and the measurement errors as  $\varepsilon_{il} \stackrel{iid}{\sim} N(0, 0.5^2)$ . For the generation of  $Y_i$ , we consider two scenarios: A Gaussian case,  $Y_i | X_i \sim N(2 \sum_{k=1}^K \xi_{ik}, 5^2)$ , and a mixture distribution,  $Y_i | X_i \sim 0.5N(\sum_{k=1}^K \xi_{ik}, 1^2) + 0.5N(3 \sum_{k=1}^K \xi_{ik}, 4^2)$ .

Regarding the measurement locations at which predictor trajectories were sampled, two settings were considered, a sparse irregular setting, where the number of measurements for each subject was chosen from  $\{4, \dots, 14\}$  with equal probability, the locations were chosen separately and were uniformly distributed on  $[0, 10]$ ; and a regular dense case, where 30 equally spaced observations on  $[0, 10]$  were generated for each subject. The method of Cardot *et al.* (2005) requires densely observed functional data, and thus will not work for the first scenario with sparse design. For comparison purposes, we created a modified version of Cardot's method, specifically for the sparse design case, and this modified version was used for the comparisons pertaining to this case. To summarize, there are four simulation scenarios, referred to in the following as Gaussian sparse, Gaussian dense, Gaussian mixture sparse and Gaussian mixture dense cases, respectively.

The smoothing bandwidths for the proposed method were chosen by generalized cross-validation. For the sparse case, these bandwidths were approximately 0.59 for the mean function and 1.1 for the covariance surface, and for the dense case, 0.6 and 0.5 for mean and covariance function, respectively, slightly differing between simulation runs. The number of included components  $p$  was chosen by AIC, using a marginal likelihood (pseudo-likelihood) of the observations. This method and several alternative selectors are discussed in Yao *et al.* (2005). Alternative methods include BIC, minimizing leave-one-curve-out prediction error, or controlling the fraction of variance explained (FVE),  $FVE = \sum_{k=1}^p \hat{\lambda}_k / \sum_{k=1}^M \hat{\lambda}_k$ , where  $M$  is chosen large. For all simulations, we used the logit link function.

For the functional quantile linear regression model proposed by Cardot *et al.* (2005), the estimator depends on the number of knots  $k$ , the degree  $q$  of the spline, the order  $m$  of the derivative chosen for the penalization term, and the smoothing parameter  $\rho$ . It seems that only  $\rho$  has a major impact on the resulting estimate, provided that the number of knots is chosen

large enough (Besse *et al.*, 1997; Cardot *et al.*, 2007). We use cubic B-splines with 8 inner knots and penalty terms based on second order derivatives, the same specifications as in Cardot *et al.* (2007) and also investigated alternatively choosing 6 and 10 knots, which led to nearly identical results as using 8 knots. We therefore only report the results for this case. The tuning parameter  $\rho$  in eq. (9) was chosen by generalized cross-validation.

In Tables 1 and 2, the results are listed for mean absolute error (MAE), based on 200 simulation runs for each simulation scenario and quantile levels  $\alpha \in (0.05, 0.1, 0.25, 0.5)$ ; estimates of the upper quantiles were found to behave similarly (not reported). Additional information can be found in Figure 2, which contains the boxplots of MAE for conditional quantile estimation,  $\alpha \in (0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95)$ , over 200 simulation runs for Gaussian (left) and mixture Gaussian (right) distributions, for densely sampled designs generated from 10 eigenbases.

The simulation results in Table 1, Table 2 and Figure 2 indicate that the proposed method outperforms the two comparison methods for all quantiles in the case of sparsely and irregularly observed predictor functions. For the case of a mixture of Gaussians, the proposed method outperforms the approach of Cardot *et al.* for all quantiles, for both sparsely and densely observed predictor functions (Table 2 and Figure 2 right part). For the case of a Gaussian setting coupled with densely observed predictor functions, the approach of Cardot *et al.* performs better than the proposed method for the median, and comes close to the proposed method for first and third quartiles, but for more extreme quantiles, its performance is clearly inferior to that of the proposed method. We note that for the growth status screening application, small quantiles such as those with levels  $\alpha = 0.05$  or  $\alpha = 0.1$  matter most, as the goal is to indicate which children might need additional health screening, based on their predicted very low adult height status. We conclude that the proposed method is preferable for this and similar applications.

## 5. Application to Berkeley Growth Data

One of the most important roles of growth monitoring is to screen the growth status of children and to identify children with underlying health problems that are reflected in stunted height growth. Growth reference charts are widely used by pediatricians and auxologists for this

purpose. We demonstrate the proposed functional conditional quantile estimation with an application to the Berkeley growth data (Tuddenham and Snyder, 1954). These data contain height measurements for 54 girls, with 31 measurements taken between ages 1 year and 18 years.

Assume we wish to make a growth assessment for a given girl at age  $T$  based on a series of unequally spaced and randomly numbered height growth measurements that are available for this particular girl in the age interval  $[0, T]$ . The ultimate outcome of interest is the girl's adult height  $Y$ . Accordingly, evaluation and auxological assessment of growth patterns in  $[0, T]$  will benefit from linking observed growth in the interval  $[0, T]$  to adult height  $Y$ , which serves as a natural summary of health status over the entire growth phase. This motivates to predict the distribution function of adult height  $Y$  conditional on the growth history  $[0, T]$ , a particularly low quantile indicating that pediatric follow-up might be needed.

To exemplify this approach, we choose  $T = 12$  years. Leave-one-out cross-validation is used to assess prediction performance, i.e. when predicting the conditional quantiles of adult height for the  $i$ th subject, we omit the data of this subject when fitting the model. The fitted coefficient functions  $\hat{\beta}_0(y)$  and  $\hat{\beta}_k(y)$ , defined in eq. (5), with logit link and  $p = 2$ , the number of components selected by the AIC criterion (explaining more than 95% of the total variance), are shown in Figure 3. The predicted  $100\alpha\%$  quantiles  $\hat{Q}_i(\alpha)$  of adult height  $Y$ , conditional on the growth history on  $[0, 12]$ , for  $\alpha \in \{0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95\}$ , were obtained from (6). The resulting predicted quantiles for four randomly selected girls are displayed in Figure 4.

Assessing the quality of estimated quantiles is not straightforward, as commonly only one pair of data  $(X_i, Y_i)$  at a particular predictor level  $X = X_i$  is available,  $X_i$  indicating the predictor function and  $Y_i$  adult height. There are no observations pertaining to the true conditional distribution. For an indirect assessment of the quality of predicted quantiles observe that for the indicator function  $I_i(\alpha) = I(Y_i \leq Q_i(\alpha))$ , it holds that  $E(I_i(\alpha) | X_i) = \alpha$ . Therefore

$$\bar{I}(\alpha) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq \hat{Q}_i(\alpha)) \quad (10)$$

should be close to  $\alpha$ , if the quantile estimation method gives reasonable estimates. The scatterplot of  $\bar{I}(\alpha)$  against  $\alpha$  is found to be indeed close to the identity line in Figure 5, indicating

that the proposed method performs reasonably well with regard to this criterion.

Another approach, previously used in Wei and He (2006), is to compare the empirical distribution of  $Y$  with the simulated distribution from the conditional distribution model. Considering conditional quantiles  $Q_i(\alpha) = F_i^{-1}(\alpha)$ , where  $F_i(y)$  denotes  $P(Y \leq y | X = X_i)$ , for  $\alpha \sim \mathcal{U}(0, 1)$  and a fixed  $X_i$ ,  $P(Q_i(\alpha) < y) = P(F_i^{-1}(\alpha) < y) = P(\alpha < F_i(y)) = F_i(y) = P(Y \leq y | X = X_i)$ . Therefore the marginal distribution of  $Q_i(\alpha)$  for randomly sampled  $X_i$  and  $\alpha \sim \mathcal{U}(0, 1)$  should be the same as the marginal distribution of the responses, which can be checked as follows: Sample  $X_i$  from the empirical distribution of  $X$  and  $\alpha$  from  $\mathcal{U}(0, 1)$ . Setting  $z = \hat{Q}_i(\alpha)$ , repeat this procedure  $L$  times and obtain a sample  $\{z_1, z_2, \dots, z_L\}$ . If the quantile estimates are reasonable, the two marginal distributions defined by the observed responses in the original data and by the simulated responses  $\{z_1, z_2, \dots, z_L\}$  should match. Choosing  $L = 3000$ , we find from the Q-Q plot in Figure 5 that the match is indeed quite good. We conclude that the proposed functional quantile estimation passes both of these diagnostic tests.

Assessing the growth status of girls at age 12, we use conventional unconditional reference charts to identify two girls, whose heights at age 12 are below the 10th cross-sectional percentile of girl's height at age 12, measuring 142.1 cm (girl A) and 141.3 cm (girl B), respectively. In order to obtain a longitudinal assessment of growth status, based on the growth history from birth to age 12, we proceed to estimate the conditional distribution of adult height for these two girls, with results given in Figure 6. These two conditional distribution functions demonstrate different patterns. Choosing as response  $Y = H_{18}$ , we obtain the conditional probabilities  $P(H_{18} < 157\text{cm} | X_{[0,12]}^A) = 0.2235$  and  $P(H_{18} < 157\text{cm} | X_{[0,12]}^B) = 0.8868$ , where 157 cm is the empirical 10th percentile of height for female adults.

Thus, the event that girl B ends up shorter than 157 cm as an adult is predicted to occur with high probability, indicating that further pediatric evaluation and possibly intervention is needed, while the probability for girl A to have a short adult height is estimated to be much lower. The actual adult height for girl A was observed at 164 cm and for girl B at 154.5 cm. The proposed method thus correctly identifies the girl with actual height corresponding to a low quantile. From the derivatives of the growth curves in the bottom panels of Figure 6, obtained

with local polynomial fitting, girl B is seen to have reached the pubertal growth spurt ahead of girl A. As a consequence, at age 12, growth of girl B was steadily decelerating, while girl A had not reached the point where pubertal growth acceleration starts. Thus the difference in the estimated distribution functions is due to differences inherent in the entire growth pattern up to age 12; it is not evident from the small difference in heights observed at age 12.

Current growth status for a girl at age  $s$  may be evaluated by comparing current height with a set of conditional quantiles estimated from model (4), trained on the interval  $[0, s - \delta]$ , for a suitable  $\delta > 0$ . We applied this idea to screen the height of girls at age 13, based on their growth history on the age interval  $[0, 12]$ , so that  $s = 13$ ,  $\delta = 1$ , and the response  $Y$  corresponds to the height at age 13. Growth trajectories on  $[0, 12]$  and height at age 13 are shown for four girls (C,D,E and F) in Figure 6, along with the corresponding estimated conditional quantiles  $Q_i(\alpha)$ , for  $\alpha \in \{0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95\}$  and the actual height attained at age 13.

The height of 157.8cm observed for girl C is far below expectation, although it is slightly above the trend of this girl's growth trajectory. Specifically, from the predicted quantiles, height at age 13 is seen to fall below the 5th percentile. So what is happening? Interestingly, girl C grew very fast subsequently, between year 14 and 15, and ended up with a normal adult height of 170.3cm. It turns out that this girl had an unusually late occurring pubertal growth spurt, in comparison with the timing of the pubertal growth spurt for the other girls across the sample. The relative delay in the pubertal growth spurt is the reason for the unexpectedly short height at age 13 and the correspondingly low conditional quantile. For the other girls, the observed height at age 13 by and large falls within the middle range of the predicted quantiles.

## 6. Asymptotic Properties

Recall the conditional distribution model

$$F(y | X) = E(Z(y) | X) = g^{-1}\left(\beta_0(y) + \sum_{k=1}^{\infty} \beta_k(y)\xi_k\right),$$

where  $Z(y) = I(Y \leq y)$ . The estimator based on the first  $p$  components is  $\hat{F}(y | X) = g^{-1}(\sum_{k=0}^p \hat{\beta}_k(y)\xi_k)$ , where we set  $\xi_0 \equiv 1$ , and  $\hat{\beta}^p(y) = (\hat{\beta}_0(y), \dots, \hat{\beta}_p(y))$  is a solution of the

following minimization problem:

$$\hat{\boldsymbol{\beta}}^p(y) = \arg \min_{(\tilde{\beta}_0, \dots, \tilde{\beta}_p)} \sum_{i=1}^n L(g^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_{ik}), Z_i(y)), \quad (11)$$

where

$$L(\theta, u) = -u \log \frac{\theta}{1-\theta} - \log(1-\theta) \quad (12)$$

is the negative log likelihood function of a binomial random variable  $U$  with mean parameter  $\theta$ .

We define  $\tilde{\eta}_p = \sum_{k=0}^p \tilde{\beta}_k \xi_k$  for any  $\tilde{\beta}_k$ , and  $\eta_p(y) = \sum_{k=0}^p \beta_k(y) \xi_k$  for the linear part with true coefficients  $\boldsymbol{\beta}^p(y) = (\beta_0(y), \dots, \beta_p(y))$ .

The following conditions will be needed.

(A.1) Assume that there exist  $\zeta > 0$  and an interval  $I$  such that  $P(I \subset D_\zeta(X)) = 1$ , where  $D_\zeta(X) = \{y \in R : \zeta \leq F(y | X) \leq 1 - \zeta\}$ .

This assumption basically requires that  $F(y | X)$  stays away from 0 and 1 with probability 1 on a suitably defined interval  $I$ . This is needed for establishing the uniform consistency of the estimated conditional distribution function on the interval  $I$ .

$$(A.2) \quad \sup_{z \in R} |(g^{-1})'(z)| \leq C.$$

This assumption requires a uniform bound on the derivative of the inverse link function. This is satisfied for many important link functions, including the logit link function.

(A.3) The functional principal components (FPCs)  $\{\xi_1, \xi_2, \dots\}$  are independent random variables.

This is for example satisfied in the commonly considered case where the predictors are Gaussian random processes. But Gaussianity of predictors is not required. Of course, the FPCs will always be uncorrelated by definition. Assumption (A.3) is used to prove that the discrepancy between link function  $g$  and actual link function  $g_{p,y}$  vanishes asymptotically.

(A.4) For any  $p \geq 1$ , define the Hessian  $p \times p$  matrix  $M_p = \frac{\partial^2}{\partial^2(\tilde{\beta}_0, \dots, \tilde{\beta}_p)} EL(g^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_k), Y)$ . There exists  $\gamma_p > 0$  such that  $\{M_p - \gamma_p I_p\}$  is positive definite, where  $I_p$  is the  $p \times p$  identity matrix.

$$(A.5) \quad \sum_{j=p+1}^{\infty} \beta_j(y)^2 \rightarrow 0, \quad \frac{\sum_{j=p+1}^{\infty} \beta_j(y)^2}{\gamma_p^2} \rightarrow 0, \quad \text{uniformly on the interval } I, \text{ for } \gamma_p \text{ as in (A.4).}$$

(A.4) and (A.5) reflect strong convexity around the maximum of  $EL(g^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_k), Y)$  and assumptions on the asymptotic behavior of the curvature around the maximum as  $p$  increases. Importantly, for the case of a logit link function, we observe that  $M_p \geq C' E(\xi^p (\xi^p)^T)$ , where  $C' > 0$  is a constant,  $\xi^p = (\xi_0, \dots, \xi_p)^T$  and  $E(\xi^p (\xi^p)^T) = \text{diag}(1, \lambda_1, \dots, \lambda_p)$ , implying that  $\gamma_p$  can be chosen as  $\min(C', C' \lambda_p)$  in this case.

Our main result is the following

**Theorem.** *Provided that conditions (A.1)-(A.5) hold, for a suitably chosen sequence  $p(n) \rightarrow \infty$ , as  $n \rightarrow \infty$ , on intervals  $I$  as in (A.1),*

$$\sup_{y \in I} |\hat{F}(y | X) - F(y | X)| \xrightarrow{P} 0$$

where  $\hat{F}(y | X) = g^{-1}(\sum_{k=0}^{p(n)} \hat{\beta}_k(y) \xi_k)$ . Therefore, for any  $0 < \alpha < 1$ , the estimator  $\hat{Q}(\alpha) = \inf\{y : \hat{F}(y | X) \geq \alpha\}$  of the 100 $\alpha$ % conditional quantile of  $Y$  given  $X$  is a consistent estimator of  $Q(\alpha) = \inf\{y : F(y | X) \geq \alpha\}$ .

The proof of this uniform consistency result can be found in the Appendix.

## 7. Discussion

We propose estimating conditional distribution functions under a generalized functional regression framework, where the response is a binary variable and the predictors are in a functional space. The corresponding generalized functional linear model approach for estimation is investigated through simulations, growth data analysis and asymptotic study. Other generalized functional regression models are also of potential interest for extensions to quantile estimation, such as generalized functional additive models or generalized functional single index models. After estimating conditional distribution functions, one obtains conditional quantiles by inverting the distribution function. Smoothing and monotonization steps can be added as desired.

The quantile linear regression approach in general works better than the inversion of estimated conditional distribution functions, when the assumptions for this approach are satisfied (Koenker, 2005). These assumptions include that the quantiles are linear in the predictors, as in the Gaussian case. The superiority of the quantile linear regression approach has been

established for the case where predictors are directly observed without errors. However, in the functional predictor case the situation is different and more complex. A distinctive feature is that one usually does not directly observe entire predictor functions, but only has sparsely or densely sampled measurements of each random function. The available measurements of each random function are made at a sample of time points (in the sparse case perhaps as few as 2-5 measurements), which are additionally contaminated with error. Therefore one needs to recover individual random functions from these measurements, for which in the approach of Cardot *et al.* B-splines are used, while in our proposed approach we use eigenfunctions and functional principal components. The estimated functional principal components or spline components are contaminated with estimation error, a distinctive feature compared with the usual situation in multivariate linear regression where the predictors are assumed to be observed without error.

To illustrate these issues in a non-functional simple bivariate linear model, consider  $Y = X + U$ , where  $X \sim N(0, \sigma^2)$  and  $U \sim N(0, \sigma_1^2)$ ;  $X$  and  $U$  are independent. Then the  $\tau$ th quantile of  $Y$  conditional on  $X$  is  $Q_{Y|X}(\tau) = X + \sigma_1 C(\tau)$ , where  $C(\tau)$  is the  $\tau$ th quantile of a standard normal. Suppose we do not observe  $X$  but instead observe a contaminated version  $\tilde{X} = X + V$ ,  $V \sim N(0, \sigma_2^2)$ , where  $V$  is independent of  $X$ . Then

$$\begin{aligned} Q_{Y|\tilde{X}}(\tau) &= \frac{\sigma^2}{\sigma_2^2 + \sigma^2} \tilde{X} + \left\{ \sigma^2 + \sigma_1^2 - \frac{\sigma^4}{\sigma^2 + \sigma_2^2} \right\}^{1/2} C(\tau) \\ &= \frac{\sigma^2}{\sigma_2^2 + \sigma^2} X + \left\{ \sigma^2 + \sigma_1^2 - \frac{\sigma^4}{\sigma^2 + \sigma_2^2} \right\}^{1/2} C(\tau) + \frac{\sigma^2}{\sigma_2^2 + \sigma^2} V, \end{aligned} \quad (13)$$

and one finds that the contamination in the predictor has the effect of altering the slope and in addition introducing a constant deviation term (note that the factor in the second term of (13) differs more and more from  $\sigma_1$  with increasing contamination variance  $\sigma_2^2$ ) and a random deviation term (the third term in (13), which has expectation 0). The constant deviation term is seen to increase for more extreme quantiles, and vanish for the median.

Considering the case with functional predictors and the simulation results for the Gaussian case (Table 1 and Figure 2 left panel), in the light of the above, the approach of Cardot *et al.* seems indeed affected by the estimation error in the components. This is seen especially for the estimation of the more extreme quantiles, as predicted by the simplified considerations

above. For example, from Table 1, for the case of  $K = 8$  underlying functional components, the approach by Cardot *et al.* is found to perform better than the proposed method for the median, and comes reasonably close although is slightly worse for the first and third quartile. For more extreme quantiles, it performs much worse than the proposed method. The comparisons for the densely sampled case are visualized in Figure 2. We note that the more extreme quantiles are of particular interest for growth charts applications, and for their estimation the proposed method is seen to be particularly advantageous.

The comparative performance of the approach of Cardot *et al.* worsens considerably in the sparse case. This case is typical for longitudinal studies, including longitudinal growth studies. The likely reason for this worsening is that this approach has been designed for densely observed functional data and is negatively affected by the increased estimation error of the B-spline coefficients that comes with the longitudinal design case, due to the sparse and noisy measurements. The proposed method is based on Principal Analysis by Conditional Expectation, a functional method that is geared towards recovering the underlying random functions well especially for the case of sparsely sampled functional data.

In the second simulation scenario the functional predictors correspond to a mixture of Gaussian random processes. In this case the quantiles are not linear, and so it is not surprising that linear quantile regression does not work very well, not only because of the estimation error in the predictor components, but also because of the violation of a central model assumption. In this situation, the proposed method outperforms the approach of Cardot *et al.* for all quantiles in both densely and sparsely sampled cases (Table 2 and Figure 2 right panel).

The proposed method relies on a functional generalized regression model with a fixed link function to model binary responses with the mean parameter  $F(y | X)$ , and also relies on some other model assumptions. In some situations, links other than the logit, e.g., nonsymmetric links, may be of interest. The varying coefficient setting provides considerable flexibility to model the conditional distribution. A completely nonparametric approach seems infeasible for functional predictors due to the low values of the small ball probabilities in function space (Hall *et al.*, 2009). The proposed functional generalized regression framework provides a sensible

balance between model flexibility and the curse of dimensionality.

In summary, the proposed method shows promise for quantile estimation in functional data analysis. It is particularly advantageous when data are longitudinally sampled, with relatively sparse and noise-contaminated measurements, and for situations where quantiles are not linear in the predictors. Since it performs relatively well for the estimation of quantiles away from the median, the proposed method is particularly suited for growth studies and other situations where functional regression quantiles beyond the median are of interest.

## Appendix: Auxiliary results and proofs

We begin by stating an auxiliary lemma.

**Lemma.** *Let  $W_j$ ,  $j = 1, \dots, n$ , be a random sample and  $h(w, \theta)$  a Borel measurable function on  $\mathbf{W} \times \Theta$ , where  $\mathbf{W} \subset \mathbb{R}^k$  is a Borel set such that  $P[W_j \in \mathbf{W}] = 1$ , and  $\Theta$  is a compact subset of  $\mathbb{R}^m$ , such that for each  $w \in \mathbf{W}$ ,  $h(w, \theta)$  is a continuous function on  $\Theta$ . Furthermore, let  $E[\sup_{\theta \in \Theta} |h(W_j, \theta)|] < \infty$ . Then*

$$\sup_{\theta \in \Theta} \left| (1/n) \sum_{j=1}^n h(W_j, \theta) - E[h(W_1, \theta)] \right| \xrightarrow{P} 0, \text{ as } n \rightarrow \infty$$

This is usually called the uniform weak law of large numbers, and details and proofs can be found in Jennrich (1969) and Bierens (2004) (Appendix to Chapter 6).

### Proof of the Theorem.

By (A.1) and (A.2), the difference between the true  $F(y | X)$  in eq. (2) and the estimator  $\hat{F}(y | X)$  in eq. (5), for any fixed  $p$ , is bounded by

$$\begin{aligned} \sup_{y \in I} |F(y | X) - \hat{F}(y | X)| &< C \sup_{y \in I} \left| \sum_{k=0}^p \hat{\beta}_k(y) \xi_k - \sum_{k=0}^p \beta_k(y) \xi_k \right| + C \sup_{y \in I} \left| \sum_{k=p+1}^{\infty} \beta_k(y) \xi_k \right| \\ &< C \sup_{y \in I} \|\beta^p(y) - \hat{\beta}^p(y)\|_2 \|\xi^p\|_2 + C \sup_{y \in I} \left| \sum_{k=p+1}^{\infty} \beta_k(y) \xi_k \right|, \end{aligned}$$

where  $\xi^p = (\xi_0, \dots, \xi_p)^T$ ,  $\beta^p(y) = (\beta_0(y), \dots, \beta_p(y))$  and  $\hat{\beta}^p(y) = (\hat{\beta}_0(y), \dots, \hat{\beta}_p(y))$ .

Now assume that  $\epsilon > 0$ , and  $\delta > 0$  are given. Owing to the fact that  $\text{var}(\|\xi^p\|_2) < 1 + \sum_{k=1}^{\infty} \lambda_k \equiv M$ , one can find  $C_\delta > 0$  such that  $M/C_\delta^2 < \delta/3$ . By Chebyshev's inequality,

$P(\|\xi^p\|_2 > C_\delta) < \delta/3$ , for all  $p$ . If one can find  $p_{\epsilon,\delta}$  and  $n(\epsilon, \delta, p_{\epsilon,\delta})$  such that

$$P(\sup_{y \in I} \left| \sum_{k=p_{\epsilon,\delta}+1}^{\infty} \beta_k(y) \xi_k \right| > \frac{\epsilon}{2C}) < \delta/3, \quad (14)$$

and

$$P(\sup_{y \in I} \|\beta^{p_{\epsilon,\delta}}(y) - \hat{\beta}^{p_{\epsilon,\delta}}(y)\|_2 > \frac{\epsilon}{2C_\delta C}) < \delta/3, \quad \text{for } n > n(\epsilon, \delta, p_{\epsilon,\delta}), \quad (15)$$

then for all  $n > n(\epsilon, \delta, p_{\epsilon,\delta})$ ,

$$\begin{aligned} & P(\sup_{y \in I} |F(y | X) - \hat{F}(y | X)| > \epsilon) \\ & \leq P(\{C \sup_{y \in I} \left| \sum_{k=0}^{p_{\epsilon,\delta}} \hat{\beta}_k(y) \xi_k - \sum_{k=0}^{p_{\epsilon,\delta}} \beta_k(y) \xi_k \right| + C \sup_{y \in I} \left| \sum_{k=p_{\epsilon,\delta}+1}^{\infty} \beta_k(y) \xi_k \right|\} > \epsilon) \\ & \leq P(\sup_{y \in I} \|\beta^{p_{\epsilon,\delta}}(y) - \hat{\beta}^{p_{\epsilon,\delta}}(y)\|_2 > \frac{\epsilon}{2C_\delta C}) + P(\|\xi^{p_{\epsilon,\delta}}\|_2 > C_\delta) + P(\sup_{y \in I} \left| \sum_{k=p_{\epsilon,\delta}+1}^{\infty} \beta_k(y) \xi_k \right| > \frac{\epsilon}{2C}) \\ & < \delta/3 + \delta/3 + \delta/3 = \delta, \end{aligned}$$

which will complete the proof.

It remains to show (14) and (15). For any  $y \in I$ , let

$$\eta_p(y) = \beta_0(y) + \sum_{k=1}^p \beta_k(y) \xi_k, \quad R_p(y) = \sum_{k=p+1}^{\infty} \beta_k(y) \xi_k, \quad Z(y) = I(Y \leq y), \quad X_p = \mu + \sum_{k=1}^p \xi_k \phi_k.$$

According to model (2), we have  $F(y | X) = E(Z(y) | X) = g^{-1}(\eta_p(y) + R_p(y))$ , where assumption (A.3) implies that  $R_p(y)$  and  $X_p$  are independent. Let  $F_{R_p(y)}$  be the distribution function of  $R_p(y)$ . Then

$$E(Z(y) | X_p) = E(E(Z(y) | X) | X_p) = E(g^{-1}(\eta_p(y) + R_p(y)) | X_p) = \int g^{-1}(\eta_p(y) + s) dF_{R_p(y)}(s),$$

which demonstrates that  $E(Z(y) | X_p)$  depends on  $X_p$  only through the linear part  $\eta_p(y)$ .

Therefore,  $E(Z(y) | X_p) = g_{p,y}^{-1}(\eta_p(y))$ , for  $g_{p,y}^{-1}(\theta) = \int g^{-1}(\theta + s) dF_{R_p(y)}(s)$ . Furthermore,

$$\text{var}(Z(y) | X_p) = g_{p,y}^{-1}(\eta_p(y))(1 - g_{p,y}^{-1}(\eta_p(y))).$$

Then  $Z(y)$ , conditional on  $X_p$ , has a Bernoulli distribution with mean parameter  $E(Z(y) | X_p) = g_{p,y}^{-1}(\sum_{k=0}^p \beta_k(y) \xi_k)$ . Following the properties of a likelihood function as in eq. (12), we

find that for any fixed  $p$ ,

$$\boldsymbol{\beta}^p(y) = (\beta_0(y), \dots, \beta_p(y)) = \arg \min_{(\tilde{\beta}_0, \dots, \tilde{\beta}_p)} EL(g_{p,y}^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_k), Z(y)). \quad (16)$$

Since for any  $\tilde{\eta}_p = \sum_{k=0}^p \tilde{\beta}_k \xi_k$ , one has  $g_{p,y}^{-1}(\tilde{\eta}_p) = \int g^{-1}(\tilde{\eta}_p + s) dF_{R_p(y)}(s)$ , we conclude

$$\begin{aligned} & |E(g^{-1}(\tilde{\eta}_p) - g_{p,y}^{-1}(\tilde{\eta}_p))| \leq |E\{ \int (g^{-1}(\tilde{\eta}_p) - g^{-1}(\tilde{\eta}_p + s)) dF_{R_p(y)}(s) \}| \\ & \leq E\{ \int |g^{-1}(\tilde{\eta}_p) - g^{-1}(\tilde{\eta}_p + s)| dF_{R_p(y)}(s) \} \leq \int (g^{-1})'(\zeta) |s| dF_{R_p(y)}(s) \\ & \leq CE(|R_p(y)|) \leq C\{ \sum_{j=p+1}^{\infty} \beta_j(y)^2 \sum_{j=p+1}^{\infty} \lambda_j \}^{1/2}, \end{aligned}$$

where  $R_p = \sum_{j=p+1}^{\infty} \beta_j(y) \xi_j$ , and (A.2) has been used.

Observing that  $L(\theta, u)$  in (12) is Lipschitz continuous in  $\theta$ , for  $\theta \in [\zeta, 1 - \zeta]$ , where  $\zeta$  is as in (A.1), one finds

$$\begin{aligned} & |EL(g_{p,y}^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_k), Z(y)) - EL(g^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_k), Z(y))| \\ & \leq C_1 |E(g^{-1}(\tilde{\eta}_p) - g_{p,y}^{-1}(\tilde{\eta}_p))| \leq C_1 C \{ \sum_{j=p+1}^{\infty} \beta_j(y)^2 \sum_{j=p+1}^{\infty} \lambda_j \}^{1/2}, \end{aligned} \quad (17)$$

where  $C_1$  is a constant which only depends on the link function  $g$  and the domain  $I$ . Using assumption (A.5) and the fact that  $\sum_{j=1}^{\infty} \lambda_j < \infty$ , we can find  $p = p_{\epsilon, \delta}$  such that both

$$C_1 C \frac{\{ \sum_{j=p_{\epsilon, \delta}+1}^{\infty} \beta_j(y)^2 \sum_{j=p_{\epsilon, \delta}+1}^{\infty} \lambda_j \}^{1/2}}{\gamma_{p_{\epsilon, \delta}}} < \frac{\epsilon}{4C_{\delta}C}, \quad \text{for any } y \in I, \quad (18)$$

and

$$\{ \sum_{j=p_{\epsilon, \delta}+1}^{\infty} \beta_j(y)^2 \sum_{j=p_{\epsilon, \delta}+1}^{\infty} \lambda_j \}^{1/2} < \frac{\delta}{3} \frac{\epsilon^2}{4C^2}, \quad \text{for any } y \in I, \quad (19)$$

are satisfied.

Therefore, by assumption (A.4), eq. (17) and eq. (18),

$$\|(\boldsymbol{\beta}^{p_{\epsilon, \delta}})^*(y) - \boldsymbol{\beta}^{p_{\epsilon, \delta}}(y)\|_2 < C_1 C \frac{\{ \sum_{j=p_{\epsilon, \delta}+1}^{\infty} \beta_j(y)^2 \sum_{j=p_{\epsilon, \delta}+1}^{\infty} \lambda_j \}^{1/2}}{\gamma_{p_{\epsilon, \delta}}} < \frac{\epsilon}{4C_{\delta}C} \quad \text{for any } y \in I, \quad (20)$$

where  $\boldsymbol{\beta}^p(y)$  is as in eq. (16) and  $(\boldsymbol{\beta}^p)^*(y)$  is defined as

$$(\boldsymbol{\beta}^p)^*(y) = \arg \min_{\tilde{\beta}_0, \dots, \tilde{\beta}_p} EL(g^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_k), Z(y)). \quad (21)$$

To connect to the sample minimization problem (11), we use a uniform weak law of large numbers as in the Lemma. First, we verify the following condition: For a fixed  $p$ ,

$$E\left[\sup_{(\tilde{\beta}_0, \dots, \tilde{\beta}_p) \in B_p, y \in I} |L(g^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_k), Z(y))|\right] < \infty,$$

where  $B_p$  is a compact subset of  $\mathbb{R}^p$  and  $\beta^p(y) \in B_p$ , for all  $y \in I$ . Note that

$$\begin{aligned} & E\left[\sup_{(\tilde{\beta}_0, \dots, \tilde{\beta}_p) \in B, y \in I} |L(g^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_k), Z(y))|\right] \\ & < 2E[\sup |\log(1 - g^{-1}(\tilde{\eta}))|] + E[\sup |\log(\frac{g^{-1}(\tilde{\eta})}{1 - g^{-1}(\tilde{\eta})})|] = I + II, \end{aligned}$$

where  $\tilde{\eta} = \sum_{k=0}^p \tilde{\beta}_k \xi_k$ . Using Taylor expansions of  $I$  and  $II$  over  $\tilde{\eta}$ , one finds that under condition (A.2),

$$\begin{aligned} & E\left[\sup_{(\tilde{\beta}_0, \dots, \tilde{\beta}_p) \in B_p, y \in I} |L(g^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_k), Z(y))|\right] \\ & < C' E\left[\sup_{(\tilde{\beta}_0, \dots, \tilde{\beta}_p) \in B_p, y \in I} |\sum_{k=0}^p \tilde{\beta}_k \xi_k|\right] < C_p \sum_{k=1}^p \lambda_k < \infty, \end{aligned}$$

where  $C_p$  is a finite constant depending on  $B_p$ . Therefore, according to the Lemma, for any fixed  $p$ ,

$$\sup_{(\tilde{\beta}_0(y), \dots, \tilde{\beta}_p(y)) \in B_p, y \in I} \left| \left(1/n\right) \sum_{i=1}^n L(g^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_{ik}), Z_i(y)) - E[L(g^{-1}(\sum_{k=0}^p \tilde{\beta}_k \xi_k), Z(y))] \right| \xrightarrow{P} 0. \quad (22)$$

Combining (22) with condition (A.4), for  $p = p_{\epsilon, \delta}$ , one can find  $n(\epsilon, \delta, p_{\epsilon, \delta})$  such that for all  $n \geq n(\epsilon, \delta, p_{\epsilon, \delta})$ ,

$$P(\sup_{y \in I} \|(\beta^{p_{\epsilon, \delta}})^*(y) - \hat{\beta}^{p_{\epsilon, \delta}}(y)\|_2 > \frac{\epsilon}{4C_{\delta}C}) < \delta/3, \quad (23)$$

where  $(\beta^p)^*(y)$  is defined in eq. (21) and  $\hat{\beta}^p(y)$  is in eq. (11). Combining (20) and (23), eq. (15) follows and combining Chebyshev's inequality and (19) leads to eq. (14), completing the proof.

## References

Besse, P., Cardot, H. and Ferraty, F. (1997) Simultaneous nonparametric regression of unbalanced longitudinal data. *Computational Statistics & Data Analysis*, **24**, 255–270.

- Bierens, H. J. (2004) *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge University Press.
- Cai, Z. (2002) Regression quantiles for time series. *Econometric Theory*, **18**, 169–192.
- Cardot, H., Crambes, C. and Sarda, P. (2005) Quantile regression when the covariates are functions. *Journal of Nonparametric Statistics*, **17**, 841–856.
- (2007) Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates. In *Statistical Methods for Biostatistics and Related Fields*. (eds. Y. Har-dle, W. and Mori and P. Vieu), 221–244. Springer Verlag, New York.
- Cardot, H. and Sarda, P. (2005) Estimation in generalized linear models for functional data via penalized likelihood. *Journal of Multivariate Analysis*, **92**, 24–41.
- Cole, T. (1988) Fitting smoothed centile curves to reference data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **151**, 385–418.
- (1994) Growth charts for both cross-sectional and longitudinal data. *Statistics in Medicine*, **13**, 2477–2492.
- Ellozy, M. (1978) Critical analysis of double and triple logistic growth-curves. *Annals of Human Biology*, **5**, 389–394.
- Escabias, M., Aguilera, A. M. and Valderrama, M. J. (2007) Functional PLS logit regression model. *Computational Statistics & Data Analysis*, **51**, 4891–4902.
- Fan, J., Yao, Q. and Tong, H. (1996) Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83**, 189–206.
- Ferraty, F., Rabhi, A. and Vieu, P. (2005) Conditional quantiles for dependent functional data with application to the climatic EL Niño phenomenon. *Sankhyā: Special Issue on Quantile Regression and Related Methods*, **67**, 378–398.

- Gasser, T. and Kneip, A. (1995) Searching for structure in curve samples. *Journal of the American Statistical Association*, **90**, 1179–1188.
- Gasser, T., Kneip, A., Binding, A., Prader, A. and Molinari, L. (1991) The dynamics of linear growth in distance, velocity and acceleration. *Annals of Human Biology*, **18**, 187–205.
- Gasser, T., Müller, H.-G., Köhler, W., Molinari, L. and Prader, A. (1984) Nonparametric regression analysis of growth curves. *The Annals of Statistics*, **12**, 210–229.
- Gasser, T., Müller, H.-G., Köhler, W., Prader, A., Largo, R. and Molinari, L. (1985) An analysis of the mid-growth spurt and of the adolescent growth spurt based on acceleration. *Annals of Human Biology*, **12**, 129–148.
- Hall, P. and Müller, H.-G. (2003) Order-preserving nonparametric regression, with applications to conditional distribution and quantile function estimation. *Journal of the American Statistical Association*, **98**, 598–608.
- Hall, P., Müller, H.-G. and Yao, F. (2009) Estimation of functional derivatives. *The Annals of Statistics*, **37**, 3307–3329.
- Hall, P., Wolff, R. C. and Yao, Q. (1999) Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, **94**, 154–163.
- James, G. (2002) Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 411–432.
- James, G. M., Hastie, T. J. and Sugar, C. A. (2000) Principal component models for sparse functional data. *Biometrika*, **87**, 587–602.
- Jennrich, R. I. (1969) Asymptotic properties of non-linear least square estimators. *Annals of Mathematical Statistics*, **40**, 633–643.
- Kneip, A. and Gasser, T. (1992) Statistical tools to analyze data representing a sample of curves. *The Annals of Statistics*, **20**, 1266–1350.

- Koenker, R. (2005) *Quantile Regression*. Cambridge University Press.
- Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica*, **46**, 33–50.
- Koenker, R., Ng, P. and Portnoy, S. (1994) Quantile smoothing splines. *Biometrika*, **81**, 673–680.
- Leng, X. and Müller, H.-G. (2006) Classification using functional data analysis for temporal gene expression data. *Bioinformatics*, **22**, 68–76.
- Li, Y., Liu, Y. and Zhu, J. (2007) Quantile regression in reproducing kernel hilbert spaces. *Journal of the American Statistical Association*, **102**, 255–268.
- Marubini, E., Resele, L. and Barghini, G. (1971) Comparative fitting of Gompertz and logistic functions to longitudinal height data during adolescence in girls. *Human Biology*, **43**, 237.
- Müller, H.-G. (2005) Functional modelling and classification of longitudinal data. *Scandinavian Journal of Statistics. Theory and Applications*, **32**, 223–240.
- Müller, H.-G. and Stadtmüller, U. (2005) Generalized functional linear models. *The Annals of Statistics*, **33**, 774–805.
- Müller, H.-G. and Zhang, Y. (2005) Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics*, **61**, 1064–1075.
- Rice, J. A. (2004) Functional and longitudinal data analysis: Perspectives on smoothing. *Statistica Sinica*, 631–647.
- Rice, J. A. and Wu, C. O. (2001) Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**, 253–259.
- Royston, P. (1995) Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal measurements. *Statistics in Medicine*, **14**, 1417–1436.
- Sheehy, A., Gasser, T., Molinari, L. and Largo, R. (1999) An analysis of variance of the pubertal and midgrowth spurts for length and width. *Annals of Human Biology*, **26**, 309–331.

- Shi, M., Weiss, R. E. and Taylor, J. M. G. (1996) An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **45**, 151–163.
- Thompson, M. and Fatti, L. (1997) Construction of multivariate centile charts for longitudinal measurements. *Statistics in Medicine*, **16**, 333–345.
- Tuddenham, R. and Snyder, M. (1954) Physical growth of California boys and girls from birth to age 18. *Calif. Publ. Child Deve.*, **1**, 183–364.
- Wei, Y. and He, X. (2006) Conditional growth charts. *The Annals of Statistics*, **34**, 2069–2097.
- Yao, F., Müller, H.-G. and Wang, J.-L. (2005) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577–590.
- Yu, K. and Jones, M. C. (1998) Local linear quantile regression. *Journal of the American Statistical Association*, **93**, 228–237.
- Zhao, X., Marron, J. S. and Wells, M. T. (2004) The functional data analysis view of longitudinal data. *Statistica Sinica*, **14**, 789–808.

Table 1: Simulation results for mean absolute error (MAE), based on 200 simulation runs, for estimating quantiles at level  $\alpha = \{0.05, 0.1, 0.25, 0.5\}$ , for Gaussian data under sparse and dense scenarios. Methods compared are the proposed conditional functional quantiles, functional quantile linear regression (Cardot *et al.*, 2005), and unconditional empirical quantiles. The best results for each scenario are bolded;  $K$  is the number of basis functions used to generate the predictor functions. For the proposed method, the number of included functional principal components is selected data-adaptively by the AIC criterion.

	Method	$\alpha = 0.05$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.50$	
		Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense
$K = 2$	Proposed	<b>1.95</b>	<b>1.57</b>	<b>1.75</b>	<b>1.39</b>	<b>1.62</b>	<b>1.25</b>	<b>1.59</b>	1.22
	Cardot et al.	6.85	5.89	6.23	4.46	5.63	2.32	5.40	<b>0.89</b>
	Unconditional	11.89	11.76	10.42	10.37	8.69	8.68	8.02	8.07
$K = 4$	Proposed	<b>3.56</b>	<b>1.79</b>	<b>3.27</b>	<b>1.64</b>	<b>3.03</b>	<b>1.51</b>	<b>2.98</b>	1.46
	Cardot et al.	8.37	5.93	8.09	4.50	7.85	2.38	7.79	<b>1.15</b>
	Unconditional	15.21	15.36	13.20	13.31	10.77	10.90	9.77	9.90
$K = 6$	Proposed	<b>5.75</b>	<b>2.11</b>	<b>5.18</b>	<b>1.97</b>	<b>4.68</b>	<b>1.85</b>	<b>4.54</b>	1.80
	Cardot et al.	9.11	5.87	8.90	4.42	8.71	2.41	8.62	<b>1.38</b>
	Unconditional	16.48	16.28	14.22	14.11	11.40	11.42	10.33	10.39
$K = 8$	Proposed	<b>6.72</b>	<b>2.59</b>	<b>5.98</b>	<b>2.45</b>	<b>5.32</b>	<b>2.32</b>	<b>5.14</b>	2.28
	Cardot et al.	9.36	5.82	9.13	4.41	8.96	2.53	8.92	<b>1.61</b>
	Unconditional	16.64	16.74	14.43	14.37	11.73	11.67	10.56	10.56
$K = 10$	Proposed	<b>6.80</b>	<b>2.62</b>	<b>5.74</b>	<b>2.50</b>	<b>5.40</b>	<b>2.38</b>	<b>5.16</b>	2.28
	Cardot et al.	9.36	5.79	9.15	4.38	8.99	2.54	8.94	<b>1.71</b>
	Unconditional	16.82	16.84	14.50	14.57	11.79	11.76	10.65	10.54

Table 2: Simulation results for mean absolute error (MAE), based on 200 simulation runs, for estimating quantiles at level  $\alpha = \{0.05, 0.1, 0.25, 0.5\}$ , for Gaussian mixture data under sparse and dense scenarios. Methods compared are the proposed conditional functional quantiles, functional quantile linear regression (Cardot *et al.*, 2005), and unconditional empirical quantiles. The best results for each scenario are bolded;  $K$  is the number of basis functions used to generate the predictor functions. For the proposed method, the number of included functional principal components is selected data-adaptively by the AIC criterion.

	Method	$\alpha = 0.05$		$\alpha = 0.10$		$\alpha = 0.25$		$\alpha = 0.50$	
		Sparse	Dense	Sparse	Dense	Sparse	Dense	Sparse	Dense
$K = 2$	Proposed	<b>2.84</b>	<b>2.45</b>	<b>2.24</b>	<b>1.91</b>	<b>2.22</b>	<b>1.95</b>	<b>1.83</b>	<b>1.64</b>
	Cardot et al.	7.46	5.08	6.88	4.32	5.96	3.29	3.95	2.01
	Unconditional	14.88	14.73	11.65	11.53	8.57	8.54	5.57	5.56
$K = 4$	Proposed	<b>4.88</b>	<b>2.90</b>	<b>3.89</b>	<b>2.31</b>	<b>3.59</b>	<b>2.57</b>	<b>2.91</b>	<b>2.09</b>
	Cardot et al.	9.39	5.57	8.90	4.92	8.05	4.02	5.64	2.69
	Unconditional	18.91	18.85	14.36	14.55	10.49	10.74	6.76	6.96
$K = 6$	Proposed	<b>7.24</b>	<b>3.37</b>	<b>5.90</b>	<b>2.79</b>	<b>5.09</b>	<b>2.96</b>	<b>3.92</b>	<b>2.34</b>
	Cardot et al.	10.25	5.69	9.76	5.09	8.90	4.29	6.26	2.86
	Unconditional	19.48	19.94	15.16	15.35	11.27	11.32	7.24	7.30
$K = 8$	Proposed	<b>8.35</b>	<b>3.84</b>	<b>6.79</b>	<b>3.30</b>	<b>5.75</b>	<b>3.30</b>	<b>4.32</b>	<b>2.58</b>
	Cardot et al.	10.67	5.89	10.16	5.31	9.27	4.53	6.52	3.01
	Unconditional	20.14	20.22	15.46	15.42	11.52	11.44	7.42	7.40
$K = 10$	Proposed	<b>8.40</b>	<b>3.82</b>	<b>6.84</b>	<b>3.34</b>	<b>5.77</b>	<b>3.32</b>	<b>4.40</b>	<b>2.59</b>
	Cardot et al.	10.63	5.84	10.15	5.29	9.25	4.56	6.51	3.08
	Unconditional	20.45	20.52	15.68	15.71	11.54	11.52	7.44	7.43

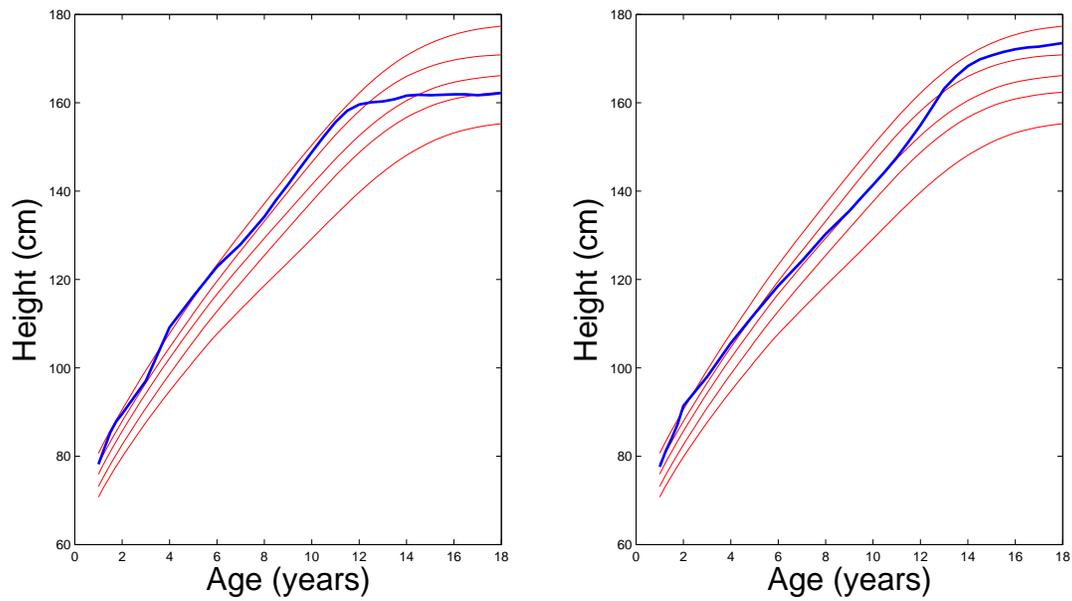


Figure 1: Conventional cross-sectional reference growth charts, for percentiles  $\alpha$  in  $\{0.05, 0.2, 0.5, 0.8, 0.95\}$ , overlaid with the growth curves of two girls from the Berkeley Longitudinal Growth Study.

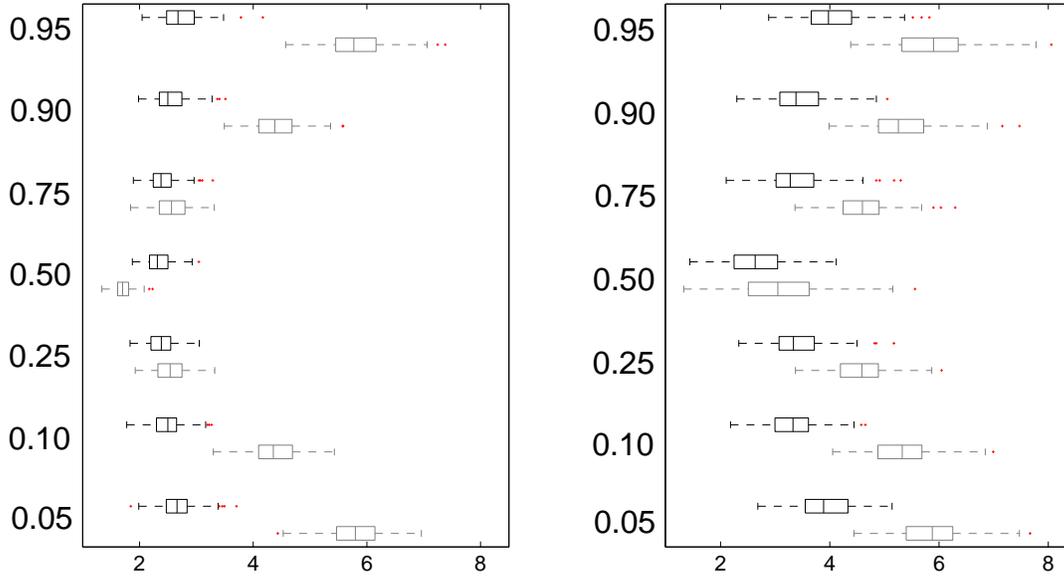


Figure 2: Boxplots of mean absolute error (MAE) over 200 simulation runs for conditional quantile estimation for  $\alpha \in \{0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95\}$ . The predictor trajectories are generated from 10 eigenfunctions and are densely sampled. Conditional distributions are either Gaussian (left) or follow a Gaussian Mixture (right). The gray boxplots correspond to the functional quantile linear regression approach (Cardot *et al.*, 2005) and the black boxplots to the proposed method.

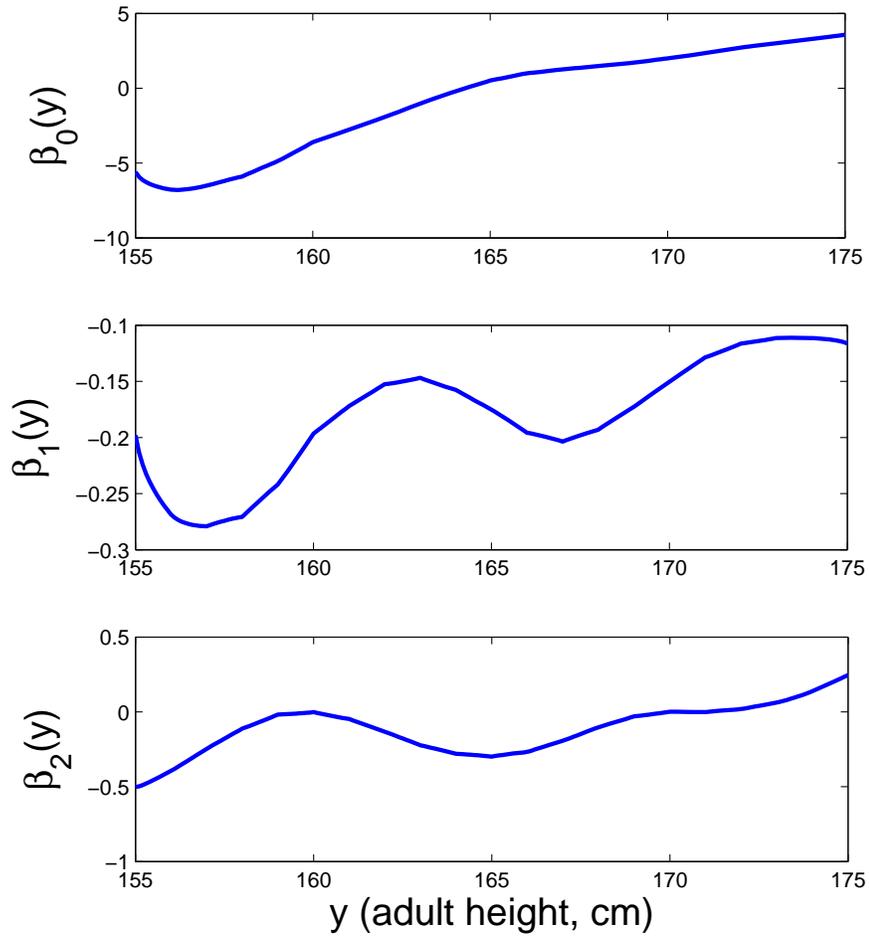


Figure 3: For the Berkeley growth data, a model with two functional principal components explains more than 95% of the variation. The corresponding three estimated coefficients functions  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\beta}_2$ , defined before eq. (5), are shown above.

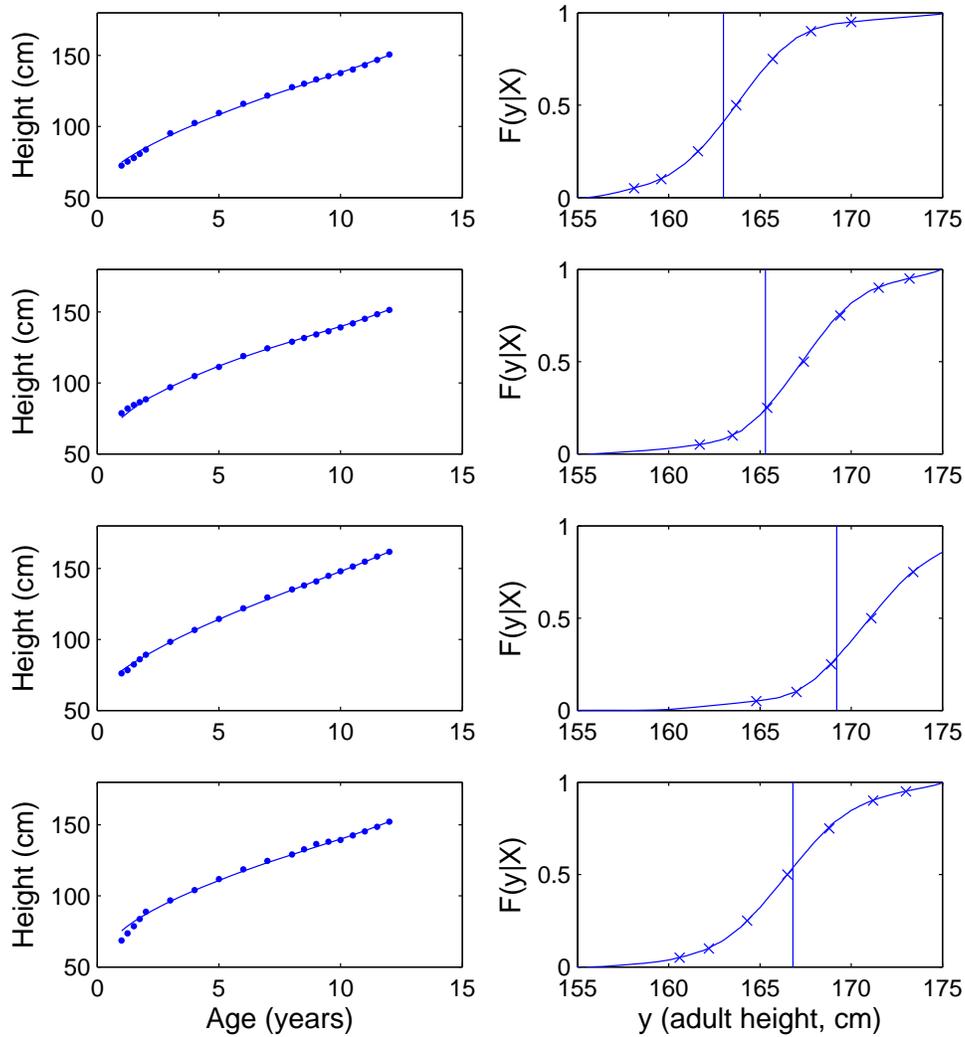


Figure 4: For four randomly selected girls, estimated growth paths obtained from functional principal component analysis (left panels) and predicted distribution functions for adult height, conditioning on the observed growth history in the interval  $[0, 12]$  (right panels). Estimated quantiles  $Q_i(\alpha)$  are marked by 'x', for  $\alpha \in [0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95]$ . The subsequently observed adult height is indicated by the vertical line.

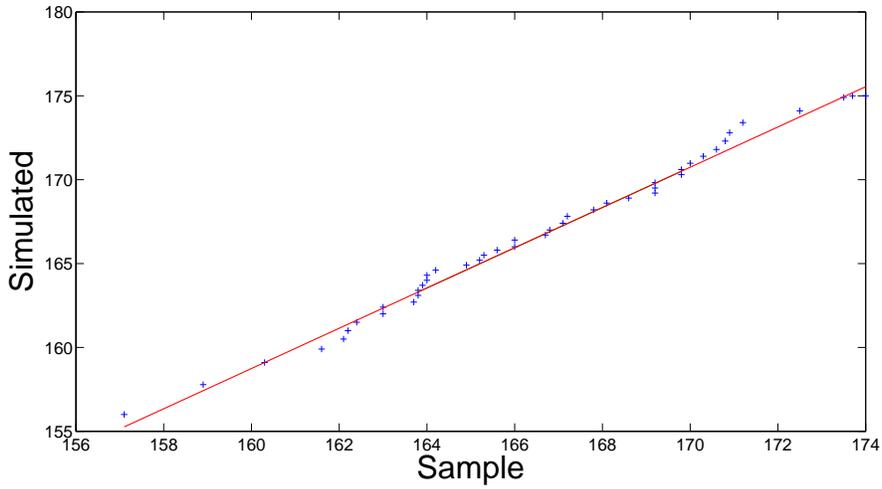
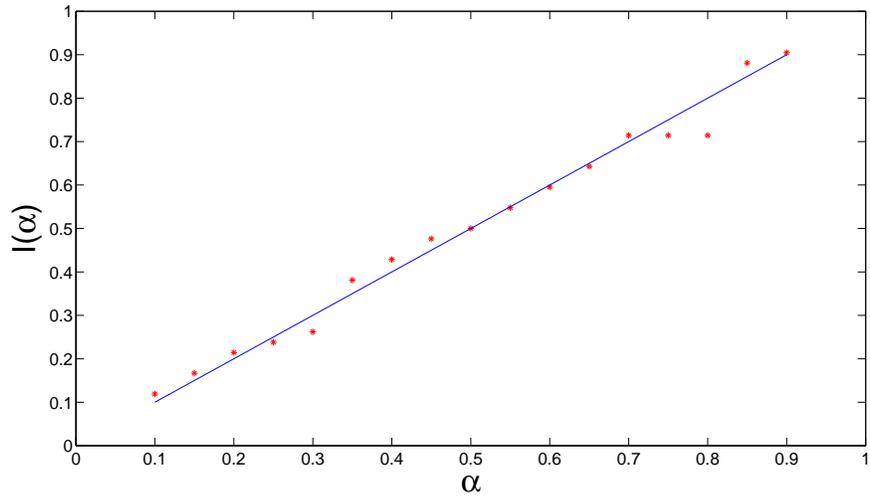


Figure 5: Diagnostic plots for the proposed functional conditional quantile estimates  $\hat{Q}_i(\alpha)$ , displaying the function  $\bar{I}(\alpha)$  (top panel, see eq. (10)) and the Q-Q plot of the empirical sample against the model based simulated data (bottom panel). Both plots should be close to the identity line if the estimation works well, and indeed there is no indication of lack of fit.

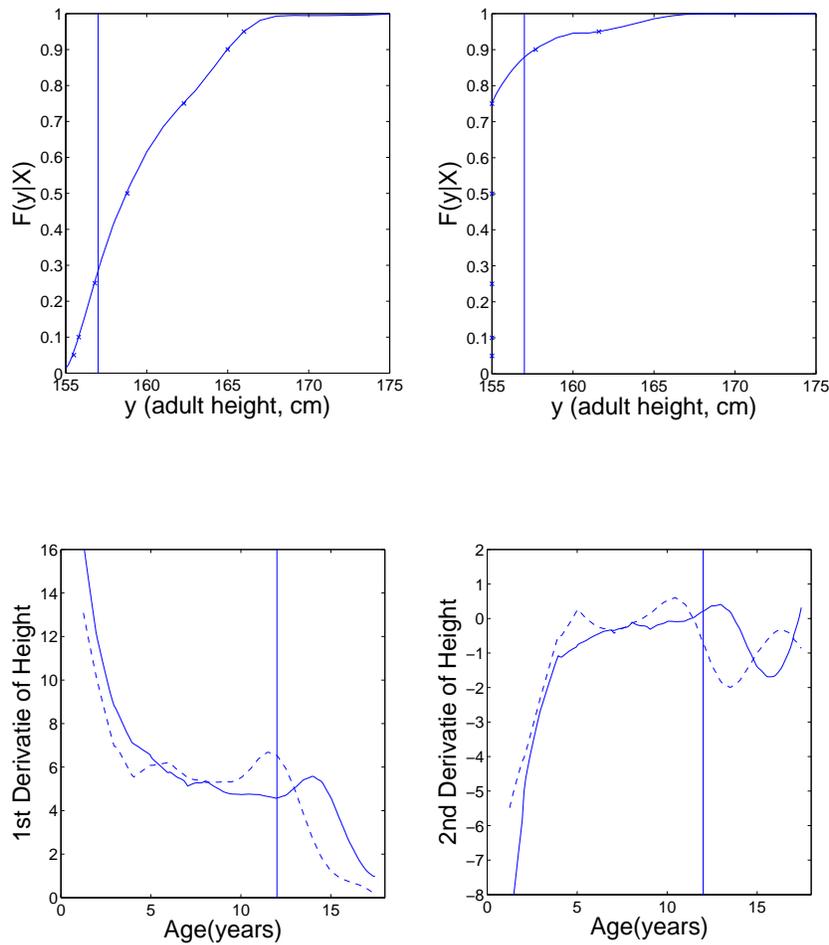


Figure 6: Estimated conditional distribution functions and quantiles  $Q_i(\alpha)$ , marked as 'x', for  $\alpha \in [0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95]$ , for adult height, based on measurements of height on  $[0, 12]$ , for girl A (left upper panel) and girl B (right upper panel), with heights at age 12 of 142.1 cm (girl A) and 141.3 cm (girl B), respectively, both below the 10th percentile of height at age 12. The vertical line indicates the 10th percentile of adult height in the sample. The estimated probability for Girl B for an adult height below the 10th quantile is about 0.9, while it is smaller than 0.3 for girl A. Adult height turned out to be 164 cm for girl A and 154.5 cm for girl B. The first (left lower panel) and second (right lower panel) derivative of the growth curves for girl A (solid line) and girl B (dashed line) show clear differences in the growth patterns of the two girls up to age 12.

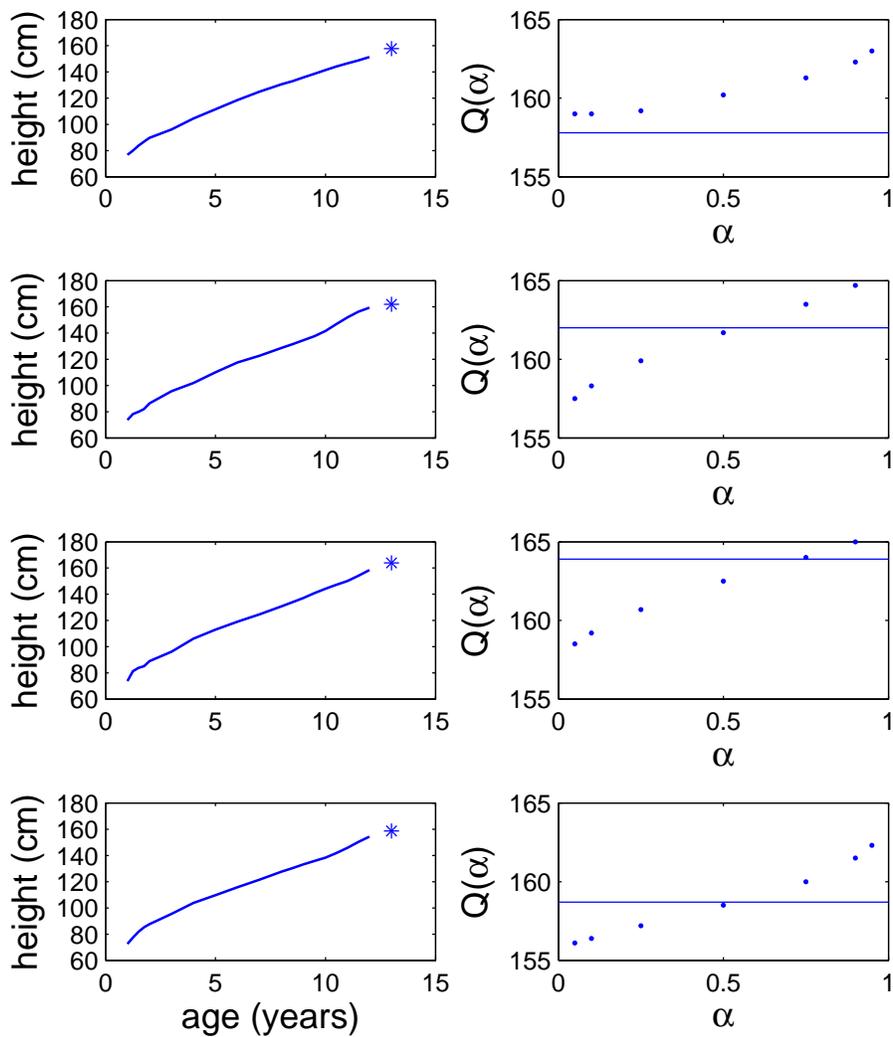


Figure 7: Growth curves of four randomly selected girls (girls C-F, from top to bottom) in the interval  $[0,12]$  years, with corresponding height at age 13 marked by an asterisk (left panels), and estimated conditional quantiles of height at age 13, obtained from model (4), for  $\alpha \in [0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95]$ , with a horizontal line indicating the actual height at age 13 (right panels).