Quantile regression for scalar and functional clustered data and data analysis with phase-amplitude separation

Maria Laura Battagliola

PhD Thesis

This thesis has been submitted to the PhD School of the Faculty of Science, University of Copenhagen

June 13, 2021

Department of Mathematical Sciences
Faculty of Science
University of Copenhagen

Maria Laura Battagliola
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
2100 København Ø
Denmark
mlbattagliola@math.ku.dk
mlbattagliola@gmail.com

Supervisor:	Prof. Helle Sørensen University of Copenhagen
Co-supervisor:	Associate Prof. Anders Tolver University of Copenhagen
Co-supervisor:	Prof. Ana-Maria Staicu North Carolina State University
Assessment committee:	Prof. Marco Geraci Sapienza - University of Rome
	Associate Prof. Lina Schelin Umeå University
	Associate Prof. Bo Markussen (chairman) University of Copenhagen

The PhD project was partly funded by the Danish Research Council (DFF grant 7014-00221).

ISBN: 978-87-7125-044-2

Abstract

In this thesis we present results arising from either quantile regression, functional data analysis or a combination of the two fields.

First of all, we study quantile regression for clustered data in the cases when the number of clusters is much larger than the observations per cluster. Via simulation studies we demonstrate that some classical estimators for the population-level quantile regression parameters exhibit bias when considering heteroskedastic models at quantile levels different from the median. We propose an estimator whose bias adjustment is based on bootstrap, which we also rely on in order to build confidence intervals. We apply the new estimation methods to data arising from a clinical study concerning AIDS.

We analyze the aforementioned framework further when functional covariates are introduced and data has a longitudinal structure. In particular, we establish the modelling setting, we propose an estimation method for the approximation of the functional coefficient, and we clearly outline how to implement estimation relying on existing software. Our work is motivated by an application in animal science, in which we study the impact of temperature, considered as functional, on low quantiles of feed intake of lactating sows, whose daily conditions were recorded several times over the lactating days, which we take as longitudinal time points.

Our last contribution concerns functional data analysis and revolves around the analysis of learning curves of mice undergoing memory-involving tasks repeatedly. We rely on existing methods that study bivariate functional objects constituted by amplitude and phase components arising from the registration of a collection of curves. The multivariate functional principal component analysis of such objects gives us an insight on the differences and similarities of the learning behaviors of two groups of mice, one where the animals were induced with a brain lesion similar to that observed in patients affected by psychiatric disorders such as schizophrenia, and a control group.

Resumé

I denne afhandling præsenterer vi resultater fra fraktilregression, funktionel dataanalyse eller en kombination af de to områder.

Første studie handler om fraktilregression for grupperede data i situationer hvor antallet af grupper (eller klynger) er meget større end antallet af observationer per gruppe. Vi undersøger estimation af populationsparametrene i en regressionsmodel for fraktilerne og påviser i simulationsstudier at flere metoder fra litteraturen fører til estimatorer med bias når data simuleres med heteroskedasticitet og analysen foretages på et andet fraktilniveau end 50% svarende til medianen. Vi udvikler og undersøger en ny estimationsmetode der justerer for bias ved hjælp af bootstrap, og de samme bootstrapdata bruges til at beregne konfidensintervaller. Vi benytter den nye estimationsmetode på data fra et klinisk studie vedrørende AIDS.

Andet studie udvider rammerne fra første studie til dels at omfatte funktionelle kovariater og dels at omfatte longitidinale data. Vi undersøger en klasse af regressionsmodeller for fraktilregression og viser hvordan modellen kan estimeres med eksisterende software. Artiklen er motiveret af data fra husdyrvidenskab om søer der giver die. Det ønskes undersøgt om og hvordan temperaturen i grisestien, målt kontinuerligt henover døgnet, påvirker søernes indtag af føde i den periode hvor de giver die.

Tredje studie er et studie i funktionel dataanalyse og omhandler estimation og sammenligning af indlæringskurver for forsøgsmus der udfører hukommelseskrævende opgaver gentagne gange. Analysen består i først først at registrere (tidsforskyde) kurverne således at fase- og amplitudevariation separares og estimeres og dernæst udførse todimensional principalkomponentanalyse af fase- og amplitudekomponenterne. Principalkomponentanalysen giver os indsigt i forskelle og ligheder i indlæring mellem to grupper af mus, nemlig kontrolmus og mus med en induceret hjernelæsion der svarer til hvad man kan observere hos patienter med psykiatriske sygdomme som fx skizofreni.

Acknowledgments

First of all, I would like to thank the members of the assessment committee for reading my work.

I consider myself lucky for being surrounded by great persons, both in my professional and private life. Among them, I definitely count Prof. Helle Sørensen, who is the supervisor one can only dream of having. She has been an example of mentorship ever since I visited her during my Master's, and I am extremely grateful for having had the chance to work with her during my PhD. I would also like to thank my co-supervisors, Prof. Anders Tolver and Prof. Ana-Maria Staicu for their guidance and the interesting conversations we have been having throughout the years. Moreover, I thank Prof. Staicu for hosting me twice in NCSU.

During my PhD studies, I had the chance to visit the Department of Biostatistics at Columbia University, where Prof. Todd Ogden warmly welcomed me. I take the chance to thank him for introducing me to the vibrant atmosphere of the department, as well as for dedicating time and interest to our discussions. I am also grateful to Prof. Jeff Goldsmith, Prof. Julia Wrobel and Erin McDonnell for their availability to answer my questions while there.

I would like to thank my friends and colleagues at the Department of Mathematics of the University of Copenhagen for making work truly enjoyable. I will cherish the memories of the cups of coffee, slices of cake, and all the other moments we had together.

I am very grateful for having been having old and new friends with me along the way. It is hard to imagine making it through the PhD without them, especially in the last and hardest moments. I thank them all, especially Angélica, Beatriz, Francesco, Giuliano, Laura and Roberta, for being there to cheer for me at the very end.

I thank Jorge from the bottom of my heart for being an outstanding partner, showing me nothing but kindness and understanding, and for having my back always. Finally, I am deeply grateful to my family and all those that I consider my family, who have been believing in me much before the start of the PhD. They have been an incredible source of support throughout the years, and they are all very precious to me.

Maria Laura Battagliola *Copenhagen, June* 2021.

Summary

The results of this PhD thesis concern two areas of statistics, namely quantile regression and functional data analysis, and they consist of the following manuscripts

A. Battagliola, M.L., Sørensen, H., Tolver, A., and Staicu, A.-M. A bias-adjusted estimator in quantile regression for clustered data.
Corrected proof available in Ecosta (https://doi.org/10.1016/j.ecosta.2021.07.003).

B. Battagliola, M.L., Sørensen, H., Tolver, A., and Staicu, A.-M. Quantile regression for longitudinal functional data with application to feed intake of lactating sows.
In progress.

C. Battagliola, M.L., Benoit, L.J., Canetta, S., Ogden, R.T. Analysis of learning curves of mice by phase-amplitude separation.
In progress.

Maria Laura Battagliola *Copenhagen, June* 2021.

Contents

Chapter 1

Introduction

The aim of this chapter is two-fold. First of all, it contains an overview of the essential mathematical ingredients used in the following chapters. In particular, this thesis has its foundation on two topics of statistics: quantile regression and functional data analysis, presented in Sections 1.1 and 1.2, respectively. Secondly, it summarises the contributions of our work to the aforementioned areas in Section 1.3.

1.1 Quantile regression

Quantile regression, first introduced by Koenker and Bassett Jr (1978), is a well-established mathematical tool used in econometrics and statistics. In this section we first of all present the derivation of the loss function used in such regression framework. Afterwards, an overview of quantile regression is given for independent observations. Finally, quantile regression for clustered data, which is a key topic for the work in Chapters 2 and 3, is discussed.

1.1.1 Origins of loss function

The loss function is a central feature to every regression framework. In what follows, we take inspiration from Koenker (2005a, Chapter 1) to explain what is the loss function used in quantile regression.
Consider a continuous random variable $Y \sim f$ taking values in $\mathbb{R}$. It's Cumulative Distribution Function (CDF) F is defined as

$$F(y) = P(Y \leq y),$$

with $y \in \mathbb{R}$. Moreover, the quantile of level $\tau \in (0,1)$ of Y is

$$Q(\tau) = F^{-1}(\tau) = inf\{y : F(y) \geq \tau\}.$$

From a probabilistic point of view, the quantiles of Y can be found by minimizing an expected loss function. First of all, for fixed $\tau \in (0,1)$, consider

$$\rho_\tau(v) = v(\tau - I_{(v<0)}). \tag{1.1.1}$$

This is called *check function*, and it is the essential pillar of quantile regression. The loss function that one aims at optimizing in order to find the τth quantile of Y is

$$E[\rho_\tau(Y - \hat{y})], \tag{1.1.2}$$

which is minimized with respect to $\hat{y} \in \mathbb{R}$. In particular, (1.1.2) corresponds to

$$\int_{\mathbb{R}} \rho_\tau(y - \hat{y}) dF(y) = \left[(\tau - 1) \int_{-\infty}^{\hat{y}} (y - \hat{y}) dF(y) + \tau \int_{\hat{y}}^{\infty} (y - \hat{y}) dF(y) \right].$$

Taking

$$\frac{d}{d\hat{y}}E[\rho_\tau(Y-\hat{y})] = -(\tau-1)\int_{-\infty}^{\hat{y}} dF(y) - \tau\int_{\hat{y}}^{\infty} dF(y) = \int_{-\infty}^{\hat{y}} dF(y) - \tau = F(\hat{y}) - \tau,$$

and imposing $\frac{d}{d\hat{y}}E[\rho_\tau(Y-\hat{y})] = 0$, the solution is $\hat{y} = F^{-1}(\tau)$ when F is strictly monotone, otherwise, by convention, it is the infimum of the interval of values minimizing (1.1.2).

In practice, one observes independent realizations $Y_1, \dots, Y_N$ of Y, and hence relies on the empirical CDF

$$F_N(y) = \frac{1}{N}\sum_{i=1}^{N} I(Y_i \leq y).$$

In such case, the empirical version of (1.1.2) becomes

$$\int_{\mathbb{R}} \rho_\tau(y-\hat{y})dF_N(y) = \frac{1}{N}\sum_{i=1}^{N}\rho_\tau(Y_i-\hat{y}).$$

Hence, the problem we consider is the minimization of

$$\sum_{i=1}^{N}\rho_\tau(Y_i-\hat{y}) \tag{1.1.3}$$

with respect to $\hat{y}$ to estimate τth quantiles of Y.
The aforementioned preliminaries are particularly important to quantile regression, which we review in the next paragraphs.

1.1.2 Quantile regression for independent observations

Consider data $(Y_i, x_i)_{i=1}^N$, with response $Y_i \in \mathbb{R}$ and vector of covariates $x_i \in \mathbb{R}^{p-1}$. Relying on (1.1.3), it is possible to build a regression framework to estimate the quantile of the distribution of the response given the independent variables. Assuming a linear structure of such quantile, we consider the following model for the ith observation and fixed level τ:

$$Q_{Y_i|X_i}(\tau) = X_i^T\beta^\tau, \tag{1.1.4}$$

with $X_i^T = (1, x_i^T)$ and $\beta^\tau = (\beta^{\tau,1}, \dots, \beta^{\tau,p}) \in \mathbb{R}^p$. The estimator $\hat{\beta}^\tau$ of the quantile regression coefficients is thus obtained by minimizing

$$\sum_{i=1}^{N}\rho_\tau(Y_i - X_i^T\beta^\tau), \tag{1.1.5}$$

with respect to β^τ. Due to the non-differentiability of (1.1.1) at $v = 0$, the minimization of the loss function in quantile regression has to be handled with caution. The optimization problem can be solved computing the directional derivatives and then evaluate them in $\hat{\beta}$. If for every direction they are non-negative, then $\hat{\beta}$ minimizes the loss function (Koenker, 2005a, Chapter 1) .

The minimization of a loss function that is not everywhere differentiable is not the only challenge in quantile regression. We take the opportunuty of introducing the problem of quantile crossing with the following example.

Example 1.1 (Location and location-shift models)**.** *Consider data* $(Y_i, x_i)_{i=1}^N$*, with* $Y_i, x_i \in \mathbb{R}$ *and*

$$Y_i = \beta_0 + \beta_1 x_i + (1 + \gamma x_i)\epsilon_i, \qquad i = 1, \dots, N,$$

with $\gamma \geq 0$ *,* $(1 + \gamma x_i) > 0$ *for every* $i = 1, \dots, N$ *and* $\epsilon \overset{iid}{\sim} f_\epsilon$*. In order to find the corresponding quantile at level* τ*, we start from*

$$P(Y_i \leq Q_{Y_i|X_i}(\tau)) = \tau,$$

which, explicitly expressing the response in terms of the independent variable and isolating the error term, becomes

$$P\left(\epsilon_i \leq \frac{Q_{Y_i|X_i}(\tau) - \beta_0 - \beta_1 x_i}{1 + \gamma x_i}\right) = \tau.$$

The, the quantity we are looking for is

$$Q_{Y_i|X_i}(\tau) = \beta_0 + \beta_1 x_i + F_\epsilon^{-1}(\tau)(1 + \gamma x_i) = \beta_0^\tau + \beta_1^\tau x_i,$$

where $F_\epsilon^{-1}(\cdot)$ *is the quantile distribution of the error term. In the quantile model we call the coefficients* $\beta_0^\tau = \beta_0 + F_\epsilon^{-1}(\tau)$ *and* $\beta_1^\tau = \beta_1^\tau + F_\epsilon^{-1}(\tau)\gamma$*. In the case in which* $\gamma = 0$*, the only coefficient that varies with quantile level* τ *is the intercept, while* $\beta_1^\tau = \beta_1$*. In this case, the model has a location shift effect. On the other hand, when* $\gamma > 0$ *both regression coefficients depend on the quantile level, and hence we refer to such model as having location-scale shift effect.*

In our work, we consider γ a measure of heteroskedasticity in the model. Notice that we imposed $(1 + \gamma x_i) > 0$. Such restriction makes sure that map $\tau \mapsto Q_{Y_i|X_i}(\tau)$ is non-decreasing. If we were working in a framework free of the restriction on the sign of $(1 + \gamma x_i)$, in order to insure the monotonicity of $Q_{Y_i|X_i}(\tau)$ we would have to consider a piece-wise linear instead of linear shape of the quantile model, namely

$$Q_{Y_i|X_i}(\tau) = \begin{cases} \beta_0 + \beta_1 x_i + F_\epsilon^{-1}(\tau)(1 + \gamma x_i) & \text{if } (1 + \gamma x_i) > 0 \\ \beta_0 + \beta_1 x_i + F_\epsilon^{-1}(1 - \tau)(1 + \gamma x_i) & \text{otherwise,} \end{cases}$$

since the quantile of $-\epsilon_i$ in τ is $-F^{-1}(1 - \tau)$. From this example we see that, even when we consider a rather simple model, not taking care of the monotonicity of quantiles may produce erroneous results due to *quantile crossing*. Quantile crossing is an issue one has to be aware of, especially when analysing quantiles at multiple levels. Several methods are available in the literature to tackle such problem, from simultaneous quantile estimation with non-crossing constraints (Bondell et al., 2010; Liu and Wu, 2011) to monotonization techniques to be applied either directly on the estimated quantile function (Chernozhukov et al., 2010) or on the estimated CDF that is then inverted (Dette and Volgushev, 2008).

1.1.3 Quantile regression for clustered observations

We now turn to a more complex data structure, namely a clustered one. In particular, for each cluster $i = 1, \dots, N$ consider observations $(Y_{ij}, x_{ij})_{j=1}^{n_i}$, with $Y_{ij} \in \mathbb{R}$ and $x_{ij} \in \mathbb{R}^{p-1}$. We assume within-cluster dependence, but independence between clusters. For each cluster $i = 1, \dots, N$ the linear quantile model at level τ is

$$Q^i_{Y_{ij}|X_{ij}}(\tau) = X_{ij}^T \beta_i^\tau,$$

with $X_{ij}^T = (1, x_{ij}^T)$ and $\beta_i^\tau = (\beta_i^{\tau,1}, \dots, \beta_i^{\tau,p}) \in \mathbb{R}^p$. Adopting the classical formulation of linear mixed effects (Laird and Ware, 1982), one can assume that only some covariates carry cluster-specific effects, while others only act on a population level. Hence, without loss of generality, assume that the first $q \leq p$ components of X_{ij}, which we call Z_{ij}, vary with clusters, while the remaining $p - q$ solely have a mean effect. With an abuse of notation, denote the latter as X_{ij}, in order to be faithful to the classical linear mixed models notation. Then, the quantile model at level τ for cluster i is

$$Q^i_{Y_{ij}|X_{ij}}(\tau) = X_{ij}^T \beta^\tau + Z_{ij}^T u_i^\tau, \tag{1.1.6}$$

where we split coefficient β_i^τ into β^τ, effect at the population level, and u_i^τ, effect at the ith cluster level. As in the literature of linear mixed effects models, we assume $(u_1^\tau, \dots, u_N^\tau)$ to be random elements whose mean is zero, such that on average their effect on the population level vanishes. The loss function to be minimized in this case is

$$\sum_{i=1}^{N} \sum_{i=j}^{n_i} \rho_\tau(Y_i - X_{ij}^T \beta^\tau - Z_{ij}^T u_i^\tau). \tag{1.1.7}$$

Before presenting possible approaches for the estimation of regression parameters β^τ, it is important to be aware that there are different interpretations of the quantile model at the population level. We show this with the following example.

Example 1.2 (Conditional and marginal models)**.** *For each cluster $i = 1, \dots, N$, consider data $(Y_{ij}, x_{ij})_{j=1}^{n_i}$, with $Y_{ij}, x_{ij} \in \mathbb{R}$, and the generating model scale-shift model*

$$Y_{ij} = u_i + \beta_0 + \beta_1 x_{ij} + (1 + \gamma x_{ij})\epsilon_{ij}, \qquad i = 1, \dots, N,\ j = 1, \dots, n_i,$$

with $\gamma > 0$, $(1 + \gamma x_{ij}) > 0$, $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$ and $u_i \overset{iid}{\sim} N(0, \sigma_u^2)$ independent of the error terms. The design matrix is $X_{ij} = (1, x_{ij})$. Two quantile models could be adopted, namely one where we condition the response with respect to both X_{ij} and u_i or solely with respect to X_{ij}. We refer to the first model as conditional while to the latter as marginal. The τth conditional quantile model corresponding to the generating data in the example is

$$Q_{Y_{ij}|X_{ij},u_i}(\tau) = u_i + \beta_0 + \beta_1 x_{ij} + \Phi^{-1}(\tau)(1 + \gamma x_{ij})\sigma_\epsilon,$$

where the computations were carried out similarly to the case of the independent observations. In order to obtain the marginal quantile $Q_{Y_{ij}|X_{ij}}(\tau)$ we first write

$$P(Y_{ij} \leq Q_{Y_{ij}|X_{ij}}(\tau)) = \tau.$$

By substituting the generating model formula and normalizing by the standard deviation of the linear combination $u_i + (1 + \gamma x_i)\epsilon_{ij}$ we obtain

$$P\left(\frac{u_i + (1 + \gamma x_i)\epsilon_{ij}}{\sqrt{\sigma_u^2 + (1 + \gamma x_i)^2 \sigma_\epsilon^2}} \leq \frac{Q_{Y_{ij}|X_{ij}}(\tau) - \beta_0 - \beta_1 x_{ij}}{\sqrt{\sigma_u^2 + (1 + \gamma x_i)^2 \sigma_\epsilon^2}} \right) = \tau.$$

Hence, the τth marginal quantile corresponding to the generating model is

$$Q_{Y_{ij}|X_{ij}}(\tau) = \beta_0 + \beta_1 x_{ij} + \Phi^{-1}(\tau)\sqrt{\sigma_u^2 + (1 + \gamma x_{ij})^2 \sigma_\epsilon^2},$$

where $\Phi^{-1}(\cdot)$ is the quantile function of the standard Gaussian distribution. One can see that $Q_{Y_{ij}|X_{ij},u_i}(\tau)$ and $Q_{Y_{ij}|X_{ij}}(\tau)$ have in general different shapes, the latter not even

being linear. However, notice that the two models coincide when $\sigma_u^2 = 0$*, namely the case in which data is generated without subject-specific effects, while the population-level quantile regression coefficients are the same in the two models, namely* β_0 *and* β_1*, when* $\tau = 0.5$*.*

As shown in the previous example, conditional and marginal models are in general different, especially for extreme quantile levels. When aiming at estimating (1.1.6), namely the quantile for each cluster i, it is important to account for the dependence within clusters by including cluster-specific effects (Koenker, 2004; Reich et al., 2009). In case one was interested in estimating the quantile on the population level, then the marginal model can be employed, provided that the within-cluster dependence is accounted for in the covariance structure (see for instance, Bossoli and Bottai (2017), and Marino and Farcomeni (2015) for an overview on the topic). We focus on conditional models, and in our work presented in Chapter 2 we demonstrated that aiming at estimating quantile regression coefficients of a conditional model with a marginal approach can lead to severe bias.

When it comes to the estimation of the population level quantile regression coefficient β^τ in (1.1.6), several methods are available in the literature. For instance, one possible way of dealing with the problem is by treating the cluster-specific effects as fixed effects, like Kato et al. (2012) and Galvao and Kato (2016). In a somewhat similar way, Canay (2011) presented an estimation procedure of parameters β^τ in two steps: first estimating the cluster-specific effects, considered as fixed, from a mean regression model, and then using them as offsets in a standard quantile regression for independent observations, having dealt with the structural dependence in the previous step. Another possible approach relies on the shrinkage of loss function (1.1.7), especially in those cases with an increasing number of clusters N. Among these works we count Koenker (2004), Lamarche (2010), Harding and Lamarche (2017) and Gu and Volgushev (2019). Finally, a broad class of estimation methods rely on an Asymmetric Laplace Distribution (ALD) working model. The ALD distribution (Yu and Zhang, 2005) is particularly suitable in such framework, given the equivalence between minimizing (1.1.5) at fixed level $\tau \in (0, 1)$ and maximizing the likelihood of an ALD distribution with location and skewness parameters equal to (1.1.4) and τ respectively. One possibility is to consider a linear quantile working model (LQMM) assuming a fully-specified working model where responses Y_{ij} are, conditionally to both X_{ij} and u_i, ALD-distributed, and the distribution of the cluster-specific effects is specified. In such context, Geraci and Bottai (2007) and Geraci and Bottai (2014) considered maximizing the marginalised joint distribution of responses and cluster-specific effects in order to estimate β^τ. On the other hand, Galarza et al. (2017) suggested an EM algorithm for the same model, exploiting an equivalence result of the stochastic representation of the ALD distribution. In a different perspective, Fasiolo et al. (2020) relied on a Bayesian approach aimed at minimizing the penalized Extended Lof-F (ELF) loss function, which consists in a smooth generalization of (1.1.7). In particular, the ELF loss is strongly linked to a class of distributions which the ALD is nested in.

We reviewed as well as tested a selection of estimators of quantile regression coefficients in a conditional model framework in our manuscript in Chapter 2. Moreover, the approach from Fasiolo et al. (2020) plays an important role in our work in Chapter 3.

1.2 Functional data analysis

Functional data analysis deals with stochastic processes that arise from smooth curves. Unlike the subject of the previous section, functional data analysis addresses the interpretation of data, rather than being a specific mathematical tool. One can hence imagine that the topic is rather broad, with several statistical techniques defined to deal with such type of observations. In the following paragraphs we overview those tools of functional data analysis that are particularly relevant in Chapters 3 and 4, taking Ramsay and Silverman (2005) as main theoretical reference.

1.2.1 Characteristics of functional data

Assume we observe $W_1, \ldots, W_H \in \mathbb{R}$. The decision to model those as values taken by an underlying function observed at H points rather than a sequence of scalar observations is based on the interpretation given to the data. In particular, one might want to treat $W_1, \ldots, W_H$ as part of one functional observation if W_h and W_{h+1}, with $h = 1, \ldots, H-1$, are believed not to vary too much from one another, as well as if every single W_h is assumed to be linked to the point it is observed at. In such case the single observation W_h is interpreted as

$$W_h = X(s_h) + \varepsilon_h, \tag{1.2.1}$$

where $X(\cdot)$ is a smooth $L^2(S)$ function, the underlying true functional observation. The domain of $X(\cdot)$ is S, and in practice it is discretized into a grid of points $(s_1, \ldots, s_H)$. Without loss of generality, we assume $S \subset \mathbb{R}$. Moreover, $\varepsilon_1, \ldots, \varepsilon_H$ are independent and identically distributed, drawn from a distribution with mean equal to zero. Hence, we interpret $W_1, \ldots, W_H$ as observations of the values of the true smooth underlying function $X(\cdot)$ over a discrete grid with some measurement noise. These characteristics differentiate functional data from multivariate observations, which are not assumed to be smooth and can be shuffled without the risk of loosing information from their ordering.

Given the underlying smooth nature of functional data, one is usually interested in recovering a smooth estimate $\hat{X}(\cdot)$ of $X(\cdot)$ from the noisy observations. When $\varepsilon_1 = \cdots = \varepsilon_H = 0$, namely when no measurement error occurs, mere interpolation of the observed values would be sufficient. However, that is not feasible in the presence of measurement noise. As a matter of fact, interpolating the values of the noisy observations would bring high point-wise variation to $\hat{X}(\cdot)$ as a result of overfitting. In such case, one has to adopt some *smoothing* technique, and several options in the literature are available, from kernel to local polynomials smoothing (Ramsay and Silverman, 2005, Chapters 3, 4 and 5). A popular way of dealing with the task is by setting known basis functions $\varphi_1, \ldots, \varphi_K$ and represent $\hat{X}(\cdot)$ with a linear combination of them, namely

$$\hat{X}(s) = \sum_{k=1}^{K} c_k \varphi_k(s). \tag{1.2.2}$$

The quality of the approximation depends on several factors, such as the shape and the number K of the basis functions, as well as coefficients $c_1, \ldots, c_K$. Regarding the first matter, a common choice of $\varphi_1, \ldots, \varphi_K$ are B-splines, since they are very flexible and hence can well approximate a wide variety of functional characteristics. Moreover, one would usually choose K high enough to be able to represent the characteristics of $X(\cdot)$, while avoiding nuisance sources of variation. In practice the estimates $\hat{c}_1, \ldots, \hat{c}_K$ for the coefficients in (1.2.2) are obtained by fitting a penalized least squares criterion.

Penalization is a way of imposing smoothness to the estimated curves and a common penalty term is

$$\int_S \left(\frac{d^2}{ds^2} \hat{X}(s) \right)^2 ds.$$

The integrand in the above formula, namely the square of the second derivative of $\hat{X}(\cdot)$, is often called curvature, given the fact that it would be equal to zero for a straight line.

Notice that the aforementioned smoothing method can be applied to single curves. In the case one has access to a collection of functional observations, another possible choice of $\varphi_1, \ldots, \varphi_K$ are the eigenfunctions arising from the eigendecomposition of the estimated covariance function of the available curves, and we are going to overview it in the next paragraph. Both of the two mentioned choices of basis functions are used in Chapter 3 and the former is used in the preprocessing of Chapter 4.

1.2.2 Univariate and Multivariate Functional Principal Component Analysis

Functional principal component analysis (FPCA) consists in an extension of principal component analysis from a multivariate to a functional setting. More specifically, consider the case in which we had access to the true curves $X_1(\cdot), \ldots, X_N(\cdot) \in L^2(S)$, and without loss of generality assume $\mathbf{E}[X_i(s)] = 0$ for $i = 1, \ldots, N$. Then, the sample covariance function can be then defined as

$$v(s, \tilde{s}) = \frac{1}{N-1} \sum_{i=1}^{N} X_i(s) X_i(\tilde{s}). \tag{1.2.3}$$

The pivotal step of FPCA consists in carrying out the spectral decomposition on covariance operator

$$(\mathcal{V}\psi)(\cdot) = \int_S v(\cdot, \tilde{s}) \psi(\tilde{s}) d\tilde{s}.$$

From the decomposition, one may extract eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$, orthonormal eigenfunctions $\psi_1, \psi_2, \ldots$ as well as scores defined as

$$\xi_{ik} = \int_S X_i(s) \psi_k(s) ds, \qquad i = 1, \ldots, N, \ k = 1, 2, \ldots.$$

Notice that, as long as $X_1(\cdot), \ldots, X_N(\cdot)$ are linearly independent, than the covariance operator has rank $N-1$, and thus the decomposition brings $\lambda_1, \ldots . \lambda_{N-1} > 0$. Moreover, any of the functions in the sample can be expressed by means of Karhunen-Loéve representation

$$X_i(s) = \sum_{k=1}^{\infty} \xi_{ik} \psi_k(s). \tag{1.2.4}$$

The idea behind (1.2.4) is the fact that the eigenfunctions represent the main sources of variation of the sample, and hence all together they should fully capture the nature of each one of the collected curves. It is common to truncate the Karhunen-Loéve representation so as to perform dimension reduction on the curves. The number of eigenfunctions to use is usually established by means of Percentage of Variance Explained (PVE), namely

$$K^{PVE} = \min \left\{ K \in \{1, \ldots, P\} : \frac{\sum_{k=1}^{K} \lambda_k}{\sum_{k=1}^{P} \lambda_k} \geq p \right\}, \tag{1.2.5}$$

where P is the maximum number of K allowed. Then, every functional observation can be approximated by

$$\hat{X}_i(s) = \sum_{k=1}^{K^{PVE}} \xi_{ik}\psi_k(s). \tag{1.2.6}$$

As mentioned in the previous section, representation (1.2.6) is also particularly relevant when in practice we observe discrete values of the underlying smooth functions. Specifically, the simplest case one can encounter in applications is when every underlying curve is observed on the same regular and dense grid of points $(s_1, \ldots, s_H)$, and we have access to observations

$$W_{ih} = X_i(s_h).$$

One can hence build $N \times H$ matrix $\mathbf{X}$, such that $[\mathbf{X}]_{ih} = W_{ih}$, as well as covariance matrix $V = \frac{1}{N-1}\mathbf{X}\mathbf{X}^T$. It is then straightforward to compute the spectral decomposition on V, given the singular value decomposition UDW^T of $\mathbf{X}$ and recognizing that $(N-1)V = WD^2W^T$. This corresponds to a multivariate principal component analysis, from which one obtains estimated eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_P \geq 0$, eigenvectors which can be interpolated to obtain eigenfunctions $\hat{\psi}_1, \ldots, \hat{\psi}_P$, and consequently scores $\{\hat{\xi}_{ik}\}$. In order to recover an estimate of the smooth underlying functions one can use (1.2.6), first establishing K^{PVE} using the estimated eigenvalues, and then employing estimated eigenfunctions and scores in the representation. In more complex cases in which grid $(s_1, \ldots, s_H)$ is possibly irregular and

$$W_{ih} = X_i(s_h) + \varepsilon_{ih}, \tag{1.2.7}$$

where measurement errors $\{\varepsilon_{ih}\}$ are iid samples from a zero-mean density, extra computational actions to carry out the eigendecomposition are needed. For instance, penalized splines-based smoothers can be applied to the discretized covariance matrix arising from noisy observations, either in the presence of a dense grid (Xiao et al., 2016) or a sparse one (Xiao et al., 2018). In particular, we use the former approach, named FACE, in Chapter 3, and we illustrate the truncated Karhunen-Loéve representation based on such approach in Example 1.3. Another possibility is to embed the estimation problem in a mixed model framework, where the scores are used as random effects and the off-diagonal elements of the estimated covariance matrix are smoothed (Yao et al. (2003), Yao and Lee (2006) and Goldsmith et al. (2013)). Moreover, smoothing techniques to be directly applied to the observed curved have been suggested, for instance by Huang et al. (2008) and Ramsay and Silverman (2005, Chapter 8).

Example 1.3 (Smoothing via truncated Karhunen-Loéve representation)**.** *To show the result of smoothing via FPCA we show an example in Figure* 1.1*, where the underlying smooth curves are simulated and available in package* `fdasrvf` *(Tucker, 2020), and were also used in Srivastava and Klassen (2016). Consider true curves* $X_1(\cdot), \ldots, X_N(\cdot)$*, with* $N = 21$ *and* $S = [-3, 3]$*, shown in the top left panel. The observed values are as in* (1.2.7)*, where* $\varepsilon_i \overset{iid}{\sim} N(0, 0.02^2)$ *and dense regular grid* $(s_1, \ldots, s_H)$ *with* $s_{h+1} - s_h = 0.06$*, for* $h = 1, \ldots, H-1$*, shown in the top right panel. The bottom panels show the reconstructed functions* $\hat{X}_1(\cdot), \ldots, \hat{X}_N(\cdot)$ *via FACE algorithm (Xiao et al., 2016) when the chosen PVE is 0.95 and 0.9999, in the left and right panels respectively. In the two truncated Karhunen-Loéve representations the selected number of eigenfunctions were* $K = 4$ *and* $K = 11$ *respectively. It is possible to notice how even when employing a low number of eigenfunctions the true curves are estimated quite faithfully, most probably due to*

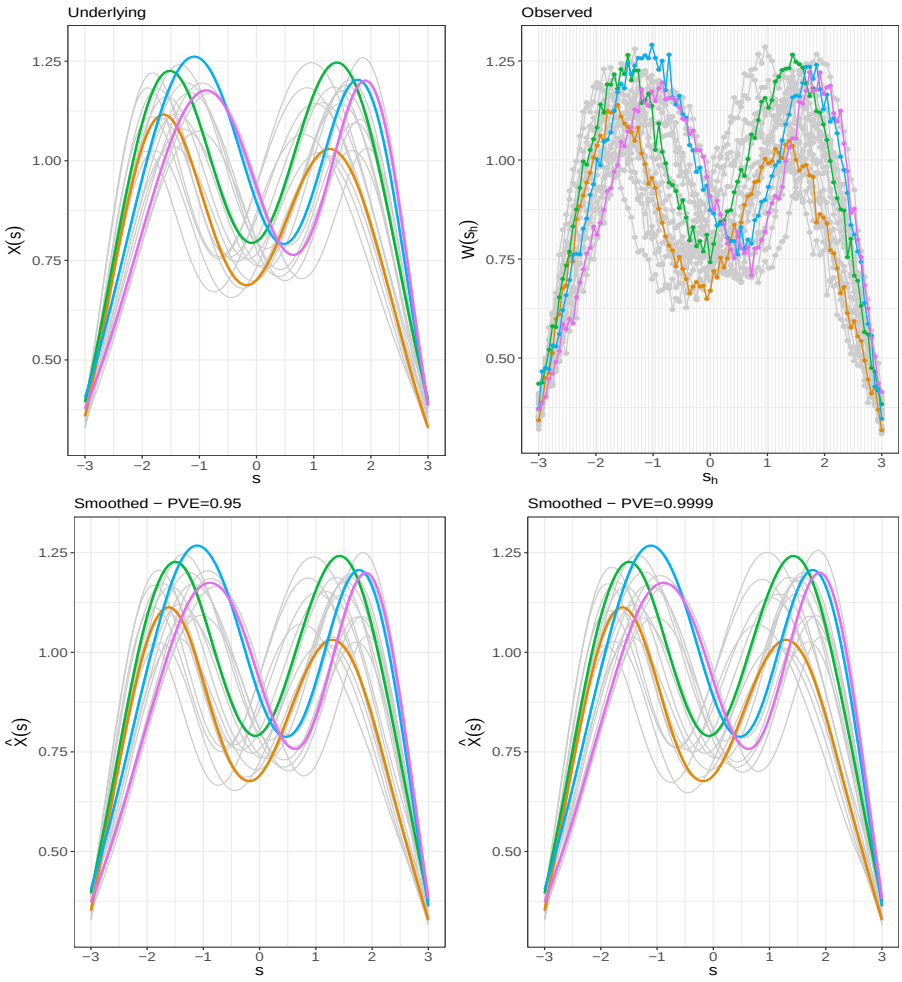

Figure 1.1: Example of smoothing via eigenfunctions.

the simple structure of $X_1(\cdot), \ldots, X_N(\cdot)$*. However, when imposing larger PVE the true underlying functions are reconstructed completely.*

In some situations, one might be interested in studying the sources of simultaneous variation of two or more sets of functions. Consider for instance the case in which we have access to functions $X_1^A(\cdot), \ldots, X_N^A(\cdot) \in L^2(S)$ and $X_1^B(\cdot), \ldots, X_N^B(\cdot) \in L^2(S)$, and without loss of generality assume that $\mathbf{E}[X_i^A] = \mathbf{E}[X_i^B] = 0$ for $i = 1, \ldots, N$. For both the two sets we can define the covariance functions $v_{AA}(\cdot, \cdot)$ and $v_{BB}(\cdot, \cdot)$ as in (1.2.3), as well as the cross-covariance functions $v_{AB}(\cdot, \cdot)$ and $v_{BA}(\cdot, \cdot)$, such that $v_{AB}(s, \tilde{s}) = v_{BA}(\tilde{s}, s)$. Then, we can consider bivariate objects $X_1(\cdot), \ldots, X_N(\cdot) \in L^2(S) \times L^2(S)$, where

$$X_i(s) = (X_i^A(s), X_i^B(s)).$$

One can endow space $L^2(S) \times L^2(S)$ with inner product $\langle\langle \cdot, \cdot \rangle\rangle = \langle \cdot, \cdot \rangle_2 + \langle \cdot, \cdot \rangle_2$, namely the sum of the two L^2 inner products, such that for any $X_i(\cdot), X_j(\cdot) \in L^2(S) \times L^2(S)$

their inner product corresponds to

$$\langle\langle X_i, X_j\rangle\rangle = \int_S X_i^A(s)X_j^A(s)ds + \int_S X_i^B(s)X_j^B(s)ds.$$

Once such metric is defined, the spectral decomposition in this case is carried out on the system of operators

$$(\mathcal{V}_A\psi)(\cdot) = \int_S v_{AA}(\cdot,\tilde{s})\psi^A(\tilde{s})d\tilde{s} + \int_S v_{AB}(\cdot,\tilde{s})\psi^B(\tilde{s})d\tilde{s},$$
$$(\mathcal{V}_B\psi)(\cdot) = \int_S v_{BA}(\cdot,\tilde{s})\psi^A(\tilde{s})d\tilde{s} + \int_S v_{BB}(\cdot,\tilde{s})\psi^B(\tilde{s})d\tilde{s}.$$

The results are, as in the univariate case, eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ and eigenfunctions $\psi_1, \psi_2, \ldots \in L^2(S) \times L^2(S)$, with $\psi_k(s) = (\psi_k^A(s), \psi_k^B(s))$. Moreover, scores are defined as

$$\xi_{ik} = \langle\langle X_i, \psi_k\rangle\rangle = \int_S X_i^A(s)\psi_i^A(s)ds + \int_S X_i^B(s)\psi_i^B(s)ds,$$

and a bivariate version of Karhunen-Loéve representation (1.2.4) holds.

In practice, in case the two sets of curves are observed on a dense grid of points $(s_1, \ldots, s_H)$, namely

$$W_{ih}^A = X_i^A(s_h),$$
$$W_{ih}^B = X_i^B(s_h),$$

then one can carry out estimation as in the univariate case, considering the linked vectors $(W_{i1}^A, \ldots, W_{iH}^A, W_{i1}^B, \ldots, W_{iH}^B)$ as observations. Once the principal component analysis is computed, it is possible to separate components A and components B of the eigenfunctions so as to return to a bivariate setting. Methods based on such concatenation were presented by Ramsay and Silverman (2005, Chapter 8), Berrendero et al. (2011), Jacques and Preda (2014) and Chiou et al. (2014). More recently, Happ and Greven (2018) proposed a flexible estimation method based in the eigendecomposition of the covariance matrix of the scores arising from univariate FPCA of the single components of the multivariate functional objects. Moreover, their approach can be extended to more general frameworks where the different components have different domains. We used this approach in the work presented in Chapter 4, in combination with phase-amplitude separation, which is overviewed in the next section. Finally, the methods used in univariate FPCA for observations with errors, on either dense or sparse grid, can be used in this multivariate setting.

1.2.3 Registration of functional data

Registration is a tool that allows to extract and separate two characteristics of a collection of curves, namely their *amplitude* and their *phase* variations. Generally speaking, these two components can be thought as the vertical and horizontal "shifts" of the curves respectively, usually with respect to a reference function. In the literature of functional data registration is often regarded as a preprocessing technique which allows to compare curves' features, such as peaks, more directly, discarding the possible phase variation in the occurrence of such features. To give a concrete example, if functional coordinate s varied with time, then it would be possible to regard the horizontal shift as "delay".

Consider curves $X_1(\cdot), \ldots, X_N(\cdot) \in \mathcal{F}(S) \subset L^2(S)$, with $S = [S_0, S_1] \subset \mathbb{R}$ and where $\mathcal{F}(S)$ is the space of absolutely continuous functions on S. Even though there are several methods to carry out registration, the results of such analysis would generally bring warpings $\gamma_1(\cdot), \ldots, \gamma_N(\cdot) \in \Gamma_S$. These functions correspond to the deformations to be applied to $X_1(\cdot), \ldots, X_N(\cdot)$ in order for them to be *aligned*, namely having characteristics as peaks and valleys occurring at similar points. For instance, Srivastava and Klassen (2016) considered a flexible class of warping functions, namely boundary-preserving diffeomorphisms, defined as

$$\Gamma_S = \{\gamma : S \to S \mid \gamma(S_0) = S_0, \gamma(S_1) = S_1, \gamma \text{ diffeomorphism}\}.$$

As mentioned earlier, the objective of group alignment is to find those warpings such that curves

$$\tilde{X}_i = X_i \circ \gamma_i \qquad i = 1, \ldots, N$$

are aligned with each other. We consider $\gamma_1(\cdot), \ldots, \gamma_N(\cdot)$ to be phase representatives, while $\tilde{X}_1(\cdot), \ldots, \tilde{X}_N(\cdot)$ are usually taken as amplitude representatives. These two sets of curves are the results of the *phase-amplitude separation*. With their work, Srivastava and Klassen (2016) gave two main contributions. The first one was setting up a framework in which amplitude is regarded as an equivalence class, in light of the desirable property that amplitude should not be changed by warpings. Their second contribution was establishing a notion of distance for the space of equivalence classes, or orbits. This is vital for the estimation of phase and amplitude components, since it is based on the optimization of a criterion measuring the optimal distance between a deformed function and the template it is aligned to. Specifically, they showed that the Fisher-Rao (FR) metric is appropriate since it has several desirable properties such as the invariance under warpings, namely the registration of two or more functions should remain the same if they are all warped with the same $\gamma \in \Gamma_S$. Moreover, they proved that the FR metric corresponds to the L^2 metric when functions $X_1(\cdot), \ldots, X_N(\cdot)$ are transformed into the square root velocity functions (SRVFs). The template used for group alignment is the center of the orbit corresponding to the Karcher mean of the equivalence classes generated by the SVRFs. These are the foundations for their proposed algorithm for registration, which we employ in our work in Chapter 4 as well as in Example 1.4.

Several other approaches to achieve phase-amplitude separation have been proposed in the literature. For instance, Ramsay and Silverman (2005, Chapter 7) presented landmark registration, where curves are aligned for some given features. The warping functions are the result of the interpolation of the points corresponding to the set of aligned landmarks. Moreover, s likelihood-based approach was proposed by Wrobel et al. (2019). Finally, approaches that rely on the functional characteristics of data are also available. Kneip and Ramsay (2008) suggested a methodology that is FPCA-based, while Tang and Müller (2008) adopted a two-step procedure in which they first perform pairwise-alignment among all the curves and then use the results to build the global registration. Moreover, Gervini and Gasser (2004) proposed a functional regression framework to estimate phase components, and Sangalli et al. (2010) suggested an algorithm to detect clusters of amplitude and phase representatives based on functional features. For an overview of phase-amplitude separation theory and existing methods, see Marron et al. (2015).

Example 1.4 (Phase-amplitude separation)**.** *Assume we have access to the collection of curves $X_1(\cdot), \ldots, X_N(\cdot)$ represented in the upper left panel of Figure* 1.1*. Figure* 1.2 *shows the registered curves $\tilde{X}_1(\cdot), \ldots, \tilde{X}_N(\cdot)$ and the corresponding warping functions*

$\gamma_1(\cdot), \ldots, \gamma_N(\cdot)$. *The phase-amplitude separation was carried out with the method from Srivastava and Klassen (2016). In the first plot one can see how features like local maxima*

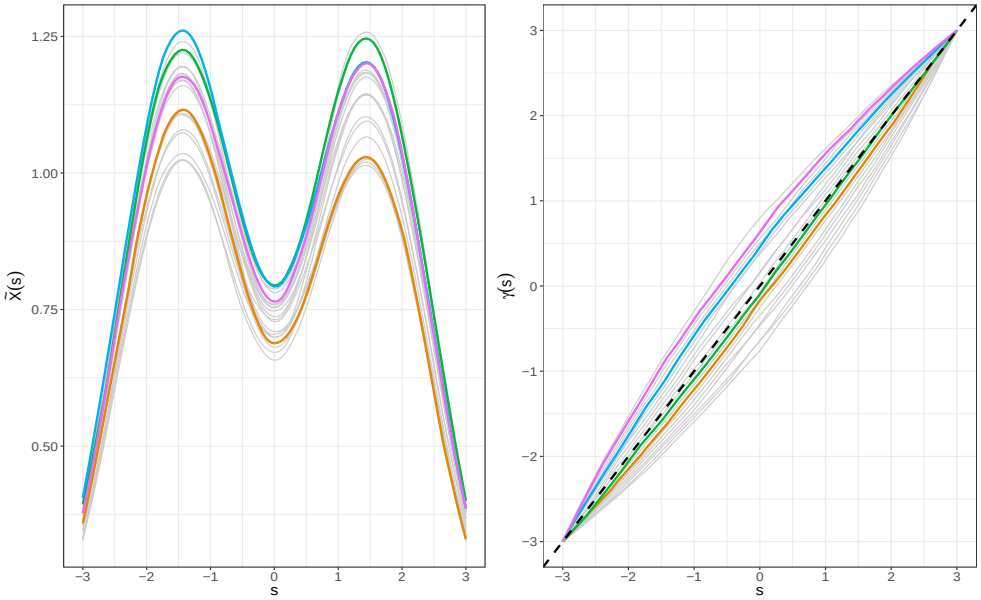

Figure 1.2: Aligned curves (on the left) and warping functions (on the right) from the registration of the curves shown in Example 1.3, with the same colour convention.

and minimum take place at the same functional coordinates for all the aligned curves. It is possible to observe, for instance, that the blue and green curves take the same value in the local minimum in $s = 0$*, which was not so clear looking at Figure* 1.2*. In the second plot the warping functions are shown. The dashed line represents* γ_{id}*, the identity element of* Γ_S*. The blue line is above* γ_{id}*, which indicates that the corresponding original function is delayed with respect to the template. On the other hand, the green curve is slightly below* γ_{id}*, which means that the corresponding original curve is a bit anticipated compared to the template.*

1.2.4 Regression with functional covariates and scalar response

Consider independent data $(Y_i, \hat{X}_i(\cdot))_{i=1}^N$, where $Y_i \in \mathbb{R}$ and $\hat{X}_i(\cdot)$ are smooth estimates of the underlying functions $X_i \in L^2(S)$, with $i = 1, \ldots, N$. For instance, in mean regression a common model is

$$\mathbf{E}[Y_i|X_i] = \int_S \beta(s)X_i(s)ds, \tag{1.2.8}$$

in which we assume that the response for the ith observations is explained by the integral of the product between the ith functional covariate and the functional coefficient $\beta(\cdot)$. Given the infinite dimension of coefficient $\beta(\cdot)$, any regression framework in which we aimed at studying some characteristic of the conditional distribution of the response given the observed functional covariates would have infinite degrees of freedom. In practice, apart from employing smooth estimates $\hat{X}_1(\cdot), \ldots, \hat{X}_N(\cdot)$ as functional covariates, one can consider representing the functional coefficient by means of linear combination of

known basis functions. More specifically, taking $\beta(s) \approx \sum_{d=1}^{D} b_d \phi_d(s)$, with known basis functions $\phi_1, \dots, \phi_D$, (1.2.8) can be approximated by

$$\sum_{d=1}^{D} b_d \int_S \phi_d(s) \hat{X}_i(s) ds = \sum_{d=1}^{D} b_d Z_{id}, \tag{1.2.9}$$

where Z_{id} correspond to the integral of the product between covariate $\hat{X}_i(\cdot)$ and basis function $\phi_d(\cdot)$. With such representation one shifts the estimation problem from an infinite dimensional functional setting to a finite dimensional one with standard covariates. The choice of the basis functions $\phi_1, \dots, \phi_D$ is arbitrary. One possibility is to adopt B-splines, and eventually penalize coefficients $b_1, \dots, b_D$ for smoothing purposes (Cardot et al., 2003; Goldsmith et al., 2011a). Another option is to rely on the first D eigenfunctions arising from the FPCA carried out on the functional observations, and in such case, we would use the score as standard covariates since $Z_{id} = \xi_{id}$, for every observation $i = 1, \dots, N$ and every basis function $d = 1, \dots, D$ (Cardot et al., 1999a).
For this overview we used mean regression model (1.2.8). However, the different approaches to the representation of $\beta(\cdot)$ hold for a broad class of regression models. Given the central role of quantile regression Chapter 3, in the following example we show how estimation of the functional coefficient can be carried out in such context.

Example 1.5 (Quantile regression with functional covariates)**.** *Consider data* $(Y_i, \hat{X}_i(\cdot))_{i=1}^N$ *and the true generating model with the underlying smooth functions* $X_1(\cdot), \dots, X_N(\cdot)$

$$Y_i = \int_S \beta(s) X_i(s) ds + \left(\gamma \int_S X_i(s) ds \right) \epsilon_i, \qquad i = 1, \dots, N$$

with $\gamma \geq 0$, $\gamma \int_S X_i(s) ds > 0$ *and* $\epsilon_i \overset{iid}{\sim} f_\epsilon$ *with mean 0. For fixed level* $\tau \in (0, 1)$*, the corresponding quantile model is*

$$Q_{Y_i|X_i}(\tau) = \int_S \left(\beta(s) + \gamma F_\epsilon^{-1}(\tau) \right) X_i(s) ds = \int_S \beta^\tau(s) X_i(s) ds.$$

Using the smooth approximations of the functional covariates and representing $\beta^\tau(\cdot)$ *with basis functions* $\phi_1, \dots, \phi_d$*, the quantile model can be approximated by*

$$Q_{Y_i|X_i}(\tau) \approx \sum_{d=1}^{D} b_d^\tau Z_{id},$$

where we used the same notation as in (1.2.9)*. For instance, if we choose to represent the* $\beta^\tau(\cdot)$ *with B-splines with penalty on the coefficients of the expansion, then estimates* $\hat{\mathbf{b}}^\tau = (\hat{b}_1^\tau, \dots, \hat{b}_D^\tau)$ *are computed as*

$$\hat{\mathbf{b}}^\tau = \arg\min_{\mathbf{b}^\tau} \left[\rho_\tau \left(Y_i - \sum_{d=1}^{D} b_d^\tau Z_{id} \right) + \gamma \left\| \mathbf{b}^\tau \right\|_B^2 \right],$$

with $\left\| \mathbf{b}^\tau \right\|_B^2 = (\mathbf{b}^\tau)^T B \mathbf{b}^\tau$*, where* B *is the penalty matrix of choice, and penalty parameter* $\gamma > 0$*.*

Finally, when dealing with smoothed functional covariates one should be aware of the identifiability issues that come with it. In particular, when employing approximation (1.2.2) for every single estimated covariate curve, then $\int_S (\beta(s) + \tilde{\beta}(s)) \hat{X}_i(s) ds =$

$\int_S \beta(s)\hat{X}_i(s)ds$, for every $\tilde{\beta} \in span(\{\varphi_1, \ldots, \varphi_K\})^{\perp}$, so that $\beta(\cdot)$ cannot be distinguished from $\beta(\cdot) + \tilde{\beta}(\cdot)$. In light of the fact that the functional coefficient is identifiable only up to elements belonging to the orthogonal complement of $span(\{\varphi_1, \ldots, \varphi_K\})$, extra caution should be given to its interpretation in a regression framework.

1.3 Contribution of this thesis

This section is dedicated to an overview of the motivations and results shown in the following chapters. More specifically, manuscript A focuses on estimation and inference in quantile regression when considering clustered-structured data. In manuscript B, we study the same framework while also considering functional covariates, and our motivating application arises from data with a longitudinal design. Finally, we present a data analysis relying on functional data techniques, such as registration and MFPCA, in manuscript C.

1.3.1 Manuscript A

In Chapter 2 we present our work regarding quantile regression applied to conditional linear quantile models for clustered scalar data. The focus of the manuscript lies in the estimation of the population-level coefficients in those cases in which the number of clusters is much larger than the number of observations per clusters. To give an example, the reference scenario in our simulation study is characterized by $N = 500$ clusters and $n = n_i = 6$ observations per cluster, and as data application we present our analysis on an ACTG study with $N = 1187$ patients whose CD4 counts was recorded $n_i \in \{2, \dots, 9\}$ times over the study.

Our first objective is to demonstrate that, in the above described framework, a selection of estimation methods in the literature fail at providing unbiased estimators of the population-level quantile regression coefficients. In particular, some of these approaches consider cluster-specific effects as fixed, and this is known to lead to the "incidental parameters problem" in mean regression (Neyman and Scott, 1948; Lancaster, 2000). However, we demonstrated that the bias related to the coefficients' estimation also occurred for methods that consider cluster-specific effects as random elements. In light of this, our second aim is to introduce a novel estimator with better properties. In particular, we propose a two-step estimator that relies on the best linear unbiased predictions (BLUPs) of subject-specific effects, treated as random, arising from the LQMM framework (Geraci and Bottai, 2007, 2014), then used as offsets in a quantile regression setting for independent observations. Even though the bias of the coefficients' estimates is improved with our estimator, our third and final contribution is to provide a bias adjustment for it. We test several bootstrap schemes, and we find that a combination of resampling of subject-specific effects combined with wild bootstrap (Wu, 1986; Liu, 1988) for the residuals is the most successful to build bias adjusted estimates and inference for the population-level coefficients. Our findings are based on extensive simulation studies, where we devoted special attention to location-scale shift heteroskedastic models ($\gamma > 0$) when estimation is carried out at a single quantile level as low as $\tau = 0.1$. Such setting is not commonly studied in the literature of quantile regression, where results are usually presented for homoscedastic models and/or the estimation is carried out at the median. Finally, we compare estimation and inference results of our suggested method and those from the LQMM setting in our application.

1.3.2 Manuscript B

The work presented in Chapter 3 is motivated by a longitudinal study from animal science. We are interested in studying the physical conditions of sows living in a commercial farm right after giving birth and along their lactation period (Park et al., 2019; Staicu et al., 2020). The surroundings of the animals, such as the temperature in the stables, can

impact their daily quantity of food intake, and scarce nutrition can severely affect the sows as well as their litter. Since those sows that eat the least are those most at risk, it comes naturally to consider a low level quantile regression model for the food intake conditional on the external temperature, adding a sow-specific shift that accounts for the longitudinal structure of the measurements as well as a smooth effect of lactation days. Moreover, since temperature was recorded frequently and regularly during each day, it is possible to regard it as a functional effect. The model we aim at is rather complex, as it requires techniques from both quantile regression and functional data analysis applied to longitudinal designs. To our knowledge there are no other works in the literature that adopt a similar framework.

First of all, we outline the aforementioned modelling setting. In particular, we consider two different approximation approaches of the functional effects, namely via splines and eigenfunctions. For the latter we propose a method for the selection of the number of basis K inspired by Kato (2012) and based on BIC. We give precise computational directions on how to implement the model of interest relying on ready available software. In particular, we adopt the flexible estimation framework proposed by Fasiolo et al. (2020), available in `R` package `qgam`, which relies on the methods developed by Wood (2017) for generalized additive models. We then test such estimation framework for different scenarios in our simulation section. Finally, in our application section we study the behavior of young and old sows for "low" and "high" temperature profiles, namely the 20% and 80% point-wise quantiles of the overall functional temperature recordings respectively. We show that later in the lactation period the temperature effect is indeed significant, and we apply bootstrap adjustments for estimation and inference inspired by Battagliola et al. (2021). Moreover, with model selection based on AIC we conclude that including both a functional and longitudinal effect is the most suitable setting for the application.

1.3.3 Manuscript C

Chapter 4 presents an analysis embedded in the functional data framework concerning neuroscience. More specifically, we base our work on that of Benoit et al. (2020), whose interest was studying working memory impairment in patients affected by psychiatric disorders. In our manuscript we analyze the learning curves of mice performing memory-involving tasks repeatedly. In particular, we compare the performances of a group of animals with an induced brain lesion aimed at emulating a similar damage of psychiatric disorders patients, and a control group.

For our study we adopt the framework outlined by Happ et al. (2019), who proposed carrying out MFPCA on bivariate functional objects whose univariate elements are the amplitude and phase components obtained by phase-amplitude separation of a collection of functions. With such analysis we are able to compare the different sources of variation, namely amplitude, phase and a combination of the two, of the two groups of mice across several stages of the overall experiment while accounting for the simultaneous variation of phase and amplitude components. In our application of interest both amplitude and phase variations carry interesting information, namely whether mice can reach a high probability of completing the tasks with success and, if so, how fast they reach such results compared to the chosen template curve. From the results we obtain, it is possible to conclude that there are indeed differences in the way the two groups of mice behave, as well as some behaviors that are common in both groups.

Chapter 2

A bias-adjusted estimator in quantile regression for clustered data

Maria Laura Battagliola, Helle Sørensen, Anders Tolver & Ana-Maria Staicu

Corrected proof available in Ecosta (https://doi.org/10.1016/j.ecosta.2021.07.003)

Abstract

The manuscript discusses how to incorporate random effects for quantile regression models for clustered data with focus on settings with many but small clusters. The paper has three contributions: (i) documenting that existing methods may lead to severely biased estimators for fixed effects parameters; (ii) proposing a new two-step estimation methodology where predictions of the random effects are first computed by a pseudo likelihood approach (the LQMM method) and then used as offsets in standard quantile regression; (iii) proposing a novel bootstrap sampling procedure in order to reduce bias of the two-step estimator and compute confidence intervals. The proposed estimation and associated inference is assessed numerically through rigorous simulation studies and applied to an AIDS Clinical Trial Group (ACTG) study.

Keywords: Linear quantile regression; Clustered data; Random effects; Bias-adjustment; Wild bootstrap; ACTG study

2.1 Introduction

Quantile regression has been introduced by Koenker and Bassett Jr (1978) as a way to describe the association between covariates and quantiles of the response distribution at pre-set quantile levels. See the comprehensive monographs by Koenker (2005a) and Koenker et al. (2017) on quantile regression. In recent years, quantile regression has for example been employed in econometrics and finance (Bayer, 2018; Wang et al., 2018b; Maciak, 2021a,b).In this article we consider linear quantile regression for clustered data, such as longitudinal data, and discuss estimation approaches that properly account for the inherent dependence of the observations within the same cluster. Research in this area has been very active, especially in econometrics, but existing methods for quantile regression estimation are proved to be asymptotically consistent only when both the number of clusters and cluster size increase to infinity. This assumption is rather strong in practice, where the common scenario is that there are many clusters of moderate to small sizes. When the cluster size is small, numerical investigations show (see Figure 2.2) that the popular quantile regression estimators may exhibit severe bias, even if there are many clusters. This represents a gap in the literature, as data settings that involve many clusters of small to moderate sizes are ubiquituos in medicine and animal science, to name a few.

Existing approaches to account for dependence in parameter estimation of quantile regression for clustered (repeated measures) data treat the cluster-specific parameters either as fixed or random. For example, Kato et al. (2012) and Galvao and Kato (2016) use cluster-specific intercepts and estimate them as fixed effects parameters together with the quantile regression parameters using the so-called fixed effects quantile regression (FE-QR) and fixed effects smoothed quantile regression (FE-SQR), respectively, while Galvao and Wang (2015) and Galvao et al. (2017) develop minimum-distance-based estimation for the same purpose. Some approaches consider shrinkage to deal with an increasing number of clusters, in the presence of cluster-specific parameters. Penalized quantile regression for longitudinal data is discussed by Koenker (2004), Lamarche (2010), Harding and Lamarche (2017) and Gu and Volgushev (2019). Canay (2011) proposes a two-step estimator, relying on mean regression estimates of cluster-specific intercepts, see also Besstremyannaya and Golovan (2019). Geraci and Bottai (2007) and Geraci and Bottai (2014) introduce a pseudo likelihood approach, where a linear quantile mixed model (LQMM) with random cluster parameters is used as a working model, and Galarza et al. (2017) develop an EM-based estimation methodology for the LQMM framework. Abrevaya and Dahl (2008) discuss estimation in a model with correlated random effects (CRE), and Luo et al. (2012) consider a fully Bayesian quantile inference using Markov Chain Monte Carlo, to account for correlated random effects. We consider a frequentist perspective and propose a novel two-step estimation approach and associated inference that rely on the LQMM framework.

When the cluster-specific parameters are treated and estimated as fixed effects parameters, estimation suffers from what is known in the literature as the "incidental parameters problem" (Neyman and Scott, 1948; Lancaster, 2000): the number of (nuisance) parameters grows with the number of clusters, leading to inconsistent joint estimation, when the cluster size is small. Not surprisingly, only asymptotic scenarios where both the number of clusters and the cluster size increase to infinity have been studied (Koenker, 2004; Kato et al., 2012; Galvao and Kato, 2016; Canay, 2011; Besstremyannaya and Golovan, 2019). To bypass the issues caused by the incidental parameter problem, the cluster-specific parameters can be modeled as random effects; however, asymptotic properties are not studied for the LQMM-based estimator (Geraci and Bottai, 2007, 2014).

Different solutions have been suggested for bias-adjustment in the case of small clusters: Galvao and Kato (2016) introduce an analytical adjustment for FE-SQR based on asymptotic analysis, nonetheless the approach requires an optimal bandwidth selection, which is challenging in practice. The authors also adapt the half-panel jackknife method (Dhaene and Jochmans, 2015) to longitudinal quantile regression. We consider the use of half-panel jackknife for bias correction in our numerical investigation. Usually, bootstrap methods have been used for construction of confidence intervals in models with cluster-specific effects (Galvao and Montes-Rojas, 2015; Canay, 2011; Geraci and Bottai, 2014), and for marginal models (without cluster-specific effects), see for example Karlsson (2009) and Hagemann (2017). We introduce a non-standard bootstrap technique for both bias-adjustment and inference of quantile regression parameters, in the context of clustered (longitudinal) data.

This paper makes three main contributions. First, we numerically demonstrate that Koenker's penalized estimator, Canay's two-step estimator and the LQMM estimator can be severely biased when clusters are small or of moderate size. Although no papers have claimed the opposite, we are the first to raise this issue. Second, we propose a new estimation methodology and associated inference for the quantile regression parameters. The point estimator is computed in two steps: (i) an LQMM framework is used to predict the cluster-specific parameters; and (ii) the predictions are used as offsets in a standard quantile regression. The two-step estimator is furthermore adjusted for bias using bootstrap, and the third contribution is the novel combination of wild bootstrap and ordinary resampling, that reduces bias and allows to construct confidence intervals that have good coverage performance. Numerical studies show that the proposed estimator has considerably smaller bias than the existing competitors, when the cluster size is small.

The structure of the paper is as follows: we set up the model framework in Section 2.2. In Section 2.3 we summarize some of the existing estimation methods in quantile regression for repeated measures data and then present the proposed estimation method. The estimation method is evaluated numerically in a thorough simulation study in Section 2.4 (with additional results in the appendix) and applied to a clinical trial regarding HIV treatments in Section 2.5. The paper concludes with Section 2.6, which discusses the main findings.

2.2 Regression framework

Let $(Y_{ij}, x_{ij})_{j=1}^{n_i}$ be the observed data for the ith cluster ($i = 1, \ldots, N$), where $x_{ij} \in \mathbb{R}^{p-1}$ is the vector of covariates corresponding to the jth observation of the ith cluster and $Y_{ij} \in \mathbb{R}$ is the respective response. Here n_i denotes the cluster size and the responses are assumed independent across different clusters but expected to be correlated within the same cluster. Let $\tau \in (0, 1)$ be a fixed quantile level of interest, and let $Q^i_{Y_{ij}|x_{ij}}(\tau)$ be the τth quantile of the conditional distribution of Y_{ij} given x_{ij} for cluster i. Consider a linear quantile regression model

$$Q^i_{Y_{ij}|X_{ij}}(\tau) = X_{ij}^T \beta_i^\tau, \tag{2.2.1}$$

where $X_{ij}^T = (1, x_{ij}^T)$ and $\beta_i^\tau = (\beta_i^{\tau,1}, \ldots, \beta_i^{\tau,p})$ is an unknown vector regression parameter that quantifies the association between the covariates and the τ-quantile of the response for cluster i. Due to the definition of X_{ij}^T, the first component of β_i^τ is the intercept; by an abuse of notation we refer to X_{ij} as the vector of covariates.

This model formulation allows for cluster-level effects for every scalar component of X_{ij}; an equivalent formulation is to represent the cluster-level effect as the sum of a population level effect and a cluster-specific deviation. Such formulation is standard in the mixed effects model representation (Laird and Ware, 1982), and we adopt it here as well. As for mean regression, all covariates are not necessarily modeled with cluster-specific levels, and the selection of variables without cluster-specific effects can be based on interpretational as well as computational arguments. Without loss of generality, assume that only the first $q \leq p$ components of X_{ij} have cluster-varying effects; denote by Z_{ij} the vector formed by the first q elements of X_{ij}. The remaining $p-q$ components of X_{ij} have only population level effect. The effects corresponding to Z_{ij} are used to account for the dependence of the observations within the same cluster; for example, Koenker (2004), Canay (2011), and Galvao and Kato (2017) used a random intercept only ($q = 1$) to model this dependence. Using the terminology from linear mixed effects we can re-write model (2.2.1) as

$$Q^i_{Y_{ij}|X_{ij}}(\tau) = X_{ij}^T\beta^\tau + Z_{ij}^T u_i^\tau, \tag{2.2.2}$$

by separating the quantile regression parameters that describe a population level effect, $\beta^\tau = (\beta^{\tau,1}, \ldots, \beta^{\tau,p-q})$, from the ones that describe cluster-specific deviations, $u_i^\tau = (u_i^{\tau,1}, \ldots, u_i^{\tau,q})$. Just like in linear mixed models, it is assumed that u_i^τ are zero mean random quantities. Our primary interest lies in the estimation of β^τ in situations with many clusters (large N) but modest cluster sizes (small n_is).

Let $\mathbf{u}^\tau = (u_1^\tau, \ldots, u_N^\tau)$ denote the collection of (unobserved) cluster-specific parameters. Moreover, let $\mathbf{Y}$ be the vector of the (observed) responses Y_{ij}. Consider the loss function

$$L(\beta^\tau, \mathbf{u}^\tau; \mathbf{Y}) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \rho_\tau(Y_{ij} - X_{ij}^T\beta^\tau - Z_{ij}^T u_i^\tau), \tag{2.2.3}$$

where $\rho_\tau(v) = v(\tau - \mathbf{1}_{(v<0)})$ is the check function (Koenker and Bassett Jr, 1978). If the values of the cluster-specific effects, u_i^τ, were observed, a natural estimator would be the linear quantile regression estimator corresponding to the covariates X_{ij} and the modified responses $Y_{ij} - Z_{ij}^T u_i^\tau$. We call this the oracle estimator,

$$\hat{\beta}^\tau_{\text{oracle}} = \arg\min_{\beta^\tau} L(\beta^\tau, \mathbf{u}^\tau; \mathbf{Y}); \tag{2.2.4}$$

evidently the estimator $\hat{\beta}^\tau_{\text{oracle}}$ enjoys the asymptotic properties of a standard quantile regression estimator (Koenker, 2005a). However, $\hat{\beta}^\tau_{\text{oracle}}$ is an unattainable estimator, as u_i^τs are not observed, and the question we consider in this paper concerns the effect of uncertainty in the cluster-specific effects on estimating the population level quantile regression parameter.

One way to address the estimation problem is to treat u_i^τs in (2.2.2) as fixed effects parameters and have them estimated jointly with β^τ using a standard quantile regression framework. The FE-QR estimation of Kato et al. (2012) minimizes the loss function (2.2.3) with respect to both β^τ and $\mathbf{u}^\tau$. With this approach, the number of parameters grows at the same rate as the number of clusters, so the estimator of β^τ is only consistent in asymptotic scenarios where n_i grows faster than N (Kato et al., 2012).

We will instead pursue an approach to estimate β^τ, when u_i^τs are treated as random. Similar to the generalized linear mixed effects framework, there are two interpretations of the covariates' effects on the response distribution quantile. On one hand, we have the conditional perspective, following from the definition (2.2.2) that $P(Y_{ij} \leq X_{ij}^T\beta^\tau +$

$Z_{ij}^T u_i^\tau | X_{ij}, u_i^\tau) = \tau$, which states that β^τ is the quantile regression parameter associated with the covariates X_{ij}, conditional on the cluster-specific effects. On the other hand, we have the marginal perspective that $P(Y_{ij} \leq X_{ij}^T \tilde{\beta}^\tau | X_{ij}) = \tau$, which describes the covariates' effect on the τ-quantile of the marginal distribution of Y_{ij}. The two quantile regression parameters (β^τ and $\tilde{\beta}^\tau$) are generally different in the same manner that a fixed effects parameter of a generalized linear mixed model has a different interpretation than its counterpart in a marginal or population average approach (Zeger et al., 1988; Neuhaus et al., 1991). The difference between the conditional and marginal quantile models is discussed more thoroughly in Reich et al. (2009), see also the simulation model in Section 2.4.

As a consequence, also pointed out in Koenker (2004), it is vital for the estimation of β^τ of a conditional perspective that the cluster-specific parameters u_i^τ are not ignored. Indeed, we illustrate in Section 2.4.2 that the simple marginal quantile regression estimator $\hat{\beta}^\tau_{\text{marg}} = \arg\min_{\beta^\tau} L(\beta^\tau, \mathbf{0}; \mathbf{Y}) = \arg\min_{\beta^\tau} \sum_{i,j} \rho_\tau(Y_{ij} - X_{ij}^T \beta^\tau)$ based on standard quantile regression (where all u_is are replaced by zero) may be severely biased for β^τ.

The conditional perspective implies that

$$P(Y_{ij} - Z_{ij}^T u_i^\tau \leq X_{ij}^T \beta^\tau | X_{ij}) = \tau,$$

where the probability is taken with respect to the joint distribution of Y_{ij} and u_i. Inspired by this equality, we propose to first predict the cluster-specific effects and then use these predictions as offset in a standard linear quantile regression model using a transformed response.

2.3 Estimation

2.3.1 Review of selected methods for estimation and bias-adjustment

Penalization of cluster-specific parameters

The model (2.2.2) was first introduced in the literature by Koenker (2004) in a simpler form, where the term $Z_{ij}^T u_i^\tau$ is replaced by only a cluster-specific intercept, call it u_{i0}, which is assumed to be quantile-invariant. For fixed quantile level τ, both the parameter β^τ and the cluster-specific intercepts, u_{i0}, are estimated by minimizing the penalized loss function

$$L(\beta^\tau, \mathbf{u}_0; \mathbf{Y}) + \lambda \sum_{i=1}^{N} |u_{i0}|, \tag{2.3.1}$$

where $\lambda \geq 0$ is a regularization parameter. Koenker (2004) uses ℓ_1 penalty in (2.3.1) due to its computational convenience; in our numerical investigation of the estimators in Section 2.4, we also use ℓ_2 penalty and find minor differences. While (2.3.1) focuses on a single quantile level, Koenker (2004) describes the estimation of the quantile regression parameters simultaneously at multiple quantile levels, by introducing quantile-level weights and minimizing a weighted penalized likelihood.

The ℓ_1-penalized estimator for β^τ is consistent and asymptotically normal, provided that $N^a/n \to 0$ for some $a > 0$ (where $n_i = n$); see Koenker (2004). Nonetheless, when the cluster size, n_i, is small the estimator may not enjoy these theoretical properties and can be seriously biased, especially for extreme quantile levels; see Section 2.4.

Canay's two-step estimator

Canay (2011) assumes a cluster-specific intercept, u_{i0}, too, but considers a two-step procedure to estimate the linear quantile regression parameter β^τ of (2.2.2). First, u_{i0} are estimated as part of the fixed parameters in a mean regression framework. Second, the quantile regression parameter β^τ is estimated using a standard quantile regression framework (Koenker and Bassett Jr, 1978) applied to adjusted responses $\widetilde{Y}_{ij} = Y_{ij} - \hat{u}_{i0}$, where $\hat{u}_{i0}$ denotes the estimated cluster-specific effects from the previous step. Equivalently, β^τ is estimated by minimizing the loss function (2.2.3) with $Z_{0i} = 1$ and $\mathbf{u}^\tau$ replaced by $\hat{\mathbf{u}}_0$, the vector containing the $\hat{u}_{i0}$s:

$$\hat{\beta}^\tau_{\text{Canay}} = \arg\min_{\beta^\tau} L(\beta^\tau, \hat{\mathbf{u}}_0; \mathbf{Y}).$$

Canay (2011) and Besstremyannaya and Golovan (2019) discuss asymptotic properties for $\hat{\beta}^\tau_{\text{Canay}}$ in scenarios where both the number of clusters and cluster size increase.

The use of the mean regression in the first step is justified in Canay's set-up because only intercepts are allowed to be cluster-specific, and the deviations from the average are assumed to be constant over quantile levels. In such case, the random effects correspond to location shifts; their estimation is quantile-invariant, which may be restrictive. Moreover, while treating u_{i0}s as fixed parameters as opposed to random may lead to negligible differences, in terms of estimation, for large clusters, the correct approach for small clusters is to treat them as random parameters. To address this issue, we propose a new quantile regression estimator in Section 2.3.2, which is inspired by Canay (2011).

Marginalization over random effects in a working model (LQMM)

Geraci and Bottai (2007, 2014) propose to embed the problem in a fully specified working model, a linear quantile mixed model (LQMM), using the duality between the quantile loss (check function) and the asymmetric Laplace distribution (ALD, Yu and Zhang (2005)). Specifically, assume $u_i \sim f(\cdot;\varphi)$ for some density f that is parameterized by a scale parameter φ and posit the following joint model for the responses Y_{ij}s and the cluster-specific u_is:

$$\begin{gathered} Y_{ij}|u_i, X_{ij} \overset{ind}{\sim} ALD(X_{ij}^T\beta^\tau + Z_{ij}^T u_i, \sigma, \tau), \qquad j = 1, \ldots, n_i \\ u_i \overset{iid}{\sim} f(\cdot, \varphi), \end{gathered} \tag{2.3.2}$$

for $i = 1, \ldots, N$, where σ is a scale parameter for the residual distribution. The conditional τ-quantile function associated to the working model is given by (2.2.2), and the conditional likelihood of Y_{ij}s given X_{ij}s and u_is takes the form (2.2.3); with $u_i^\tau = u_i$.

Estimation of model parameters $(\beta^\tau, \sigma, \varphi)$ is based on maximizing the pseudo likelihood of $\mathbf{Y}$ obtained by integrating the joint density of $(Y_{i1}, \ldots, Y_{in_i}, u_i)$ with respect to the distribution of latent random effects u_i. In practice, the random effects are assumed to be drawn either from a Gaussian distribution $N(0, \varphi^2)$ or a Laplace distribution $ALD(0, \varphi, 1/2)$, see Geraci (2014) for details about the computations. In the special case of random intercepts only, when the Laplace distribution is used for the cluster-specific parameters u_{i0}s, maximizing the joint model (2.3.2) is equivalent to minimizing Koenker's penalized loss function, while if the Gaussian distribution is used, then maximizing the joint model (2.3.2) is equivalent to minimizing the ℓ_2-penalized criterion. From this perspective, the tuning parameters using Koenker's penalization approach are scale parameters in the joint model framework and thus can be estimated with increased

computational efficiency. Finally, once the parameters β^τ, σ and φ are estimated, the random effects can be predicted using best linear predictors (BLPs), see equation (12) in Geraci and Bottai (2014). These predictions are essential ingredients for the new estimator suggested in Section 2.3.2; note that the computed predictions vary with the level τ even though u_i in the model (2.3.2) does not.

Geraci and Bottai (2007) and Geraci and Bottai (2014) do not discuss asymptotics for the LQMM estimator, but if the working assumptions are true (ALD for the within-cluster distribution and Gaussian or Laplace distribution for the random effects), then the LQMM estimator is the maximum likelihood estimator, and the usual asymptotic results hold. On the other hand, the bias of the LQMM estimator may be non-negligible, even when N is large, if the data generating process does not coincide with the working model. This will be illustrated in Section 2.4.2.

Jackknife-based bias-adjustment for an existing estimator

Since the estimators above show bias when used for clustered data, a bias reduction adjustment would be appropriate. There are various ways to do this; one approach to reduce the bias of an estimator is by using a jackknife bias-adjustment. The half-panel jackknife was first introduced in Dhaene and Jochmans (2015) as a method for bias correction for mean regression in longitudinal settings with many subjects and fixed panel size. Later, it was applied to the FE-SQR estimator for longitudinal quantile regression (Galvao and Kato, 2016); we describe it here for clustered data.

We randomly split the dataset into two sub-datasets, each containing half of the observations from every cluster. Denote the quantile regression estimator from the two sub-datatsets by $\hat{\beta}_1^\tau$ and $\hat{\beta}_2^\tau$, respectively, and let $\hat{\beta}^\tau$ be the estimator from the full dataset. Then, the half-panel jackknife estimator $\hat{\beta}^\tau_{\text{jackknife}}$ is defined as

$$\hat{\beta}^\tau_{\text{jackknife}} = \hat{\beta}^\tau - \left(\frac{1}{2}(\hat{\beta}_1^\tau + \hat{\beta}_2^\tau) - \hat{\beta}^\tau\right) = 2\hat{\beta}^\tau - \frac{(\hat{\beta}_1^\tau + \hat{\beta}_2^\tau)}{2}. \tag{2.3.3}$$

To gain some intuition about the bias reduction of this estimator, assume that all clusters have equal size n and that the asymptotic bias of the initial estimator $\hat{\beta}^\tau$ is of the form $C/n + o(n^{-1})$ for some constant C. Then the asymptotic bias of the jackknife estimator $\hat{\beta}^\tau_{\text{jackknife}}$ is of order $o(n^{-1})$, so the order of the bias is reduced. Nonetheless, empirical studies indicate that while the adjustment indeed reduces the bias, the resulting variance of the estimator is increased; see Galvao and Kato (2016).

2.3.2 Proposed quantile estimation with reduced bias

A new two-step estimator (unadjusted)

We propose to estimate the linear quantile regression parameter β^τ using a new approach, which is inspired by the LQMM estimation framework and Canay (2011). It consists of two steps:

Step 1: Use the LQMM framework to predict the cluster-specific random effects by the best linear predictors (BLPs) and center them; denote the centered prediction for cluster i by $\tilde{u}_i^\tau$;

Step 2: Transform the responses to $\widetilde{Y}_{ij} = Y_{ij} - Z_{ij}^T\tilde{u}_i^\tau$ and use the standard quantile regression framework for the new responses $\widetilde{Y}_{ij}$ and covariates X_{ij} to estimate β^τ.

There are two key differences between the proposed approach and Canay (2011): 1) Canay estimates the cluster-specific effects using a mean regression framework, whereas we use a quantile regression model, and 2) Canay estimates the cluster-specific effects by treating them as fixed parameters; in contrast we view and estimate them as random parameters. We illustrate in Section 2.4 that these differences have a large impact in terms of the estimation quality of quantile regression parameters.

Figure 2.1 shows a comparison between true random effects (x-axis) and their predicted values (y-axis) for the first cluster from 200 simulated data sets representing the benchmark scenario in Section 2.4. The BLPs capture the variation among clusters quite well, but it is clear that some degree of shrinkage takes place as more extreme random effects are drawn towards zero.

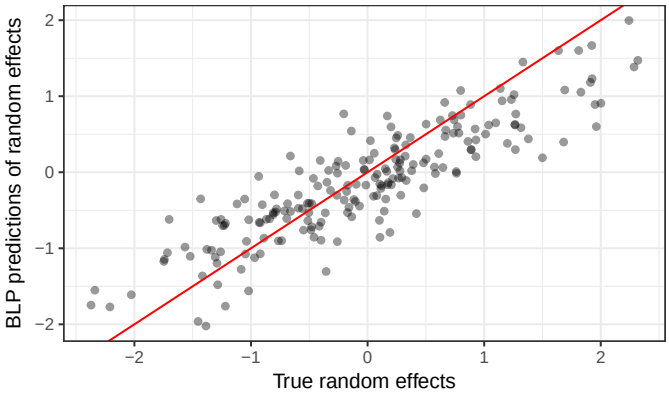

Figure 2.1: Comparison of the true random effects (on x-axis) and centered BLP predictions (on y-axis) for the first cluster from 200 simulated datasets from the standard scenario in Section 2.4. The red line is the line with slope one through the origin.

The second step consists of standard quantile regression applied to $Y_{ij} - Z_{ij}^T \tilde{u}_i^\tau$; equivalently the quantile regression parameter is estimated by minimizing the loss function (2.2.3), with $\mathbf{u}$ fixed at value $\tilde{\mathbf{u}}^\tau$, the vector containing $\tilde{u}_i^\tau$s:

$$\hat{\beta}^\tau_{\text{two-step}} = \arg\min_{\beta^\tau} L(\beta^\tau, \tilde{\mathbf{u}}^\tau; \mathbf{Y}).$$

Our two-step estimator turns out to have considerably smaller bias than the LQMM estimator; yet, the deviation between the true and estimated random effects introduces some bias. To bypass this issue, we propose a bias-corrected adjustment based on bootstrap as explained below. The second step can be carried out with standard software, which typically provides standard errors for each component of the vector β^τ. However, it is important to recognize that these uncertainty estimates are not necessarily reliable, as they only account for the sampling variability of $\hat{\beta}^\tau_{\text{two-step}}$ conditional on the random effects, not for the extra variation due to the uncertainty in predicting the random effects. We propose to use bootstrap to estimate the total variation of $\hat{\beta}^\tau_{\text{two-step}}$. We describe the bootstrap procedures used for bias-adjustment and estimation of variability in the following.

Bootstrap sampling for bias-adjustment

We propose a semi-parametric-type of bootstrap, which combines non-parametric bootstrap and wild bootstrap and relies on the linearity of the quantile regression model.

Let $\mathcal{U} = \{\tilde{u}_1^\tau, \ldots, \tilde{u}_N^\tau\}$ be the sample of predicted cluster-specific effects obtained with two-step estimation procedure and for each i and j denote the observed residuals by $\varepsilon_{ij} = Y_{ij} - X_{ij}^T \hat{\beta}^\tau_{\text{two-step}} - Z_{ij}^T \tilde{u}_i^\tau$.

We define the bootstrap sample as $\{(Y_{ij}^*, X_{ij}, Z_{ij})_{j=1}^{n_i}, u_i^{\tau,*}\}_{i=1}^N$ where $u_i^{\tau,*}$s are obtained by resampling with replacement from $\mathcal{U}$ and Y_{ij}^* is defined by

$$Y_{ij}^* = X_{ij}^T \hat{\beta}^\tau_{\text{two-step}} + Z_{ij}^T u_i^{\tau,*} + \varepsilon_{ij}^*, \qquad i = 1, \ldots, N, \;\; j = 1, \ldots, n_i, \tag{2.3.4}$$

where ε_{ij}^*s are attained by wild bootstrap; see Wu (1986) and Liu (1988) who introduced this method in the context of mean regression. Specifically, let $\varepsilon_{ij}^* = w_{ij}|\varepsilon_{ij}|$, where w_{ij}s are drawn independently from the following distribution:

$$w = \begin{cases} 2(1-\tau), & \text{with probability } 1-\tau \\ -2\tau, & \text{with probability } \tau \end{cases} \tag{2.3.5}$$

which has the τ-quantile equal to 0. The idea of scaling the residuals by weights drawn from an asymmetric distribution was proposed by Feng et al. (2011); as Wang et al. (2018a) also recognized, the wild bootstrap captures asymmetry and homoscedasticity better than ordinary resampling of residuals. Notice that the coupling between covariates and residuals is maintained in the equation (3.3.12) in the sense that each residual is used to generate a bootstrap value for its own observation.

Bootstrap methods have been used for inference on quantile regression for longitudinal data. Most of the approaches rely on non-parametric resampling where complete clusters are sampled with replacement, by sampling the covariates and the outcomes jointly (Canay, 2011; Kato et al., 2012; Galvao and Montes-Rojas, 2015; Geraci and Bottai, 2014; Karlsson, 2009). This method is useful for evaluation of an estimator's variation, and thus for computation of standard errors and confidence intervals. However, we expect such bootstrap estimators to be centered around the estimate from the observed data, and they would therefore not be useful for bias-adjustment. In contrast, our bootstrap procedure ensures that the resampled observations are generated from a distribution with $\hat{\beta}^\tau_{\text{two-step}}$ as the "true" parameter; therefore, we can measure bias as the deviation between $\hat{\beta}^\tau_{\text{two-step}}$ and the bootstrap estimates. Details are given below. Our proposed bootstrap method (abbreviated RW, for standard Resampling and Wild) is compared with resampling of complete clusters and two additional approaches in Section 2.4.

The RW bootstrap sampling procedure ensures that, conditional on the resampled random effects, the model assumption about the association between the covariates and the quantile at level τ is satisfied with $\beta^\tau = \hat{\beta}^\tau_{\text{two-step}}$ (obtained from the observed data). Furthermore, if the random effects were known then all observations were independent, and the distribution of the bootstrap estimators obtained with wild bootstrap would represent the sampling distribution of $\hat{\beta}^\tau_{\text{two-step}}$ (Feng et al., 2011; Wang et al., 2018a). However, due to the potential deviation between the working model in LQMM and the true data generating model, the empirical distribution of LQMM predictors of the random effects may not fully represent the cluster-to-cluster variation, and since this variation is driving the bias, the proposed estimator does not completely remove the bias of the initial estimator asymptotically.

Once a bootstrap sample is available, the quantile regression estimator is obtained by using the proposed two-step estimation approach. At this part, information about the resampled cluster-specific effects are ignored; nonetheless these terms are used in a subsequent step, when we estimate the estimator's variability. The bootstrap estimate of the quantile regression parameter is obtained by averaging the estimates in B such

bootstrap samples. If $\hat{\beta}^{\tau,*}_{\text{two-step},b}$ denotes the bth bootstrap replicate then the overall bootstrap estimate of the quantile regression parameter is $\bar{\beta}^{\tau,*}_{\text{two-step}} = \sum_{b=1}^{B} \hat{\beta}^{\tau,*}_{\text{two-step},b}/B$. The deviation $\bar{\beta}^{\tau,*}_{\text{two-step}} - \hat{\beta}^{\tau}_{\text{two-step}}$ between the overall bootstrap estimate and the original estimate is regarded as an estimate of the bias, so an adjusted estimator (Efron and Tibshirani, 1993, Chapter 10.6) is defined by

$$\hat{\beta}^{\tau}_{\text{adj}} = \hat{\beta}^{\tau}_{\text{two-step}} - \left(\bar{\beta}^{\tau,*}_{\text{two-step}} - \hat{\beta}^{\tau}_{\text{two-step}}\right) = 2\hat{\beta}^{\tau}_{\text{two-step}} - \bar{\beta}^{\tau,*}_{\text{two-step}}. \tag{2.3.6}$$

As illustrated by numerical studies, this quantile regression estimator has reduced bias compared to the (unadjusted) two-step estimator.

Confidence intervals

An important advantage of using a bootstrap-based estimator is that it allows to study the variability of the estimator, and we now discuss construction of the confidence intervals for the quantile regression parameter for each component k of the p-dimensional parameter β^{τ}. We consider two approaches: the first approach is based on the so-called basic bootstrap method to construct confidence intervals and the second approach capitalizes on the availability of the bootstrap sample of the cluster-specific effects, which is obtained at each step of the bootstrap procedure.

The basic bootstrap $100(1-\alpha)\%$ confidence intervals (Davison and Hinkley, 1997, eq. 5.6) for β^{τ}_{k} are defined as

$$\left(2\hat{\beta}^{\tau}_{\text{two-step},k} - \beta^{\tau,*}_{1-\alpha/2,k}\,;\, 2\hat{\beta}^{\tau}_{\text{two-step},k} - \beta^{\tau,*}_{\alpha/2,k}\right), \quad k = 1, \ldots, p,$$

where $\beta^{\tau,*}_{\alpha/2,k}$ and $\beta^{\tau,*}_{1-\alpha/2,k}$ are the $\alpha/2$ and $(1-\alpha/2)$ quantiles, respectively, in the bootstrap sample of $\hat{\beta}^{\tau,*}_{\text{two-step},k}$.

The second approach to construct confidence intervals relies on a normal asymptotic distribution for the quantile regression estimator and the bootstrap-based estimate of the variance of the quantile regression estimator. However, in contrast to most bootstrap-based confidence intervals constructed this way, the bootstrap standard error alone, $\text{SD}_{\text{two-step},k} = \sqrt{\sum_{b=1}^{B}(\hat{\beta}^{\tau,*}_{\text{two-step},k,b} - \bar{\beta}^{\tau,*}_{\text{two-step},k})^2/(B-1)}$, fails to accurately quantify the full variability of the quantile regression estimator of β^{τ}. This is due to the shrinkage phenomenon of the LQMM predicted cluster-specific effects, which is further perpetuated in the bootstrap samples of $u^{\tau,*}_{i}$s and incorporated in the bootstrap replicates $\hat{\beta}^{\tau,*}_{\text{two-step},b}$.

To bypass this issue, we consider an adjustment. In this regard, denote by $\text{SE}_{\text{obs},k}$ the estimated standard error of the kth component of $\hat{\beta}^{\tau}_{\text{two-step}}$ reported by the standard quantile regression (Koenker and Bassett Jr, 1978) with the cluster-specific effects set to the LQMM predicted values and using the accordingly transformed data (step 2 of our procedure). Recall that this quantity ignores the variability of the cluster-specific effects, and thus underestimates the true variability of the regression estimator. Fortunately, our bootstrap algorithm, by resampling from the empirical distribution of the predicted cluster-effects, allows us to track the variability of the regression estimator induced by the uncertainty in predicting these effects. Let $\hat{\beta}^{\tau,*}_{\text{oracle},b}$ denote the oracle-type quantile regression estimator based on the bth bootstrap sample, i.e. the Y^{*b}_{ij}s, and by using the "true" values of the cluster-specific effects, i.e. the $u^{\tau,*b}_{i}$s. As before, for each component k denote by $\bar{\beta}^{\tau,*}_{\text{oracle}} = \sum_{b=1}^{B} \hat{\beta}^{\tau,*}_{\text{oracle},b}/B$ and $\text{SD}_{\text{oracle},k} = \sqrt{\sum_{b=1}^{B}(\hat{\beta}^{\tau,*}_{\text{oracle},k,b} - \bar{\beta}^{\tau,*}_{\text{oracle},k})^2/(B-1)}$

the mean and standard deviation, respectively, of the oracle-type quantile regression estimator.

We define the adjusted standard error of the kth component of the two-step quantile regression estimator as

$$\mathrm{SE}_{\mathrm{adj},k} = \mathrm{SD}_{\text{two-step},k} \, \frac{\mathrm{SE}_{\mathrm{obs},k}}{\mathrm{SD}_{\mathrm{oracle},k}} \quad k = 1, \ldots, p.$$

Since both terms of the ratio are based on keeping the cluster-specific constant, the ratio is used to account for the shrinkage phenomenon. Another way to understand the adjusted standard error is to view it as a multiplicative factor to the standard error that is reported in our step 2, $\mathrm{SE}_{\mathrm{obs},k}$: in this case the ratio $\mathrm{SD}_{\text{two-step},k}/\mathrm{SD}_{\mathrm{oracle},k}$ measures the extra variation of the quantile regression estimator due to estimation of the random cluster-specific effects.

The $100(1-\alpha)\%$ confidence intervals for β_k^τ based on the adjusted standard errors are computed as

$$\hat{\beta}^\tau_{\mathrm{adj},k} \pm q_{1-\alpha/2} \cdot \mathrm{SE}_{\mathrm{adj},k}, \tag{2.3.7}$$

where $q_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of $N(0,1)$. These confidence intervals will later be referred to as SE-adjustment confidence intervals.

We summarize our procedures for estimation and inference in Algorithm 1.

2.3.3 Software

The two-step quantile regression estimator is computed using two different R (R Core Team, 2020a) packages. For the first step, the LQMM estimation method is implemented by the `lqmm()` function from the package `lqmm` (Geraci, 2014; Geraci and Bottai, 2014). For the second step, we use standard quantile regression implemented by the function `rq()` from the `quantreg` package (Koenker, 2020). Bootstrap datasets are generated with standard sampling functions. An R function for the complete estimation and inference process is available from the corresponding author's website.

2.4 Simulations

2.4.1 Data generating model

We consider a data generating model inspired by the simulation designs in Koenker (2004) and Geraci and Bottai (2014). Specifically,

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + (1 + \gamma x_{ij}) e_{ij}, \qquad i = 1, \ldots, N, \;\; j = 1, \ldots, n_i, \tag{2.4.1}$$

where $u_i \overset{iid}{\sim} N(0, \sigma_u^2)$, $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$, x_{ij} are uniformly distributed on $(0,1)$ and $\gamma \geq 0$ is a homoscedasticity-departure parameter. Notice that $1 + \gamma x_{ij}$ is always positive. When $\gamma \neq 0$, the covariate has both a location shift and a scale effect (Koenker, 2004). In the homoscedastic case (i.e. $\gamma = 0$), the correlation between observations from the same cluster is $\frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$. With a slight abuse of notation, we refer to this ratio as the interclass correlation coefficient (ICC) even when $\gamma > 0$.

Model (2.4.1) implies the following quantile regression model

$$Q_{Y_{ij}|x_{ij},u_i}(\tau) = \beta_0^\tau + \beta_1^\tau x_{ij} + u_i, \tag{2.4.2}$$

Consider data $\{(Y_{ij}, X_{ij}, Z_{ij})_{j=1}^{n_i} : i = 1, \ldots, N\}$;
Using LQMM framework, obtain the centered BLPs of the random effects: $\{\tilde{u_i}^{\tau}\}$;
Use data $\{(\widetilde{Y}_{ij}, X_{ij})_{j=1}^{n_i} : i = 1, \ldots, N\}$, where $\widetilde{Y}_{ij} = Y_{ij} - Z_{ij}^T \tilde{u_i}^{\tau}$ and get the (unadjusted) estimate, $\hat{\beta}^{\tau}_{\text{two-step}}$, and its estimated standard error, SE_{obs};
For all i, j compute residuals as $\varepsilon_{ij} = Y_{ij} - X_{ij}^T \hat{\beta}^{\tau}_{\text{two-step}} - Z_{ij}^T \tilde{u}_i^{\tau}$;
forall $b = 1 : B$ **do**
  Draw weights w_{ij} from the weight distribution (2.3.5);
  Use wild bootstrap on ε_{ij}: $\varepsilon_{ij}^{*b} = w_{ij}|\varepsilon_{ij}|$;
  Resample $\tilde{u_i^{\tau}}$ with replacement to get $u_i^{\tau,*b}$;
  Construct the bootstrap sample: $[\{(Y_{ij}^{*b}, X_{ij}, Z_{ij})_{j=1}^{n_i}, u_i^{\tau,*b}\} : i = 1, \ldots, N]$ where $Y_{ij}^{*b} = Z_{ij}^T u_i^{\tau,*b} + X_{ij}^T \hat{\beta}^{\tau}_{\text{two-step}} + \varepsilon_{ij}^{*b}$;
  Use data $\{(\widetilde{Y}_{ij}^{*b}, X_{ij})_{j=1}^{n_i} : i = 1, \ldots, N\}$, where $\widetilde{Y}_{ij}^{*b} = Y_{ij}^{*b} - Z_{ij}^T u_i^{\tau,*b}$ and standard linear quantile regression estimation to get $\hat{\beta}^{\tau,*}_{\text{oracle},b}$;
  Use data $\{(Y_{ij}^{*b}, X_{ij}, Z_{ij})_{j=1}^{n_i} : i = 1, \ldots, N\}$ and the proposed two-step estimation to get $\hat{\beta}^{\tau,*}_{\text{two-step},b}$;
end
Compute the two-step bootstrap mean, $\bar{\beta}^{\tau,*}_{\text{two-step}}$;
For each component $k = 1, \ldots, p$, calculate the standard deviation for the two-step and oracle estimators, $\text{SD}_{\text{two-step},k}$ and $\text{SD}_{\text{oracle},k}$, respectively;
For specified α, for each component $k = 1, \ldots, p$ in part calculate:

- $100(1-\alpha)\%$ basic confidence interval: $\left(2\hat{\beta}^{\tau}_{\text{two-step},k} - \beta^{\tau,*}_{1-\alpha/2,k}\,;\, 2\hat{\beta}^{\tau}_{\text{two-step},k} - \beta^{\tau,*}_{\alpha/2,k}\right)$
- $100(1-\alpha)\%$ SE adjusted confidence interval: $\hat{\beta}^{\tau}_{\text{adj},k} \pm q_{1-\alpha/2} \cdot \text{SE}_{\text{adj},k}$, where $\text{SE}_{\text{adj},k} = \text{SD}_{\text{two-step},k} \frac{\text{SE}_{\text{obs},k}}{\text{SD}_{\text{oracle},k}}$

Algorithm 1: Pseudo code for implementation of the bootstrap adjusted two-step estimator and related confidence intervals.

where $\beta_0^{\tau} = \beta_0 + \sigma_e \Phi^{-1}(\tau)$ and $\beta_1^{\tau} = \beta_1 + \gamma\sigma_e\Phi^{-1}(\tau)$, with Φ denoting the cumulative distribution function for the $N(0,1)$ distribution. In particular, the quantiles are of the same form as (2.2.2), with $X_{ij} = (1, x_{ij})$ and $Z_{ij} = 1$,and with u_i^{τ} not depending on τ. When $\gamma = 0$ the slope parameter of the quantile is constant across τ, i.e., $\beta_1^{\tau} = \beta_1$, while the covariate effect differs between quantile levels when $\gamma \neq 0$. Irrespective of the choice of γ, the regression parameter for the median, $\beta_1^{0.5}$, does not depend on γ, since $\Phi^{-1}(0.5) = 0$.

Notice that the data generating model implies that the marginal-type quantile at level τ of Y_{ij} given x_{ij} (but not conditional on u_i) is given by

$$\beta_0 + \beta_1 x_{ij} + \Phi^{-1}(\tau)\sqrt{\sigma_u^2 + (1+\gamma x_{ij})^2\sigma_e^2}. \tag{2.4.3}$$

In the heteroscedastic setting ($\gamma > 0$) this expression is not linear in x_{ij}, in contrast with (2.4.2), and a linear approximation has parameters that are different from β_0^{τ} and β_1^{τ}. This shows that a marginal estimation approach aims at different parameters compared to those in (2.4.2).

We are going to compare our proposed estimators to the marginal estimator and the other estimation methods discussed in Section 2.3. To implement the approaches we use the function `rq()` of the `quantreg` package (Koenker, 2020) to perform standard quantile regression and the `lqmm()` function of the package `lqmm` (Geraci, 2014) to perform LQMM. More specifically, we use Gauss-Hermite quadrature (option `lqmmType="normal"` in `lqmm`) with 15 quadrature points (`nK=15`) and derivative-free optimisation (`lqmmMethod="df"`). Quantile regression with ℓ_1 and ℓ_2 penalization and cross validation for selection of the penalty parameter is implemented in the function `cv.hqreg()` of the `hqreg` package (Yi, 2017). We use five-fold cross validation. Finally, we use $B = 100$ bootstrap replications for bias-adjustment, where applicable.

2.4.2 Comparison of estimation methods

Overall comparison for a benchmark scenario

In the model (2.4.2), we consider true (mean) parameters $\beta_0 = \beta_1 = 1$, homoscedasticity departure parameter $\gamma = 0.4$, variances $\sigma_u^2 = \sigma_e^2 = 1$, and thus ICC = 0.5. The main focus is on the quantile level $\tau = 0.1$ that is somewhat extreme; then true parameter values amount to $\beta_0^\tau = -0.281$ and $\beta_1^\tau = 0.487$. Define the "benchmark scenario" by the case with $N = 500$ clusters of size $n_i = 6$ ($i = 1, \ldots, N$); we use this scenario to study the performance of the estimators in the situation with $N \gg n_i$.

Figure 2.2 shows the boxplots of the bias for β_0^τ (left) and β_1^τ (right) corresponding to quantile levels $\tau = 0.5$ (top) and $\tau = 0.1$ (bottom), based on 200 Monte Carlo simulations. We compare the proposed two-step estimator and its adjusted version (`twostep` and `adj`, respectively), the estimator from Canay (2011) (`canay`), the LQMM estimator (`lqmm`) and its jackknife-based adjustment (`jackknife`), the estimators arising from penalized quantile regression, both with ℓ_1 and ℓ_2 penalties (`l1pen` and `l2pen`, respectively), the marginal estimator arising from standard quantile regression (`marg`), and the estimator from (2.2.4) where the actual random effects are used in the computations (`oracle`). The oracle estimator is unfeasible in practice, but is used as a reference to study the effect of random effects being latent.

All nine estimators have similar distributions for $\tau = 0.5$, except the jackknife-adjusted estimator, which has slightly larger variation for both parameters. The results are more interesting for $\tau = 0.1$. Focusing first on the methods developed in this paper, the unadjusted two-step estimator has a smaller bias (component-wise) than the other estimators studied; yet, there is still some bias left compared to the oracle estimator. The bias-adjusted estimator, on the other hand, has a very small bias (for each component) and variance that is slightly larger than that of the oracle estimator, but comparable to the other competitors.

The estimator proposed by Canay (2011) has a comparable bias to the other estimators when it comes to the slope, but it shows positive (but small) bias for the intercept. The variance is small for both components of the quantile regression parameters. Results for the LQMM estimators and the estimators from Koenker (2004) based on ℓ_1 penalisation are similar and show a small bias for both components. The estimator based on ℓ_2 penalisation has the same properties for the slope, but has a larger bias for the intercept. The jackknife-based adjustment of the LQMM estimator reduces the bias for the slope parameter, but not for the intercept, and generally, it has large variation.

As expected, the standard quantile regression estimator, which completely ignores the cluster structure, leads to increased bias. The bias is particularly severe for the intercept, whereas the bias for the slope is comparable to that of Canay's estimator, the LQMM

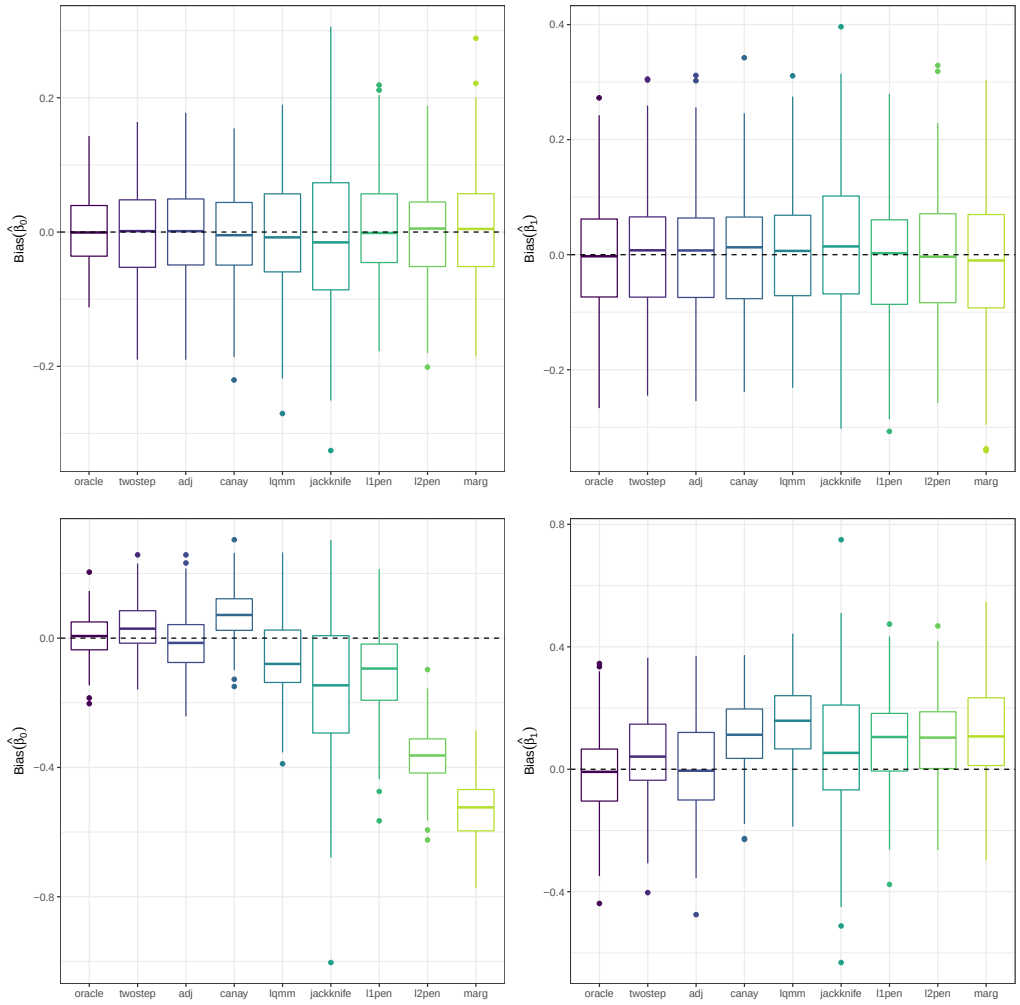

Figure 2.2: Bias for different estimators of β_0^τ (left) and β_1^τ (right) for 200 datasets from the benchmark scenario. The quantile level is 0.5 (top) and 0.1 (bottom). The true parameter values are $\beta_0^{0.5} = \beta_1^{0.5} = 1$ and $\beta_0^{0.1} = -0.281$, $\beta_1^{0.1} = 0.487$, respectively.

estimator, and the penalization-based estimators. This is interesting, as it indicates that these latter estimators effectively estimate the slope coefficient in (a linearized version of) a marginal quantile model rather than in the conditional quantile model.

Additional simulation results are included in the appendix; Tables 2.4–2.7 show results for settings where (N, n_i) differ from the benchmark scenario, and for quantile levels $\tau = 0.1, 0.5$. The conclusions from Figure 2.2 are confirmed; in particular an advantage of the proposed estimators is observed for $\tau = 0.1$ (Tables 2.5 and 2.7). In passing, we note that the ℓ_1-penalized estimator is preferable to the ℓ_2-penalized estimator in all settings, and that the jackknife estimator reduces bias for $\hat{\beta}_1^\tau$ but increases bias for $\hat{\beta}_0^\tau$ and has larger variance. For those reasons we do not study the ℓ_2-penalized and the jackknife estimators any further. The remaining estimators are discussed in more detail in the next section.

The average computing time per simulated dataset for the bootstrap-adjusted two-step estimator was 18.83 seconds. By comparison, the computation time for the LQMM estimator was 0.15 seconds. The difference reflects the additional $B = 100$ iterations involving LQMM estimation and the construction of the confidence intervals that are required by the proposed method. The average computation time for Canay's estimator was 0.58 seconds. The average computation time for the ℓ_1-penalized estimator was 72.29 seconds, partly due to the cross-validation step. The computation time for the ℓ_2-penalized estimator was close to that of the ℓ_1-penalized, and computations for the jackknife adjusted estimator took about three times longer than computations for LQMM. Computations were run on a commodity PC with 2.9 GHz Dual–Core Intel Core i5 processor 5287U.

Bias for LQMM, ℓ_1-penalized, and Canay's estimator for extreme quantile levels

For quantile level 0.1, the bias of the LQMM, ℓ_1-penalized, ℓ_2-penalized and Canay's estimators in the bottom of Figure 2.2 is quite large. This flaw is reported for Canay's estimator in a simulation study with N much larger than n_i and varying quantile levels (Canay, 2011); however, to the best of our knowledge, the bias has not been documented thoroughly in the literature for the other estimators. The ℓ_1-penalized estimation is carried out in Koenker (2004) for a simulation model similar to ours, but only for the median ($\tau = 0.5$) where all estimators are unbiased. LQMM estimation is analyzed in Geraci and Bottai (2014) in many simulation scenarios with good overall performance, but the dependence on bias of sample size (N and n_i) is not studied in the presence of heteroscedasticity.

Figure 2.3 shows boxplots of the bias for the LQMM, the ℓ_1-penalized, and Canay's estimator for various number of clusters, N, cluster sizes, n_i, and at different quantile levels, τ; results are based on 200 replications. We vary one factor at a time, while keeping the others fixed at their benchmark values ($N = 500$, $n_i = 6$, $\tau = 0.1$). As a consequence, the benchmark scenario appears in each panel. The top plots show the results for the intercept, while the bottom row shows results for the slope.

Generally, the magnitude of the bias decreases as the number of observations per cluster increases for fixed N (central panels): this confirms the existing asymptotic results (Koenker, 2004; Canay, 2011). However, when the cluster size, n_i, is fixed (left most panels), there is non-negligible bias for these estimators, as the sample size, N, increases. The results are valid for both parameter components, but in particular for the slope (bottom panel). In other words, the estimators are not consistent for β_1^τ in the asymptotic scenario with a fixed (and small) number of repeated measurements and increasing the number of clusters. The bias behavior is worse for quantile levels closer to the boundaries, $\tau = 0.1$ or $\tau = 0.9$, than for levels closer to the median, $\tau = 0.5$ (right panels).

The three methods are comparable for estimation of β_1^τ whereas there are subtle differences for β_0^τ: LQMM and ℓ_1-penalized estimators behave similarly, except for small values of N; Canay's estimator has bias of opposite sign and of smaller size as well as smaller variation compared to the two other methods. Further simulation scenarios are presented in Tables 2.4 and 2.5 in the appendix, showing similar results.

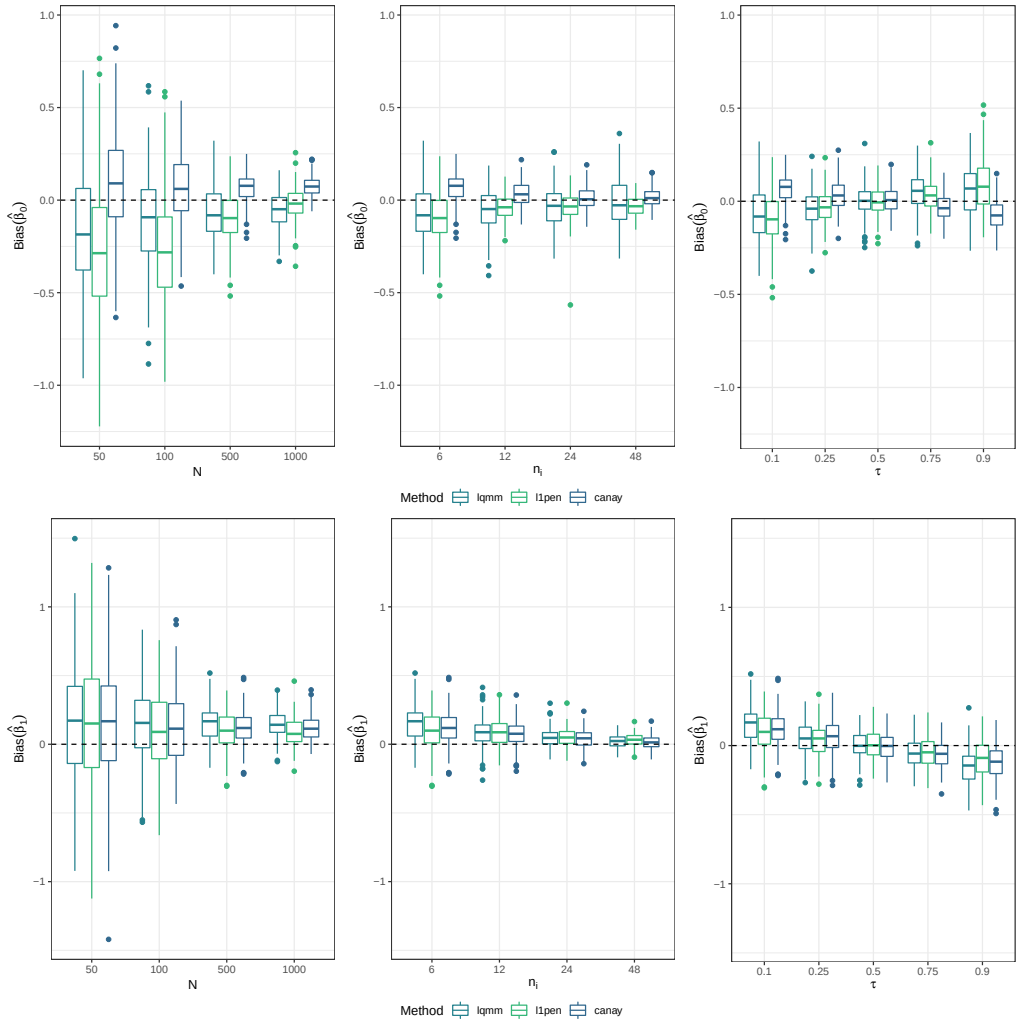

Figure 2.3: Boxplots for estimators of β_0^τ (top) and β_1^τ (bottom) for 200 datasets from a selection of the traditional methods with varying N (left panels), n_i (middle panels) and τ (right panels). The factors that do not vary are kept fixed at benchmark values: $N = 500$, $n_i = 6$, $\tau = 0.1$.

2.4.3 Performance of the proposed estimators

Bias and variation

We now turn to a more detailed study of our proposed estimators. Figure 2.4 has the same structure as Figure 2.3, but now includes the oracle estimator (as an infeasible point of reference), the LQMM estimator (as a representative of the existing methods, cf. Figure 2.3, and as starting point of our two-step procedure), and the unadjusted and adjusted two-step estimators. Results are based on 1000 replications. The benchmark scenario ($N = 500$, $n_i = 6$, $\tau = 0.1$) was also considered in Figure 2.2, but notice that the results of Figure 2.4 summarize performance in 1000 simulations, while only 200 simulations were considered in Figure 2.2, due to the increased computational burden

required by some of the alternative methods.

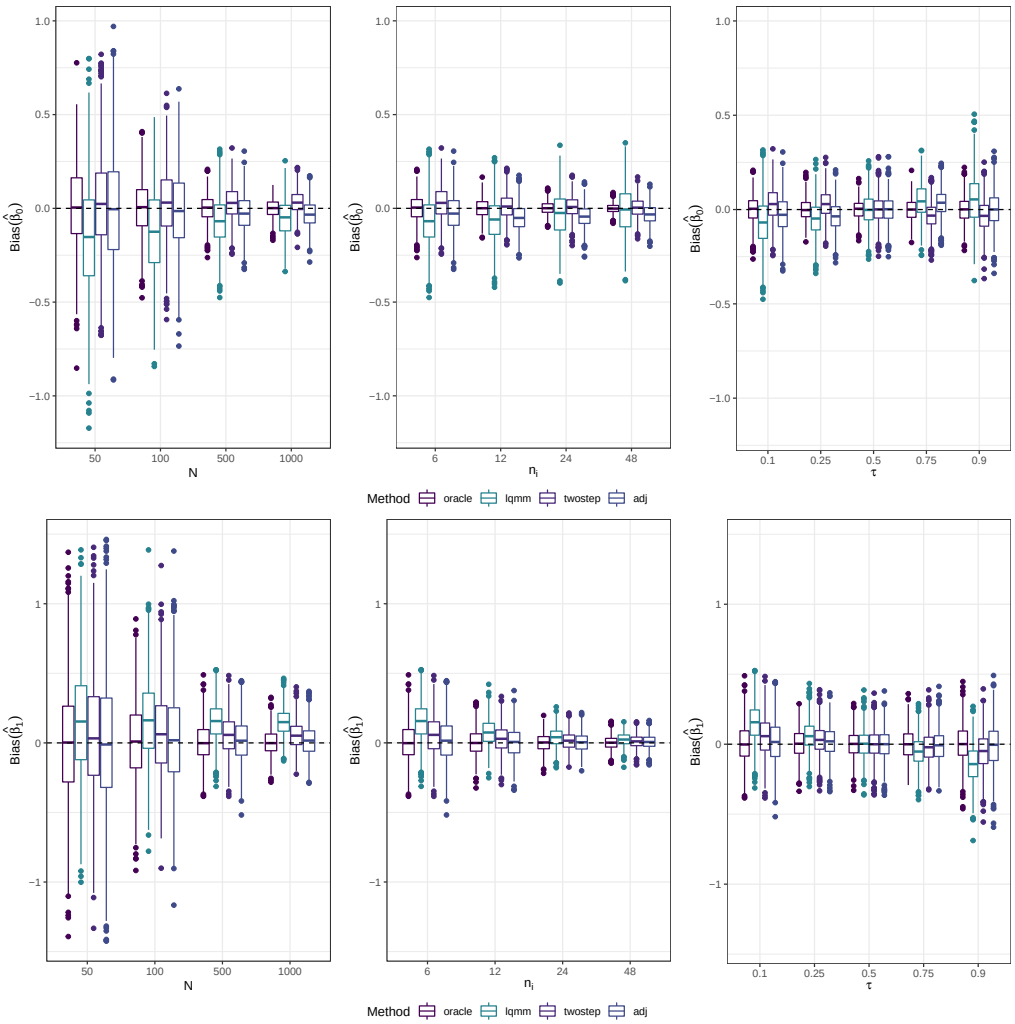

Figure 2.4: Boxplots for estimators of β_0^τ (top) and β_1^τ (bottom) for 1000 datasets from the oracle estimator, the LQMM estimator and the adjusted two-step estimator with varying N (left panels), n_i (middle panels) and τ (right panels). Parameters that do not vary are kept fixed at benchmark values: $N = 500$, $n_i = 6$, $\tau = 0.1$.

For the slope quantile regression parameter, β_1^τ (bottom panels), the bias is reduced for the two-step estimator compared to the LQMM estimator and is almost completely removed in all scenarios for the bias-adjusted estimator. The variability is only slightly larger than the variability of the oracle estimator. For the intercept quantile regression parameter, β_0^τ, the bias is considerably reduced for the proposed two-step estimators compared to the LQMM estimator when the cluster size is small (top left panel). For large clusters the unadjusted two-step estimator shows the best performance in terms of both bias and variance (top central panel).

Results for more combinations of N, n_i and τ are reported in the appendix. For the median, $\tau = 0.5$ (Table 2.6), all three estimators are unbiased and show similar

variability. For $\tau = 0.1$ (Table 2.7), the situation is more complex. Nonetheless, the proposed two-step estimators (without adjustment) yields a smaller RMSE than the LQMM counterpart. Consider the estimation of the slope parameter β_1^τ: all estimators seem to show similar variability, however the two-step estimators indicate a considerably improved bias behavior compared to the LQMM estimator. The numerical studies show that the cluster size has a larger impact on estimation performance than the number of clusters; compare the RMSE when the number of observations is kept fixed to say 3000 composed by 1) $N = 1000$ clusters of size $n_i = 3$ and 2) $N = 500$ clusters of size $n_i = 6$.

Figure 2.5 compares the two-step estimators with the oracle and LQMM for three extra scenarios that have larger heteroscedasticity ($\gamma = 1$) or larger within-cluster relative variance ($\sigma_u^2 = 1.5$, $\sigma_e^2 = 0.5$ yielding ICC $= 0.75$), or larger total variation ($\sigma_u^2 = \sigma_e^2 = 1.5$) compared to the benchmark scenario. All other simulation parameters are kept fixed to the values from the bechmark setting. The changed parameter settings have larger impact on the distribution of the LQMM estimator than on the distribution of the two-step estimators. In particular, the two-step estimation results in improved bias performance compared to the LQMM estimator, irrespective of the setting.

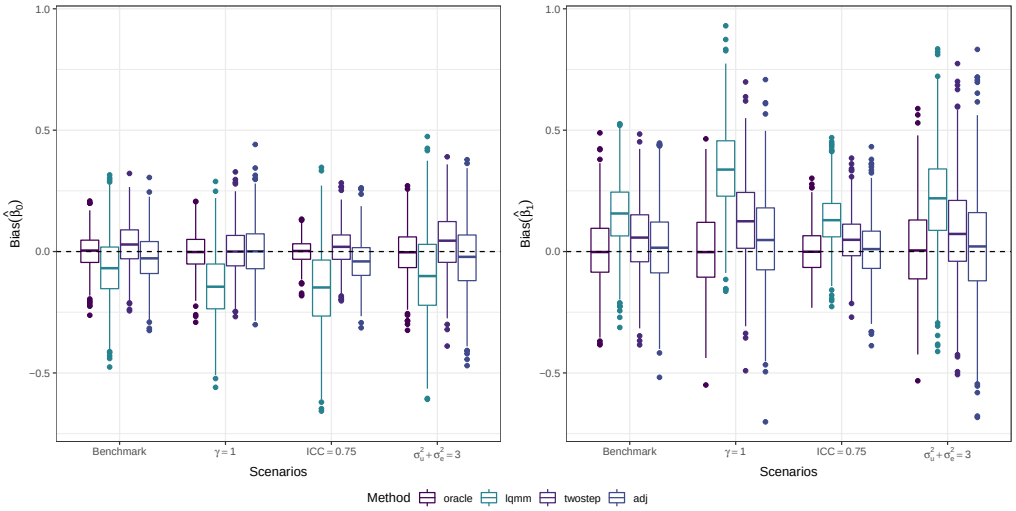

Figure 2.5: Boxplots of the estimates of β_0^τ (left) and β_1^τ (right) obtained using oracle method, LQMM, and the two-step estimators with and without adjustment for two-step estimation for the benchmark scenario and scenarios with larger homoscedasticity, larger ICC, and larger variance. All the other simulation factors are kept constant to their values of the benchmark scenario. Results are based on 200 simulations.

Confidence intervals and comparison of bootstrap strategies

Next, we turn to evaluating the proposed bootstrap scheme for decreasing the estimator's bias and construction of confidence intervals. We compare the proposed mixture of standard and wild resampling (denoted by RW) with other types of data resampling, with respect to bias-adjustment in estimating the parameters, as well as the actual coverage and average length of the confidence intervals.

Resample random effects and residuals (RRR) A bootstrap sample takes the form $\{(Y_{ij}^{*b}, X_{ij}, Z_{ij})_{j=1}^{n_i}, u_i^{\tau,*b}\}_{i=1}^{N}$ where $Y_{ij}^{*b} = Z_{ij}^T u_i^{\tau,*b} + X_{ij}^T \hat{\beta}^\tau_{\text{two-step}} + \varepsilon_{ij}^{*b}$, with ε_{ij}^{*b}

obtained from a standard sampling with replacement procedure from the observed residuals, $\{\varepsilon_{ij}\}_{i,j}$, and $u_i^{\tau,*b}$ is sampled from $\mathcal{U}$. In contrast to RW sampling, there is no coupling between covariates and residuals. Carpenter et al. (2003) has proposed the method for mean regression for multilevel data. Notice that residuals could also be sampled cluster-wise in order to maintain within-cluster dependence not accounted for by the random effect, but we do not consider this.

Resample clusters (RC) The clusters are sampled with replacement in a completely non-parametric way. More specifically, $i_1^*, \ldots, i_N^*$ are sampled with replacement from $\{1, \ldots, N\}$, and a bootstrap dataset consists of $(Y_{ij}^*, X_{ij}^*, Z_{ij}^*) = (Y_{i_i^* j}, X_{i_i^* j}, Z_{i_i^* j})$, $i = 1, \ldots, N,\ j = 1, \ldots, n_i$. Within-cluster dependence is maintained because complete clusters are sampled. The method, also known in the literature as *cross-sectional resampling* (Galvao and Montes-Rojas, 2015), is used by Canay (2011) and Geraci and Bottai (2014) to construct confidence intervals. Karlsson (2009) uses RC in an attempt to correct for estimation bias in a nonlinear quantile regression for longitudinal data, using a marginal perspective, but experienced limited gain.

Cluster-wise wild bootstrap (CW) The idea is to use wild bootstrap for the sum of random effects and error terms. Specifically, let $r_{ij} = Y_{ij} - X_{ij}^T \hat{\beta}^\tau_{\text{two-step}}$ be the residuals corresponding to the two-step estimation, and let w_is be a random sample from (2.3.5). The bootstrap sample is $\{(Y_{ij}^{*b}, X_{ij}, Z_{ij})_{j=1}^{n_i}\}_{i=1}^N$, where $Y_{ij}^{*b} = X_{ij}^T \hat{\beta}^\tau_{\text{two-step}} + w_i |r_{ij}|$. In contrast to the residuals ε_{ij} used for RW, r_{ij} are defined without subtraction of predicted random effects (often referred to as "level zero residuals"). Also, same weight w_i is used for all the observations within cluster i in order to preserve dependence within clusters. This resampling scheme is used by Modugno and Giannerini (2015) in the context of multilevel models for mean regression, but does not appear to have been used for quantile regression.

The RW and RRR sampling schemes use bootstrap to approximate the joint distribution of (u_i^τ, Y_{ij}), whereas the other two bootstrap methods approximate the distribution of Y_{ij} only. As RC- and CW-based approaches do not involve generation of random effects, SE-adjustment confidence intervals are only applicable for RW and RRR. The bias-adjusted estimator and basic confidence intervals, on the other hand, can be computed for any of the four bootstrap schemes.

Table 2.1 shows bias and actual coverage rates for confidence intervals with an intended level of 95%. We employ the benchmark scenario, except for a varying number of clusters (same simulated data as in the left part of Figure 2.4). Results are based on 1000 simulated datasets. SE-adjustment confidence intervals generated with the RW bootstrap method give the best coverage rates, close to the nominal 95% in all scenarios. Basic confidence intervals with RW bootstrap are also good for β_0^τ when N is large, whereas coverage rates are below 0.90 for β_1^τ. RRR and RW produce similar coverage rates for β_1^τ, but no bias-adjustment is obtained with RRR (bias is equivalent to bias for the unadjusted two-step estimator, not reported). For β_0^τ the coverage rates are slightly smaller for RRR compared to RW. As expected, bootstrap method RC gives no bias reduction, neither for β_0^τ nor β_1^τ, and coverage rates are consequently never above 0.90. The CW bootstrap method has poor coverage rates. For β_1^τ the main reason is that the adjusted estimator has large variability which is not properly taken into account, whereas the explanation for β_0^τ is that CW introduces a large bias such that the confidence interval is located far from the true value.

In summary, the semi-parametric bootstrap sampling methods using the additive model structure for the quantiles (RW and RRR) with SE-adjusted confidence intervals

		β_0^τ				β_1^τ			
N		RW	RRR	RC	CW	RW	RRR	RC	CW
50	Bias	-0.01	-0.03	0.03	-0.94	0.01	0.05	0.05	-0.03
	Coverage, basic	0.87	0.90	0.88	0.10	0.86	0.90	0.86	0.42
	Coverage, SE-adj.	0.95	0.93	—	—	0.95	0.96	—	—
100	Bias	-0.02	-0.03	0.03	-0.9	0.03	0.07	0.07	< 0.01
	Coverage, basic	0.89	0.90	0.90	0.02	0.88	0.90	0.89	0.39
	Coverage, SE-adj.	0.95	0.92	—	—	0.94	0.95	—	—
500	Bias	-0.02	-0.03	0.03	-0.92	0.02	0.06	0.06	-0.03
	Coverage, basic	0.94	0.90	0.90	< 0.01	0.89	0.89	0.88	0.36
	Coverage, SE-adj.	0.96	0.91	—	—	0.93	0.92	—	—
1000	Bias	-0.03	-0.02	0.03	-0.91	0.02	0.05	0.05	-0.04
	Coverage, basic	0.94	0.88	0.86	< 0.01	0.89	0.88	0.89	0.31
	Coverage, SE-adj.	0.95	0.89	—	—	0.93	0.90	—	—

Table 2.1: Bias and coverage rates of 95% confidence intervals for the adjusted two-step method for different bootstrap schemes (RW, CW, RC, RRR) for 1000 datasets. Basic confidence intervals are used for all bootstrap schemes, whereas SE-adjusted confidence intervals are only defined for RW and RRR. Cluster size is fixed at $n_i = 6$ and the quantile level is $\tau = 0.1$.

show the best coverage properties. Nonetheless, the proposed two-step with RW-based adjustment results in the greatest bias reduction.

Geraci and Bottai (2014) and Canay (2011) use RC bootstrap for construction of confidence intervals (Canay also uses asymptotic results), and Table 2.2 compares coverage rates and average lengths for their confidence intervals and our SE-adjusted confidence intervals based on RW sampling. The simulated data are the same as those used for Table 2.1. Geraci and Bottai (2014) and Canay (2011) present estimation and inference results regarding different settings than the ones considered here, but our results are well in line with theirs. The LQMM and Canay confidence intervals loose coverage for large N because the estimators are biased. For small N the coverage is close to the nominal level (bias plays a minor role because variation is large), and the confidence intervals are shorter than those based on SE-adjustment, most likely because extra variability is introduced with the bias adjustment.

2.4.4 Additional simulation studies

At the suggestion of an anonymous reviewer, we further investigate the proposed method when the errors e_{ij} are generated from a non-Gaussian distributions. Specifically, we use a scaled t_3-distribution and an $ALD(0, \sigma_0, \tau_0)$ with $\tau_0 = 0.1$ and $\sigma_0 = \frac{(1-\tau_0)\tau_0}{\sqrt{1-2\tau_0+2\tau_0^2}} = 0.09939$. Both distributions are scaled to have unit variance in order to make fair the comparison with the standard normal errors scenarios considered previously. When sampling from the ALD distribution, we consider both the benchmark scenario and a departure from it, corresponding to $\gamma = 0$. Notice that the true values of β_0^τ and β_1^τ change compared to the standard normal case. The results are shown in Table 2.8 in the appendix and should be compared to the relevant scenarios in Table 2.7.

In the case of scaled t-distributed errors, the bias is reduced for the two-step estimator, compared to the LQMM estimator, but it is not completely removed. The RW bootstrap correction reduces the bias even further for β_1^τ, but surprisingly it increases the bias for β_0^τ. This may be due to the inflated residuals that are obtained with the wild bootstrap

N		β_0^τ adj (RW)	β_0^τ lqmm	β_0^τ Canay	β_1^τ adj (RW)	β_1^τ lqmm	β_1^τ Canay
50	Bias	-0.01	-0.16	0.07	0.01	0.15	0.13
	Coverage	0.95	0.93	0.93	0.95	0.94	0.93
	Av. Length	1.29	1.21	0.98	2.11	1.67	1.55
100	Bias	-0.02	-0.13	0.07	0.03	0.17	0.13
	Coverage	0.95	0.90	0.93	0.94	0.93	0.91
	Av. Lengtvh	0.88	0.93	0.70	1.38	1.22	1.09
500	Bias	-0.02	-0.07	0.07	0.02	0.15	0.13
	Coverage	0.96	0.90	0.84	0.93	0.83	0.81
	Av. Length	0.42	0.52	0.31	0.56	0.57	0.48
1000	Bias	-0.03	-0.05	0.07	0.02	0.15	0.13
	Coverage	0.95	0.90	0.73	0.93	0.70	0.69
	Av. Length	0.30	0.40	0.22	0.38	0.41	0.34

Table 2.2: Bias, coverage rates of 95% confidence intervals and average length of confidence intervals for our adjusted two-step method as well as LQMM and Canay's methods for 1000 datasets. Cluster size is fixed at $n_i = 6$ and the quantile level is $\tau = 0.1$.

scheme, as they can be large in the situation of heavy-tailed errors, and therefore have large impact on the estimation of bias for the intercept.

In the case of heteroscedastic ALD errors ($\gamma > 0$), the bias of the LQMM estimator for β_1^τ is reduced considerably compared to the Gaussian case (Table 2.7). The estimators' variability is also reduced in this setting, in spite of the error variance remaining fixed, because quantiles are generally estimated with higher precision when the model is ALD than when it is Gaussian. The two-step estimator and the adjusted two-step estimator have almost the same distributions as the LQMM estimator. For estimating the intercept, the performance of the proposed estimators is superior to that of the LQMM, in terms of reduced bias and variability.

When the errors come from a homoscedastic ALD ($\gamma = 0$), the working distribution for the LQMM estimation approach coincides with the data generating mechanism. As expected, the LQMM estimator of β_1^τ has a very good performance: no bias and small variance. The two-step estimators are also unbiased, but have slightly larger variance. For estimating the intercept parameter, surprisingly, the LQMM estimator shows a behavior comparable to the heteroscedastic ALD case; in contrast the two-step estimators have a much smaller bias and variance.

Finally, we also consider a quantile regression model involving both a random intercept and a random slope. To be specific, the data are generated from the model $Y_{ij} = \beta_0 + u_i + (\beta_1 + v_i)x_{ij} + (1 + \gamma x_{ij})e_{ij}$, where u_i is generated as described in (2.4.1) and $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$. Out of the existing methods, only LQMM allows to incorporate random slopes in the quantile regression; thus we compare the results of the two-step estimation with LQMM solely. Table 2.3 shows the results. We see that irrespective of the sample size or cluster size, the two-step estimation without adjustment improves or maintains the RMSE compared to LQMM estimation. The adjusted two-step estimator generally shows the smallest bias, but at the expense of increased variability; for the estimation of the intercept parameter in the case of $n_i = 12$ the unadjusted two-step estimator has the smallest bias and variance.

			β_0^τ			β_1^τ		
N	n_i		lqmm	two-step	adj (RW)	lqmm	two-step	adj (RW)
500	6	Bias	-0.03	-0.01	0.00	0.15	0.14	0.04
		SD	0.13	0.09	0.12	0.18	0.17	0.24
		RMSE	0.14	0.09	0.12	0.24	0.22	0.24
1000	6	Bias	-0.03	-0.02	-0.02	0.18	0.17	0.07
		SD	0.08	0.08	0.11	0.15	0.15	0.21
		RMSE	0.09	0.08	0.11	0.23	0.22	0.22
500	12	Bias	-0.07	-0.02	-0.05	0.06	0.10	0.04
		SD	0.12	0.07	0.08	0.14	0.11	0.13
		RMSE	0.14	0.07	0.10	0.15	0.15	0.14

Table 2.3: Bias, standard deviation, and RMSE for the LQMM estimator (lqmm), the two-step estimator (two-step), and bootstrap-adjusted two-step estimator (adj) where bootstrap samples are generated with the RW method, and we consider the model with random intercept as well as random slope.The quantile level is $\tau = 0.1$, and results are from 200 replications.

2.5 Data application

AIDS Clinical Trial Group (ACTG) Study 193A (Henry et al., 1998) is a randomized and double-blinded study of patients affected by AIDS at severe immune suppression stage, with CD4 counts of less than 50 cells/mm^3. There are 1309 patients, who were assigned to one of four treatments, namely: 600 mg of zidovudine daily alternating monthly with 400 mg of didanosine (double treatment 1); 600 mg of zidovudine as well as 2.25 mg of zalcitabine, both daily (double treatment 2); 600 mg of zidovudine as well as 400 mg of didanosine, both daily (double treatment 3); the combination of 600 mg of zidovudine, 400 mg of didanosine and 400 mg of nevirapine, all of them daily (triple treatment). The CD4 counts were recorded at a baseline visit and at the follow-up visits during the subsequent 40 weeks. The measurements were intended to be taken every eight weeks, but occasionally there were dropouts or skipped medical appointments; see Figure 2.6. After excluding the subjects with a single measurement (baseline), there are $N = 1187$ subjects remaining in the study; their number of repeated measurements, n_i, varies between two and nine with a median of four. The data has been previously used as an illustrative application for mean regression frameworks in Fitzmaurice et al. (2012) and it is available at the associated webpage (https://content.sph.harvard.edu/fitzmaur/ala2e/).

Our aim is to study the progression of the infection under the four treatment regimes for patients at different stages of immune suppression. Since CD4 counts are proxies for the stage of suppression—with lower CD4 counts corresponding to later stages—this can be obtained by studying the time trend for each treatment at different quantile levels. More specifically, an effective treatment reduces the decrease in CD4 counts, yielding a time trend closer to zero than a less effective treatment, and the effect may be different for early-stage patients (corresponding to high quantile levels) than late-state patients (corresponding to low quantile levels). Figure 2.6 shows that subjects tend to have low or high CD4 counts throughout, suggesting incorporation of subject-specific intercepts in the model.

As it is common in the literature, we log-transform the observed values and denote by Y_{ij} the log(CD4 count + 1) for patient i at the jth hospital visit and by t_{ij} the time of the jth visit, which is recorded by the number of weeks since the patient's baseline visit. We use dummy variables Treat_h $(h = 1, \ldots, 4)$ to indicate the assigned treatment, where Treat_1 corresponds to the triple therapy, and Treat_2, Treat_3 and Treat_4 correspond

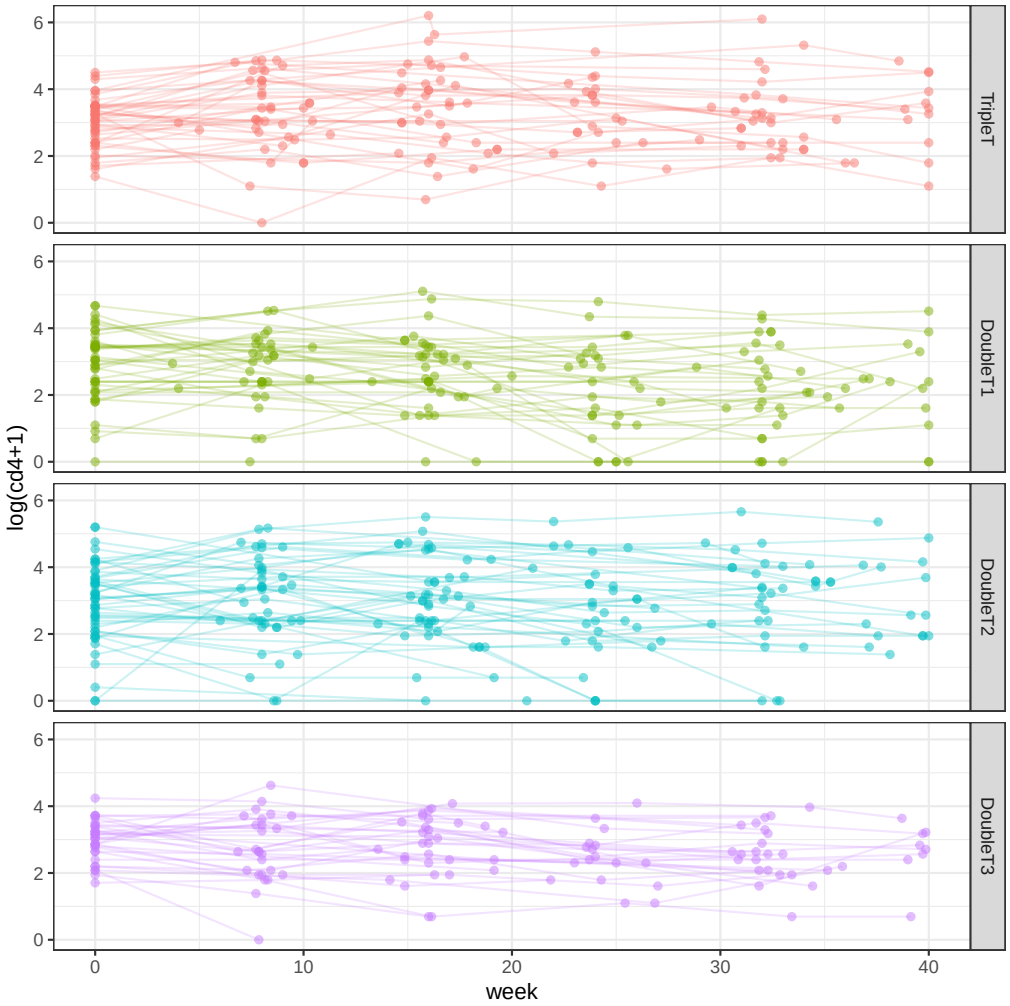

Figure 2.6: Transformed CD4 counts for 200 patients, showing the records of 50 random subjects from each of the four treatment groups. Observations from the same patients are connected with lines.

to the three double treatments. We account for age at baseline (variable Age) and sex (variable Sex, zero for females and one for males) as well. For simplicity of notation, we collect covariates relative to the ith patient at the jth follow-up visits into X_{ij} such that $X_{ij}^T = (\text{Treat}_{1,i}, \text{Treat}_{2,i}, \text{Treat}_{3,i}, \text{Treat}_{4,i}, \text{Age}_i, \text{Sex}_i, t_{ij})$. To study the time-varying effect of treatment at quantile level τ of the response, let u_i^τ be a subject-specific random effect associated with the quantile level τ and posit the following linear quantile regression model:

$$Q_{Y_{ij}|X_{ij},u_i^\tau}(\tau) = \sum_{h=1}^{4} \beta_{0,h}^\tau \cdot \text{Treat}_{h,i} + \sum_{h=1}^{4} \beta_{1,h}^\tau \cdot \text{Treat}_{h,i} \cdot t_{ij} + \beta_2^\tau \cdot \text{Age}_i + \beta_3^\tau \cdot \text{Sex}_i + u_i^\tau. \quad (2.5.1)$$

The slope parameters $\beta_{1,1}^\tau, \ldots, \beta_{1,4}^\tau$ describe the behavior of CD4 counts over time, conditional on subject, and represent the main object of interest. As our interest is in the

time varying effect of each treatment we are using the so-called "explicit parameterization"; as a result, the model specification does not require a common intercept parameter. Estimation and inference are carried out using the proposed two-step estimation with adjustment; the results are compared with LQMM.

The estimated slope parameters for each treatment in part are plotted in Figure 2.7 for varying quantile levels. The left panels show the two-step estimates with adjustment and the corresponding 95% confidence intervals for quantile levels $\tau \in \{0.1, 0.15, \ldots, 0.9\}$ (separate analyses). We used 100 RW bootstrap samples for the computations. The top panels concern the triple treatment: since the confidence band, corresponding to the two-step estimator, includes zero at all the quantile levels, it indicates that this therapy maintains an almost constant CD4 count during the study for subjects at any stage of their condition. For the other three treatments the situation is different. As depicted in the remaining panels, the two-step estimated coefficients $\hat{\beta}^{\tau}_{1,2}$, $\hat{\beta}^{\tau}_{1,3}$ and $\hat{\beta}^{\tau}_{1,4}$ are negative and significant at all the quantile levels, indicating that patients treated with either one of the double therapies must expect to see their CD4 count decrease over time. Notice that there is a slight increase in the estimated $\hat{\beta}^{\tau}_{1,2}$ over quantile levels, which indicates that double treatment 1 makes the CD4 counts decrease faster for patients in the most severe conditions (lower quantile levels), whereas double treatments 2 and 3 appear to have more homogenous effects across patient groups.

In order to compare the treatments more directly we consider contrasts of the form $\hat{\beta}^{\tau}_{1,h} - \hat{\beta}^{\tau}_{1,1}$, which describe the difference in the effects between each double treatment and the triple treatment at quantile level τ. The middle panels in Figure 2.7 show the estimated contrasts and the corresponding 95% confidence intervals. Except for a single quantile level for double treatment 3, confidence intervals exclude zero, showing that the triple therapy is the most efficient treatment for patients in all infection stages. Fitzmaurice et al. (2012) reported similar results for the mean.

For comparison, the LQMM estimates and confidence intervals for the contrasts are shown in the right panels of Figure 2.7. Confidence intervals are based on 100 RC bootstrap samples. LQMM estimates are in the same range as the adjusted two-step estimates, albeit in general closer to zero. Moreover, the confidence bands are much wider, implying that the LQMM method does not find evidence for significant treatment differences for double treatments 2 and 3. This should not be surprising, since our numerical investigation showed that LQMM confidence intervals are wider (and coverage lower) than those corresponding to the adjusted two-step estimator, when the number of subjects is much larger than the number of repeated measurements; recall Table 2.1.

While these results are interesting, we acknowledge one aspect of the data that our analysis does not account for: missing data. Out of the 1187 patients in the study, only 795 of them have measurements past the 30th week since their baseline. Missing data is not uncommon in ACTG studies and previous quantile regression analyses with longitudinal data have approached the problem by incorporating weights into the estimating equations (Lipsitz et al., 1997), employing hierarchical Bayesian models (Huang and Chen, 2016; Feng et al., 2011), or by considering a linear quantile mixed hidden Markov model with a missing data indicator (Marino et al., 2018). Incorporation of such methods falls beyond the scope of this paper, but could be an interesting avenue for future research.

2.6 Discussion

We have identified a gap in the literature concerning mixed effects models for quantile regression for clustered data: existing estimation methods may yield severely biased

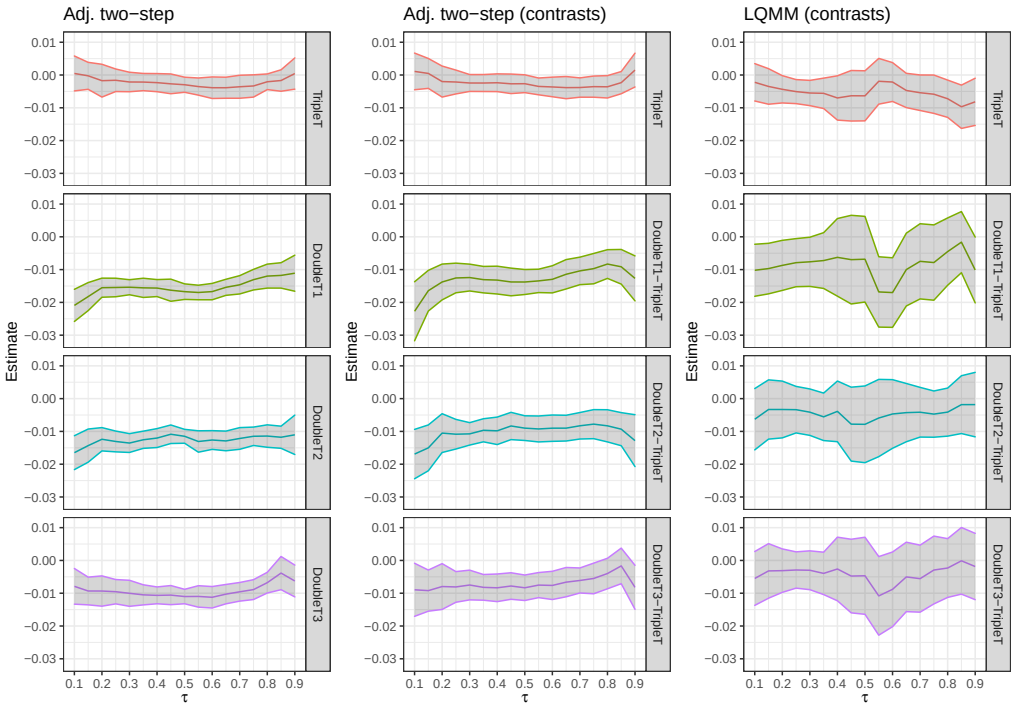

Figure 2.7: Estimated coefficients and pointwise 95% confidence bands at varying τ for model (2.5.1). The left panels show results for slope coefficients $\beta^{\tau}_{1,h}$ ($h = 1, \ldots, 4$, adjusted two-step method) whereas the central and right panels show results for contrasts with triple therapy as reference (adjusted two-step method in the centre, LQMM to the right).

estimators for fixed effects parameters in situations with many, but small clusters. In this paper, we propose a new estimation method that relies on predicted random effects computed by using an LQMM working framework (in particular, at the quantile level of interest), standard quantile regression with offsets, and a bias-adjustment by means of a novel bootstrap sampling technique. In the simulation study, the proposed estimator shows considerably smaller bias compared to the available competitors, especially in situations with small clusters. The RW adjustment appears to be particularly beneficial for estimating slope parameters, while the results are less clear for the intercept and could be studied further. The two-step estimation procedure may be seen as the onset in an iterative procedure alternating between estimation of the regression parameters for fixed random effects and prediction of random effects for fixed regression parameters. An ALD working model with random effects only (no fixed effects) can be used in the second step, and this requires minor modifications of the current implementation of the `lqmm()` function.

Hitherto, the literature for quantile regression for clustered data has focused on studying asymptotics for increasing both the number of clusters and the cluster size (Koenker, 2004; Kato et al., 2012; Canay, 2011; Besstremyannaya and Golovan, 2019). In such case, the cluster-specific parameters are asymptotically "eliminated" as stated by Canay (2011) or "concentrated out" as stated by Kato et al. (2012) and act as known quantities for the asymptotics of β^{τ}. On the other hand, the theoretical study of the

estimators is inherently challenging, when cluster size is fixed, and only the number of clusters increases to infinity. Results from (generalized) linear mixed models do not carry over for primarily two reasons. First, the criterion functions constructed from the check function is not differentiable. Second, the distributional assumptions are typically held to the minimum and focus on the relationship between the covariates and the quantile of interest. In particular, Geraci and Bottai (2007, 2014) do not mention any attempts to derive asymptotic results for the LQMM estimator and rely on bootstrap methods for inference. Neither do we provide asymptotic results for our estimators, nor claim that bias is *removed* asymptotically. The main difficulty lies in the prediction accuracy of the random effect predictors, which are used as one of the main ingredients in the bootstrap sampling procedure. If the predicted random effects do not accurately capture the variation of the cluster-specific random effects, then the estimated bias may not represent the bias of the unadjusted estimator. Therefore, when we are neither assuming an increasing cluster size nor considering a specific data generating model, then it is difficult to prove asymptotic results for our estimators, and we leave this for future research.

Mean regression models for longitudinal data often incorporate more complex within-subject dependence structures than the one modeled by random intercepts alone (compound symmetry). Similar attempts do not seem to exist for quantile regression. The two-step estimator is not readily modified to take a serial dependence into account, but the RW bootstrap sampling could be easily adapted such as by sampling the weights for wild bootstrap at the subject level rather than at the measurement level. Moreover, longitudinal studies may involve drop-outs and occasional missing data, with data not missing at random, and how to incorporate such missingness in quantile regression in an appropriate way remains an open research problem.

One direction that the proposed methodology opens up is to consider quantile regression for time series data (one long series rather than many shorter series), see Xiao (2017). In such case, the quantile model would be $Q_{Y_t|X_t}(\tau) = X_t^T\beta^\tau + u_t^\tau$ where Y_t and X_t denote the response and covariate, respectively, at time t ($t = 1, \ldots, T$), and $\{u_t^\tau\}_{t=1,\ldots,T}$ is a latent series which describes (random) fluctuations of quantiles over time. Another direction is to extend the approach to multi-level data with multiple levels of nested random effects or data with several, but non-nested random effects. The ideas behind the methods from this paper (existing as well as our proposed method) would carry over to such situations, but a rigorous investigation of this extension is left for future research.

2.7 Acknowledgements

The project was partly funded by the Danish Research Council (DFF grant 7014-00221).

2.8 Appendix

The appendix contains additional numerical results from the simulation study with data generated from model (2.4.1). The results are discussed in the main text. Tables 2.4 and 2.5 compare various existing approaches when both the number of clusters and the cluster size vary; other simulation parameters are specified by their level at the benchmark scenario. Estimation is carried out for quantile levels $\tau = 0.5$ (Table 2.4) and $\tau = 0.1$ (Table 2.5), respectively, with results based on 200 replications. It is not possible to compute the jackknife estimator when $n_i = 3$ because clusters cannot be split into two subsets with several observations per cluster. Furthermore, in the scenario with $N = 1000$,

$n_i = 12$ and $\tau = 0.1$ there were convergence problems for the ℓ_1-penalized estimator for two datasets, and the results for this estimator are based on the remaining 198 replications. Table 2.6 and Table 2.7 have the same structure as described above and consider the same scenarios; they evaluate the performance of the LQMM estimator and our two proposed methods in 1000 replications. Notice the difference in the number of replications; as mentioned in Section 2.4.2 it is due to the computational burden of some of the traditional estimators. Finally, Table 2.8 summarizes the results for the case when the error terms in (2.4.1) are either sampled from a scaled t-distribution in the benchmark scenario, from an ALD distribution in the benchmark scenario or an ALD distribution when $\gamma = 0$. Results correspond to the quantile level $\tau = 0.1$ and are based on 200 replications.

					β_0^τ						β_1^τ			
N	n_i		oracle	Canay's	jackknife	ℓ_1-pen	ℓ_2-pen	marg	oracle	Canay's	jackknife	ℓ_1-pen	ℓ_2-pen	marg
500	3	Bias	< 0.01	0.01	—	0.01	0.01	0.01	-0.01	-0.02	—	-0.02	-0.02	-0.02
		SD	0.07	0.08	—	0.09	0.09	0.09	0.14	0.15	—	0.15	0.15	0.17
		RMSE	0.07	0.08	—	0.09	0.09	0.09	0.14	0.15	—	0.15	0.15	0.17
1000	3	Bias	< 0.01	0.01	—	0.01	0.01	0.01	< 0.01	< 0.01	—	-0.01	-0.01	-0.01
		SD	0.05	0.06	—	0.07	0.07	0.07	0.09	0.10	—	0.11	0.11	0.12
		RMSE	0.05	0.06	—	0.07	0.07	0.07	0.09	0.10	—	0.11	0.11	0.12
500	6	Bias	< 0.01	< 0.01	-0.01	< 0.01	< 0.01	0.01	-0.01	0.01	0.01	-0.01	< 0.01	-0.02
		SD	0.05	0.07	0.11	0.07	0.07	0.08	0.11	0.10	0.13	0.11	0.11	0.13
		RMSE	0.05	0.07	0.11	0.07	0.07	0.08	0.11	0.10	0.13	0.11	0.11	0.13
1000	6	Bias	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	-0.01	< 0.01	-0.01	< 0.01	-0.01	-0.01
		SD	0.03	0.05	0.08	0.05	0.05	0.05	0.07	0.07	0.08	0.07	0.08	0.09
		RMSE	0.03	0.05	0.08	0.05	0.05	0.05	0.07	0.07	0.08	0.07	0.08	0.09
500	12	Bias	< 0.01	< 0.01	< 0.01	-0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
		SD	0.03	0.06	0.13	0.07	0.06	0.07	0.07	0.07	0.09	0.07	0.07	0.09
		RMSE	0.03	0.06	0.13	0.07	0.06	0.07	0.07	0.07	0.09	0.07	0.07	0.09
1000	12	Bias	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
		SD	0.03	0.04	0.10	0.05	0.04	0.05	0.05	0.05	0.06	0.05	0.05	0.06
		RMSE	0.03	0.04	0.10	0.05	0.04	0.05	0.05	0.05	0.06	0.05	0.05	0.06

Table 2.4: Bias, standard deviation, and RMSE for the oracle, Canay's, the jackknife, the ℓ_1-penalized, the ℓ_2-penalized and the marginal estimators. The quantile level is $\tau = 0.5$, and results are based on 200 replications.

N	n_i		oracle	Canay's	β_0^τ jackknife	ℓ_1-pen	ℓ_2-pen	marg	oracle	Canay's	β_1^τ jackknife	ℓ_1-pen	ℓ_2-pen	marg
500	3	Bias	< 0.01	0.16	—	-0.42	-0.48	-0.51	< 0.01	0.23	—	0.10	0.10	0.09
		SD	0.10	0.10	—	0.14	0.12	0.13	0.18	0.17	—	0.21	0.20	0.23
		RMSE	0.10	0.19	—	0.44	0.50	0.53	0.18	0.29	—	0.23	0.23	0.25
1000	3	Bias	< 0.01	0.16	—	-0.38	-0.46	-0.52	< 0.01	0.24	—	0.11	0.12	0.11
		SD	0.07	0.07	—	0.10	0.09	0.10	0.14	0.13	—	0.15	0.16	0.18
		RMSE	0.07	0.18	—	0.40	0.47	0.53	0.13	0.27	—	0.19	0.20	0.21
500	6	Bias	0.01	0.07	-0.15	-0.10	-0.36	-0.53	-0.01	0.11	0.05	0.09	0.10	0.12
		SD	0.07	0.08	0.22	0.13	0.09	0.10	0.13	0.13	0.21	0.14	0.14	0.16
		RMSE	0.07	0.11	0.27	0.17	0.37	0.54	0.13	0.17	0.22	0.17	0.17	0.20
1000	6	Bias	< 0.01	0.07	-0.14	-0.02	-0.34	-0.53	< 0.01	0.13	0.03	0.09	0.10	0.12
		SD	0.04	0.06	0.19	0.10	0.06	0.07	0.09	0.10	0.16	0.10	0.10	0.12
		RMSE	0.04	0.09	0.23	0.10	0.35	0.53	0.09	0.16	0.17	0.14	0.14	0.17
500	12	Bias	< 0.01	0.03	-0.06	-0.04	-0.34	-0.52	< 0.01	0.06	-0.01	0.07	0.10	0.12
		SD	0.05	0.07	0.21	0.09	0.08	0.09	0.10	0.09	0.15	0.09	0.10	0.12
		RMSE	0.05	0.08	0.22	0.10	0.35	0.53	0.09	0.11	0.15	0.12	0.14	0.17
1000	12	Bias	< 0.01	0.04	-0.04	-0.03	-0.34	-0.52	< 0.01	0.06	< 0.01	0.07	0.10	0.12
		SD	0.03	0.05	0.16	0.08	0.06	0.06	0.07	0.07	0.10	0.07	0.07	0.09
		RMSE	0.03	0.06	0.17	0.08	0.34	0.52	0.06	0.09	0.10	0.10	0.12	0.15

Table 2.5: Bias, standard deviation, and RMSE for the oracle, Canay's, the jackknife, the ℓ_1-penalized, the ℓ_2-penalized and the marginal estimators. The quantile level is $\tau = 0.1$, and results are based on 200 replications.

			β_0^τ			β_1^τ		
N	n_i		lqmm	two-step	adj (RW)	lqmm	two-step	adj (RW)
500	3	Bias	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
		SD	0.09	0.09	0.09	0.14	0.15	0.15
		RMSE	0.09	0.09	0.09	0.14	0.15	0.15
1000	3	Bias	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
		SD	0.07	0.06	0.07	0.10	0.11	0.11
		RMSE	0.07	0.06	0.07	0.10	0.11	0.11
500	6	Bias	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
		SD	0.07	0.06	0.07	0.10	0.10	0.10
		RMSE	0.07	0.06	0.07	0.10	0.10	0.10
1000	6	Bias	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
		SD	0.06	0.05	0.05	0.07	0.07	0.07
		RMSE	0.06	0.05	0.05	0.07	0.07	0.07
500	12	Bias	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
		SD	0.09	0.06	0.06	0.07	0.07	0.07
		RMSE	0.09	0.06	0.06	0.07	0.07	0.07
1000	12	Bias	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
		SD	0.06	0.04	0.04	0.05	0.05	0.05
		RMSE	0.06	0.04	0.04	0.05	0.05	0.05

Table 2.6: Bias, standard deviation, and RMSE for the LQMM estimator (lqmm), the two-step estimator (two-step), and bootstrap-adjusted two-step estimator (adj) where bootstrap samples are generated with the RW method. The quantile level is $\tau = 0.5$, and results are based on 1000 replications.

			β_0^τ			β_1^τ		
N	n_i		lqmm	two-step	adj (RW)	lqmm	two-step	adj (RW)
500	3	Bias	0.02	0.09	0.06	0.25	0.10	0.05
		SD	0.16	0.11	0.14	0.21	0.20	0.23
		RMSE	0.16	0.15	0.15	0.33	0.22	0.23
1000	3	Bias	0.04	0.09	0.05	0.26	0.10	0.05
		SD	0.11	0.08	0.10	0.15	0.14	0.16
		RMSE	0.12	0.12	0.12	0.30	0.18	0.17
500	6	Bias	-0.07	0.03	-0.02	0.15	0.06	0.02
		SD	0.13	0.08	0.10	0.15	0.14	0.15
		RMSE	0.15	0.09	0.10	0.21	0.15	0.16
1000	6	Bias	-0.05	0.03	-0.03	0.15	0.05	0.02
		SD	0.10	0.06	0.07	0.10	0.09	0.11
		RMSE	0.11	0.07	0.08	0.18	0.11	0.11
500	12	Bias	-0.06	0.01	-0.05	0.08	0.03	< 0.01
		SD	0.12	0.07	0.07	0.10	0.10	0.11
		RMSE	0.13	0.07	0.09	0.12	0.10	0.11
1000	12	Bias	-0.05	0.01	-0.05	0.07	0.03	< 0.01
		SD	0.09	0.05	0.05	0.07	0.07	0.08
		RMSE	0.11	0.05	0.07	0.10	0.07	0.08

Table 2.7: Bias, standard deviation, and RMSE for the LQMM estimator (lqmm), the two-step estimator (two-step), and bootstrap-adjusted two-step estimator (adj) where bootstrap samples are generated with the RW method. The quantile level is $\tau = 0.1$, and results are based on 1000 replications.

				β_0^τ			β_1^τ		
e_{ij}	N	n_i		lqmm	two-step	adj (RW)	lqmm	two-step	adj (RW)
			Bias	-0.27	-0.04	-0.15	0.14	0.00	0.00
t_3	500	6	SD	0.14	0.08	0.10	0.12	0.11	0.14
			RMSE	0.30	0.09	0.18	0.18	0.12	0.14
			Bias	-0.25	-0.05	-0.16	0.15	0.06	0.01
t_3	1000	6	SD	0.11	0.06	0.07	0.09	0.09	0.12
			RMSE	0.28	0.07	0.17	0.17	0.11	0.12
			Bias	-0.18	-0.05	-0.15	0.10	0.04	0.02
t_3	500	12	SD	0.13	0.07	0.08	0.08	0.09	0.11
			RMSE	0.22	0.08	0.17	0.13	0.10	0.11
			Bias	-0.12	-0.10	-0.05	0.06	0.05	0.04
ALD	500	6	SD	0.13	0.06	0.07	0.06	0.08	0.09
			RMSE	0.17	0.12	0.09	0.08	0.09	0.10
			Bias	-0.08	-0.10	-0.06	0.05	0.05	0.04
ALD	1000	6	SD	0.10	0.04	0.05	0.05	0.05	0.06
			RMSE	0.12	0.11	0.07	0.07	0.07	0.07
			Bias	-0.08	-0.07	-0.03	0.03	0.03	0.02
ALD	500	12	SD	0.12	0.05	0.05	0.04	0.04	0.05
			RMSE	0.15	0.09	0.06	0.05	0.05	0.05
			Bias	-0.13	-0.06	-0.04	0.00	0.00	-0.01
ALD	500	6	SD	0.14	0.05	0.06	0.05	0.06	0.07
($\gamma = 0$)			RMSE	0.19	0.08	0.07	0.05	0.06	0.07
			Bias	-0.07	-0.07	-0.04	0.00	0.00	0.00
ALD	1000	6	SD	0.11	0.04	0.05	0.04	0.04	0.05
($\gamma = 0$)			RMSE	0.13	0.08	0.06	0.04	0.04	0.05
			Bias	-0.07	-0.05	-0.02	0.00	0.00	0.00
ALD	500	12	SD	0.13	0.05	0.05	0.03	0.04	0.04
($\gamma = 0$)			RMSE	0.14	0.07	0.06	0.03	0.04	0.04

Table 2.8: Bias, standard deviation, and RMSE for the LQMM estimator (lqmm), the two-step estimator (two-step), and bootstrap-adjusted two-step estimator (adj) where bootstrap samples are generated with the RW method. The residuals are sampled from a scaled t_3 when $\gamma = 0.4$ (top part), and from an ALD when either $\gamma = 0.4$ (central part) or $\gamma = 0$ (bottom part). The quantile level is $\tau = 0.1$, and results are based 200 replications.

Chapter 3

Quantile regression for longitudinal functional data with application to feed intake of lactating sows

Maria Laura Battagliola, Helle Sørensen, Anders Tolver & Ana-Maria Staicu

In progress

Abstract

Our work is motivated by a study on lactating sows, where the main interest is about the influence of temperature, measured throughout the day, on the lower quantiles of the feed intake. We propose a model framework and estimation methodology for quantile regression in scenarios with clustered or longitudinal data and functional covariates. The proposed quantile regression model includes subject-specific intercepts to incorporate within-subject dependence, and it allows for time-varying coefficient functions. Estimation relies on basis representations of the unknown coefficient functions, either with a spline basis or a data-driven basis, and can be carried out with existing software. The proposed method is studied numerically in a simulation study that covers a wide range of situations, and we introduce bootstrap procedures for bias adjustments and computation of standard errors. Analysis of the lactation data indicates, among others, that the influence of temperature increases during the lactation period.

Keywords: Bootstrap; Clustered data; Functional principal component analysis; Penalized splines, Subject-specific effects

3.1 Introduction

This paper studies quantile regression for clustered or longitudinal data in the presence of functional covariates. It is motivated by data on the feed intake of lactating sows, where the aim is to study how temperature in the stable, or cell, during the day affects the feed intake, in particular for sows that eat scarcely. This is of interest because poor nutrition in the lactation period may lead to health downsides, both for the sows and the piglets, and production inefficiency. Temperature is measured every fifth minute and is therefore naturally treated as a functional covariate, and the study is longitudinal since since feed intake and temperature is registered over up to 21 days for each sow.

Quantile regression, first introduced by Koenker and Bassett Jr (1978), is a well-established framework from statistics and econometrics. It is suitable when the analysis aims at describing and quantifying the association between covariates and quantiles of the distribution of the response variable. In particular, it allows to robustly target not only the central parts of the response distribution, but also the more extreme regions. For overviews, see the seminal monograph by Koenker (2005b) and Koenker et al. (2017) for more recent developments.

Analyses of longitudinal data, including quantile regression, must account for the dependence between observations from the same subject in order to provide valid results. A common approach is to include subject-specific effects in the model for the quantiles and use penalization, see for example Koenker (2004), Lamarche (2010), Harding and Lamarche (2017), Gu and Volgushev (2019), and Fasiolo et al. (2020), and we adopt the same approach for this paper. Alternatives include Kato et al. (2012) and Galvao and Kato (2016), who treated subject-specific parameters as fixed effects without penalization, and Canay (2011), who used a two-step procedure where subject-specific parameters are first estimated as fixed effects and then plugged in as offsets in a standard quantile regression (see also Besstremyannaya and Golovan (2019)). The close link between the loss function used in quantile regression and the log-density for the asymmetric Laplace distribution (ALD, Yu and Zhang (2005)) has been used to define a working model where subject-specific effects could be integrated out (Geraci and Bottai, 2014), for hierarchical Bayesian models (Luo et al., 2012), or for an EM algorithm (Galarza et al., 2017). Moreover, Battagliola et al. (2021) proposed a bias-adjustment to the estimator from (Geraci and Bottai, 2014), and we use the same idea in the application.

Quantile regression for functional covariates, similar to scalar-on-function mean regression, quantifies the association with the functional covariate involving the integral $\int \beta(s)X(s)\,ds$ for an unknown coefficient function $\beta(\cdot)$. As it is common in nonparametric regression, we approximate the function using finite basis representations for $\beta(\cdot)$, and thus the infinite-dimensional estimation problem is converted to a finite-dimensional one. Pre-specified spline functions and eigenfunctions obtained from the spectral decomposition of the functional covariates' covariance operator are the most popular choices for selecting the basis functions, and they have both been used for quantile regression. For example, Cardot et al. (2005) and Park et al. (2019) used splines, whereas Kato (2012), Chen and Müller (2012) and Li et al. (2016) used eigenfunctions. A related research area is additive quantile regression where the effect of a scalar covariate is modeled via a smooth function (Fenske et al., 2013; Greven and Scheipl, 2017; Geraci, 2019; Fasiolo et al., 2020).

In this paper we propose functional quantile regression for scalar response and functional covariates, which are both observed repeatedly for multiple clusters or subjects. To the best of our knowledge no papers in the literature are devoted to this situation. We first consider a set-up with clustered data, and then extend to a longitudinal set-up, where we account for the time at which the repeated measures are made and furthermore by

using a coefficient function that evolves over time. We contrast the spline approach and the eigenfunction approach to handle the functional covariates and use penalized cluster- or subject-specific intercepts to account for the dependence within clusters or subjects. The resulting model can be represented in a framework that can be easily implemented using existing software (Fasiolo et al., 2020).

The main contributions of the paper are threefold: First, we develop modeling and associated estimation methodology for quantile regression in the complex sampling situation involving scalar response and functional covariates, both observed repeatedly. The proposed method is studied numerically in simulation studies that cover a wide range of situations. Second, we point out bias and variance issues of the estimators and propose adjustments obtained with bootstrap, using resampling techniques from Battagliola et al. (2021) for bias adjustment and from Galvao and Montes-Rojas (2015) for computation of standard errors. Third, with the new methodology we are able to give further insight to the eating behavior of lactating sows in the application. In particular, the analysis indicates that the association between temperature in the stable becomes stronger as time goes by after delivery.

The paper is structured as follows: In Section 3.2 we introduce the model framework, in particular the model with cluster-specific intercepts for clustered data. We describe the estimation methodology in Section 3.3 and the practical implementation in Section 3.4. We study the estimation methods on simulated data in Section 3.5, and devote Section 3.6 to the analysis of the lactation data. Finally, we summarise and discuss finding in 3.7, Additional material can be found in Section 3.9.

3.2 Framework

3.2.1 Quantile regression model

In this paper we focus on quantile regression for clustered (and longitudinal) data as well as on the inclusion of functional covariates. We consider scalar responses and functional predictors $\{(Y_{ij}, X_{ij}(\cdot))\}_{ij}$, where $i = 1, \dots, N$ denote clusters, and $j = 1, \dots, n_i$ denote repeated measurements within cluster i. Covariates $X_{ij}(\cdot)$ are square-integrable functions with domain $S \subset \mathbb{R}$, i.e., $X_{ij}(\cdot) \in L^2(S)$.

We model covariate effects through integrals $\int_S \beta^\tau(s) X_{ij}(s)\, ds$ for an unknown coefficient function $\beta^\tau : S \to \mathbb{R}$. We assume that a change in the functional covariates affect all clusters in the same way, but allow for cluster-specific terms such that all observations within a cluster can have a higher/lower quantile compared to an average cluster. For a fixed quantile level $\tau \in (0, 1)$, we therefore consider the following quantile regression model:

$$Q_{Y_{ij}|X_{ij},u_i^\tau}(\tau) = u_i^\tau + \alpha^\tau + \int_S \beta^\tau(s) X_{ij}(s)\, ds, \qquad i = 1, \dots, N,\ j = 1, \dots, n_i. \tag{3.2.1}$$

Similar to mixed models, the cluster-specific terms u_i^τ are considered as random variation between clusters. This is also indicated by the notation: $Q_{Y_{ij}|X_{ij},u_i^\tau}(\tau)$ is the quantile in the conditional distribution of Y_{ij} given $X_{ij}(\cdot)$ and u_i^τ. From this point of view, the cluster-specific intercepts introduce correlation between repeated measures from the same cluster, while clusters are assumed to be independent.

The primary interest lies in the quantile regression coefficient $\beta^\tau(\cdot)$ which determines the predicted effect of a change in the functional covariate on the level τ quantile common to all clusters. In terms of an intervention study the target parameter allows us to

determine the causal effect on the level τ quantile for all subjects/clusters in response to a change of the functional covariate. Importantly, this may be different from the population averaged marginal change of the level τ quantile. We elaborate on this point in Section 3.2.2.

The regression coefficient $\beta^\tau(\cdot)$ is identifiable only up to an additive component in the orthogonal complement of the vector space spanned by the functional covariates $X_{ij}(\cdot)$. Further, equation (3.2.1) does not specify the full conditional distribution of Y_{ij}, only its τ-quantile. We are going to use it for one or a few quantiles levels of particular interest, but further restrictions could be imposed to avoid crossing quantiles if necessary. In particular, a common u_i^τ across all τ would correspond to shifts of the whole distributions between clusters.

Our simplest scenario consists of model (3.2.1) in combination with observations $X_{ij}(s_h)$ of the covariate functions on a dense grid $\{s_1, \ldots, s_H\} \subset S$ for a large H. Further complexity is introduced when data are longitudinal, such that the repeated measurements for each cluster (subject) are observed in a chronological order along time. We introduce dependence of time into the α^τ and $\beta^\tau(\cdot)$ coefficients of (3.2.1), and the added complexity allows us to predict quantiles along the longitudinal time. This is vital for our application. Moreover, in most real world applications we only have access to covariates observed with noise and/or covariates that are incomplete. Then, we perform smoothing before proceeding with estimation.

3.2.2 Comparison between conditional and marginal quantile models

Before we move on to estimation, we emphasize the distinction between conditional and marginal quantile models, in particular that the marginal quantiles may not inherit the linear structure from the conditional ones. This is important to bear in mind if we aim at the conditional model, since a marginal estimator, ignoring dependence within cluster, may lead to bias.

We illustrate the point with the following example, which generalizes a model from Battagliola et al. (2021) to include a functional covariate. Assume that the response is generated as

$$Y_{ij} = u_i + \alpha + \int_S \beta(s) X_{ij}(s) ds + \left(1 + \gamma \int_S X_{ij}(s) ds\right) e_{ij},$$

where $\gamma \geq 0$ and $1 + \gamma \int_S X_{ij}(s) ds > 0$ with probability one, and $u_i \overset{iid}{\sim} N(0, \sigma_u^2)$ are mutually independent of $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$. Then, the level τ conditional quantile of Y_{ij} given both $X_{ij}(\cdot)$ and u_i takes the form (3.2.1) with intercept $\alpha^\tau = \alpha + \sigma_e \Phi^{-1}(\tau)$ and functional coefficient given by $\beta^\tau(s) = \beta(s) + \sigma_e \gamma \Phi^{-1}(\tau)$. On the other hand, the quantile of Y_{ij} only conditional on $X_{ij}(\cdot)$, is

$$mQ_{Y_{ij}|X_{ij}}(\tau) = \alpha + \int_S \beta(s) X_{ij}(s) ds + \sigma_e \Phi^{-1}(\tau) \sqrt{\frac{\sigma_u^2}{\sigma_e^2} + \left(1 + \gamma \int_S X_{ij}(s) ds\right)^2}. \quad (3.2.2)$$

We will refer to this as the marginal quantile even though it is conditional on $X_{ij}(\cdot)$, and thus distinguish between conditional and marginal quantiles depending on whether conditioning with respect to u_i takes place or not.

If there is no dependence across repeated measurements ($\sigma_u^2 = 0$ and all u_is equal to zero) then the conditional and the marginal quantile coincide. Moreover, the coefficient functions in the conditional and marginal model coincide at the median ($\tau = 0.5$), and

equal β from the data generating model, since $\Phi^{-1}(0.5) = 0$. However, if $\sigma_u^2 > 0$ and $\tau \neq 0.5$, then the dependence of $X_{ij}(\cdot)$ takes a different functional form in the marginal quantiles; it is not even a functional linear relationship.

A linearization of the square root in (3.2.2) as a function of $\int_S X_{ij}(s)ds$ can give an indication of the target value for the marginal linear quantile regression model. In particular, if σ_u^2/σ_e^2 is large compared to the variation of $(1 + \gamma \int_S X_{ij}(s)ds)^2$, then the regression coefficient function from the linearised marginal model differs from the regression coefficient function from the conditional model; the deviation is constant over S with the sign depending on $\Phi^{-1}(\tau)$.

Thus, a marginal analysis, that ignores the cluster structure, has a different target than the conditional model, and that it is therefore important to incorporate the cluster structure in the estimation process if we aim at the conditional quantile. We return to the model in Section 3.5.2.

3.3 Estimation methodology

Two main challenges arise for the estimation of the model (3.2.1) compared to classical quantile regression for independent data with scalar covariates: how to handle the cluster-specific intercepts u_i^τ and how to represent the functional coefficient $\beta^\tau(\cdot)$ appearing in the integral. We manage the cluster-specific effects by regularization and penalize the u_i^τs with an ℓ_2 penalty (corresponding to ridge regression). For the coefficient function $\beta^\tau(\cdot)$ we use basis representations of the form

$$\beta^\tau(s) \approx \sum_{d=1}^{D} b_d^\tau \varphi_d(s) \tag{3.3.1}$$

and present two strategies for this approximation: one in terms of penalized splines and one using eigenfunctions from the eigendecomposition of the covariance operator of the functional covariates. Thereafter, we extend the model and the estimation approaches to the more complex situation with longitudinal data. For all approximations the quantile regression model involves terms of the form

$$\int_S \beta^\tau(s) X_{ij}(s)ds \approx \sum_{d=1}^{D} b_d^\tau \int_S \varphi_d(s) X_{ij}(s)ds.$$

If $X_{ij}(\cdot)$s are observed at a dense grid the integrals are well approximated by Riemann sums or by quadrature rules.

3.3.1 Representing the functional coefficient with a pre-specified basis

Our first proposal is to use a spline representation for the functional coefficient $\beta^\tau(\cdot)$. We therefore let $\{\varphi_d\}_{d=1}^D$ in (3.3.1) denote a predefined spline basis, and introduce the notation $Z_{d,ij} = \int_S \varphi_d(s) X_{ij}(s)\, ds$. Then the expression (3.2.1) for the quantile is

$$Q^{\text{spline}}_{Y_{ij}|Z_{ij},u_i^\tau}(\tau) = u_i^\tau + \alpha^\tau + \sum_{d=1}^{D} b_d^\tau Z_{d,ij} \tag{3.3.2}$$

Hereby, the infinite-dimensional estimation problem has been turned into a finite-dimensional estimation problem with the coefficients $b_1^\tau, \ldots, b_D^\tau$ as the unknown parameters. The number of basis functions used in the approximation should be large

enough to guarantee a proper approximation of $\beta^\tau(\cdot)$, but a large D could lead to overfitting. In order to handle this problem, coefficients $b_1^\tau, \ldots, b_D^\tau$ are penalized as is common in functional regression (Marx and Eilers, 1999; Cardot et al., 1999b; Goldsmith et al., 2011b), see details below.

In standard quantile regression, parameters are estimated by minimizing an objective function defined as an empirical loss $\sum_{i,j} l_\tau(Y_{ij} - Q_{ij}^\tau)$ where Q_{ij}^τ is short for the level τ quantile for observation j of cluster i and depends on the model parameters, and l_τ is an appropriate loss function, typically the check function loss $v \mapsto v(\tau - \mathbb{1}_{(v<0)})$. We modify this approach in two ways, following Fasiolo et al. (2020). First, we use a smooth approximation of the check function loss, namely

$$l_{\tau,\lambda,\sigma}(v) = \frac{\tau - 1}{\sigma} v + \lambda \log\left(1 + \exp\left(\frac{v}{\sigma\lambda}\right)\right). \tag{3.3.3}$$

where λ determines the degree of smoothing, and the check function is recovered as $\lambda \to 0$. The loss function in (3.3.3) will be referred to as the Extended log-F (ELF) loss because $v \mapsto \exp(-l_{\tau,\lambda,\sigma}(v))$ is proportional to the density of an extended log-F distribution. Second, we penalize the subject-specific intercepts and the spline coefficients, and estimate the parameters by minimizing the penalized empirical loss

$$L^{\text{spline},\tau}(\alpha^\tau, \boldsymbol{b}^\tau, \mathbf{u}^\tau) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} l_{\tau,\lambda,\sigma}(Y_{ij} - Q^{\text{spline}}_{Y_{ij}|Z_{ij},u_i^\tau}(\tau)) + \frac{1}{2}\gamma_u ||\boldsymbol{u}^\tau||^2 + \frac{1}{2}\gamma_b ||\boldsymbol{b}^\tau||_B^2 \tag{3.3.4}$$

with $\boldsymbol{u}^\tau = (u_1^\tau, .., u_N^\tau)^T \in \mathbb{R}^N$, $\alpha^\tau \in \mathbb{R}$, $\boldsymbol{b}^\tau = (b_1^\tau, .., b_D^\tau)^T \in \mathbb{R}^D$ and penalty parameters $\gamma_u, \gamma_b > 0$. The penalty matrix B defining the norm $||\boldsymbol{b}^\tau||_B^2 = (\boldsymbol{b}^\tau)^T B \boldsymbol{b}^\tau$ is of size $D \times D$, positive semi-definite, and selected by the analyst. Importantly, and in contrast to the approach to be discussed in Section 3.3.2, the idea is to use a rich basis for $\beta^\tau(\cdot)$, i.e. a large D, and avoid overfitting by penalization of the coefficients.

We adopt the Bayesian approach for additive quantile regression proposed by Fasiolo et al. (2020). The method works with a belief-update principle (Bissiri et al., 2016). The ELF loss is differentiable and thus allows for common computational optimizers like the Newton method. In particular, this loss function is proportional to the negative log-density of the Gibbs posterior from the belief-update framework when imposing priors $\boldsymbol{u}^\tau \sim N(\mathbf{0}, \gamma_u^{-1}\mathbb{I}_N)$, where $\mathbb{I}_N$ is the $N \times N$ identity matrix, and $\boldsymbol{b}^\tau \sim N(\mathbf{0}, (\gamma_b B)^-)$, whose covariance matrix is an appropriate generalized inverse of $\gamma_b B$. Hence, the estimated coefficients $\hat{\boldsymbol{u}}^\tau$, $\hat{\alpha}^\tau$, and $\hat{\boldsymbol{b}}^\tau$ obtained by minimizing (3.3.4) correspond to the maximum a posteriori estimates. Notice that, apart from the penalty parameters γ_u and γ_b, the function (3.3.4) depends on two additional tuning parameters: $\sigma > 0$ and $\lambda > 0$ are the inverse of the learning rate and the smoothing level of the ELF loss function, respectively. Selection of the tuning parameters is based on a marginal loss criterion with integration over β^τ (for γ_u and γ_b), calibration using a Bayesian sandwich covariance estimator (for σ), and minimization of an asymptotic mean squared error of the estimated quantile regression coefficients (for λ). The selection procedures are implemented as part of the R package `qgam` accompanying Fasiolo et al. (2020).

3.3.2 Representing the functional coefficient with a data-driven basis

A common alternative to using a fixed, pre-specified basis for $\beta^\tau(\cdot)$ is to use a data-driven basis designed to capture the primary modes of variation for the covariates. Functional principal component analysis (FPCA) provides an algorithm to obtain such a basis, see for example Ramsay and Silverman (2005).

Consider a distribution on $L^2(S)$. The eigendecomposition of the covariance operator provides eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$ satisfying $\lambda_{k-1} \geq \lambda_k$ and $\lambda_k \geq 0$ for all k, and orthonormal eigenfunctions $\{\phi_k(\cdot)\}_{k=1}^{\infty}$. Any random function from the distribution can be reproduced as $X(s) = \mu(s) + \sum_{k=1}^{\infty} \xi_k \phi_k(s)$ where $\mu(s) = \mathrm{E}X(s)$ is the pointwise mean function and $\xi_k = \int_S \big(X(s) - \mu(s)\big)\phi_k(s)\, ds$. The eigenfunctions, usually referred to as principal components, describe the main directions of variation, and the eigenvalue λ_k quantifies the fraction of variance explained by ϕ_k. The coefficients ξ_k are called principal component scores.

In practice the mean function, eigenvalues, and eigenfunctions must be estimated from the data, in our case from $\{X_{ij}(s_h)\}_{ijh}$. We borrow methods from standard FPCA for independent data despite the cluster/longitudinal structure (but correlation within clusters/subjects will partly carry over as correlation among scores). The same approach was used by Goldsmith et al. (2012) for mean regression for longitudinal functional data and is not inappropriate since regression is carried out conditionally on the covariate functions. As an alternative, implementations of FPCA specially targeted at multilevel and longitudinal data exist (Di et al., 2009; Greven et al., 2010; Park and Staicu, 2015). With slight abuse of notation, we leave out "hats" from the notation even though the objects are estimated rather than known. The scores, denoted $\xi_{ij,k}$, are computed by numerical integration (when observations are dense as we assume), and moreover truncated representations $\widehat{X}_{ij}^K(s) = \mu(s) + \sum_{k=1}^{K} \xi_{ij,k}\phi_k(s)$ with only the first K terms are used. We consider the value of K fixed for now, but discuss the selection of it below.

When K is large enough, then the major part of the variation of the $X_{ij}(\cdot)$s is also present in the truncated representations $\widehat{X}_{ij}^K(\cdot)$, and since $\int_S \widehat{X}_{ij}^K(s)\tilde{\beta}(s)\, ds = 0$ for $\tilde{\beta}(\cdot)$ in the orthogonal complement of $\Phi = \mathrm{span}(\{\phi_1, \ldots, \phi_K\})$, it is natural to consider representations of $\beta^\tau(\cdot)$ that belong to Φ. Therefore, as alternative to the spline representation, our second proposal is to consider $\beta^\tau(s) = \sum_{k=1}^{K} c_k^\tau \phi_k(s)$ for unknown coefficients $c_1^\tau, \ldots, c_K^\tau$. Due to orthonormality of $\phi_1, \ldots, \phi_K$, the quantile in (3.2.1) is now approximated by

$$Q^{\mathrm{fpca},K}_{Y_{ij}|X_{ij},u_i^\tau}(\tau) = u_i^\tau + \gamma^\tau + \sum_{k=1}^{K} c_k^\tau \xi_{ij,k}. \tag{3.3.5}$$

It is natural here to think of the intercept γ^τ as an approximation to $\alpha^\tau + \int_S \beta^\tau(s)\mu(s)\, ds$ such that it incorporates the mean function, and an estimate of the original intercept parameter is computed from $\hat{\gamma}^\tau$ and $\hat{\beta}^\tau(\cdot)$.

As for the spline approach, we adopt the estimation procedure of Fasiolo et al. (2020) and minimize the ELF loss function with ℓ_2 penalty on random effects:

$$L^{\mathrm{fpca},\tau}(\gamma^\tau, \mathbf{c}^\tau, \mathbf{u}^\tau \mid K) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} l_{\tau,\lambda,\sigma}(Y_{ij} - Q^{\mathrm{fpca},K}_{Y_{ij}|X_{ij},u_i^\tau}(\tau)) + \frac{1}{2}\gamma_u ||\boldsymbol{u}^\tau||^2 \tag{3.3.6}$$

where $\boldsymbol{c}^\tau = (c_1^\tau, .., c_K^\tau)^T$. Expressions (3.3.2) and (3.3.5) for the quantiles are equivalent, except for the scalar values $Z_{ij,d}$ and $\xi_{ij,k}$, constructed using spline basis functions and eigenfunctions, respectively (and the notation for the unknown parameters). The objective functions (3.3.4) and (3.3.6) are also similar but differ in an important way: there is no penalty on $c_1^\tau, \ldots, c_K^\tau$ in (3.3.6). This is because regularization is carried out through the choice of K; a smaller K implies a more smooth estimate of $\beta^\tau(\cdot)$.

We now turn to selection of K. In FPCA, the percentage of variance explained (PVE), as determined by the eigenvalues, is often used as the criterion. More specifically,

$$K^{\text{PVE}} = \min\left\{K \geq 1 : \frac{\sum_{k=1}^{K}\lambda_k}{\sum_{k=1}^{\infty}\lambda_k} \geq p\right\},$$

where p is for example 0.95, 0.99 or even larger if only a small degree of smoothing is wanted. The PVE criterion is based on the functional covariates only. However, as a prevention towards poor predictive performance caused by overfitting, we consider the interplay with the outcome. One possible approach for model based selection of K is to use information criteria. Kato (2012) studied and compared Akaike's and the Bayesian information criterion (AIC and BIC) and the generalized approximate cross-validation criterion (GACV) from Yuan (2006) for quantile functional regression for independent data. He found that BIC was the most stable. Lee et al. (2014) studied model selection for quantile regression for high-dimensional data (many scalar covariates) and demonstrated that an adjusted version of the BIC was more appropriate; hence we decided to adopt it in our algorithm.

For speed of computation we carry out the BIC comparison in a marginal model, i.e., without cluster-specific effects. More specifically, for a fixed K the FPCA based marginal model approximation is given by

$$mQ^{\text{fpca},K}_{Y_{ij}|X_{ij}}(\tau) = \tilde{\gamma}^{\tau} + \sum_{k=1}^{K}\tilde{c}_k^{\tau}\xi_{ij,k}.$$

Notice, that the interpretation of coefficients $(\tilde{\gamma}^{\tau}, \tilde{c}_1^{\tau}, \ldots, \tilde{c}_K^{\tau})$ is different from the interpretation of $(\gamma^{\tau}, c_1^{\tau}, \ldots, c_K^{\tau})$, cf. Section 3.2.2. We use an objective function given by

$$L^{\text{marg},\tau}(\tilde{\gamma}^{\tau}, \tilde{\mathbf{c}}^{\tau} \mid K) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} l_{\tau,\lambda,\sigma}(Y_{ij} - mQ^{\text{fpca},K}_{Y_{ij}|X_{ij}}(\tau))$$

where $\tilde{\mathbf{c}}^{\tau} = (\tilde{c}_1^{\tau}, \ldots, \tilde{c}_K^{\tau})$, which does not include cluster-specific parameters and is therefore much faster to compute (and minimize) than $L^{\text{fpca}}(\tilde{\gamma}^{\tau}, \tilde{\mathbf{c}}^{\tau}, \mathbf{u}^{\tau} \mid K)$.

Recall that the objective function is constructed from ELF loss, so there is a (pseudo) log-likelihood value associated to the minimizers of $L^{\text{marg}}(\cdot,\cdot|K)$. We denote this value $LL^{\text{marg}}(K)$, and base our BIC criteria on this value. Without correction for high-dimensional data, the BIC value is defined as

$$\text{BIC}^{\tau}(K) = -2 \cdot LL^{\text{marg}}(K) + (K+1)\log(M)$$

where $M = \sum_{i=1}^{N} n_i$ is the total number of observations and $K+1$ is the number of parameters (excluding tuning parameters λ and σ), while the adjusted version of BIC from Lee et al. (2014) amounts to

$$\text{BIC}^{\text{adj},\tau}(K) = -2 \cdot LL^{\text{marg}}(K) + (K+1)\log(K+1)\log(M).$$

The adjusted version, with the logarithm of the number of covariates multiplied to the usual penalization term, was demonstrated to give good results in the simulation studies and data analysis in Lee et al. (2014). Notice that they used the penalization term in combination with the log-likelihood from an ALD working model, whereas we use it together with the ELF log-likelihood.

In practice we use the median for selection of K, i.e. minimize $\text{BIC}^{\text{adj},0.5}$, and we suggest to minimize over the set $\{2, 3, \ldots, K^{\text{PVE}}\}$. Using the median (instead of the level of interest, τ) makes the evaluation more robust, and using a model without cluster-specific parameters makes the selection faster since there is a considerable computational cost of including the many (penalized) extra parameters. We emphasize that these simplifications are used only in the preliminary step for selection of an appropriate number of scores to include in the model. As soon as K is selected we fit the model with penalized cluster-specific effects at the quantile level of interest. The complete estimation procedure is described in Algorithm 2.

Consider data $\{(Y_{ij}, X_{ij}(s_h))\}_{ijh}$;
Perform FPCA and select K^{PVE} according to the PVE criterion for a prespecified p;
forall $K = 2 : K^{\text{PVE}}$ **do**
  Take the first K scores for each observation: $\xi_{ij,1}, \ldots, \xi_{ij,K}$;
  Compute $\text{BIC}^{\text{adj},0.5}(K)$;
end
Set $\hat{K} = \arg\min_K \text{BIC}^{\text{adj},0.5}(K)$;
Minimize $L^{\text{fpca},\tau}(\gamma^\tau, \mathbf{c}^\tau, \mathbf{u}^\tau \mid \hat{K})$ from (3.3.6), and get $(\hat{\gamma}^\tau, \hat{\mathbf{b}}^\tau, \hat{\mathbf{u}}^\tau)$.

Algorithm 2: Pseudo code for implementation of model selection with the adjusted BIC criterion when the the functional coefficient is approximated with eigenfunctions.

3.3.3 Quantile regression with time-varying coefficients

The approaches for clustered data presented in Sections 3.3.1 and 3.3.2 are now extended to longitudinal data, where $X_{ij}(\cdot)$ is assumed to correspond to a time point t_{ij} varying in a time range $T \subset [0, \infty)$. Hence, data consist of $\{(Y_{ij}, X_{ij}(s_h), t_{ij})\}_{ijh}$.

It is natural to consider time-varying coefficients, i.e. allow for both the intercept and the coefficient function to change with longitudinal time, and consider

$$Q_{Y_{ij}|t_{ij},X_{ij},u_i^\tau}(\tau) = u_i^\tau + \alpha^\tau(t_{ij}) + \int_S \beta^\tau(s, t_{ij}) X_{ij}(s) ds \tag{3.3.7}$$

as an extension of (3.2.1). We assume that $t \mapsto \alpha^\tau(t)$ and $(s, t) \mapsto \beta^\tau(s, t)$ are smooth functions, and therefore use tools from additive models (Wood, 2017). More specifically, we approximate the smooth intercept as $\alpha^\tau(t) \approx \sum_{l=1}^L a_l^\tau \psi_l(t)$, where $\psi_1, \ldots, \psi_L$ are L basis functions of choice. In practice, we use cubic splines. Similarly to Section 3.3.1, we add an extra term to the loss function to address the penalization of coefficients $a_1^\tau, .., a_L^\tau$.

For the coefficient function $\beta^\tau(\cdot, \cdot)$, we can go in either of two direction as in Sections 3.3.1 and 3.3.2 and represent it with penalized splines or eigenfunctions. In the penalized splines approach we model $\beta^\tau(\cdot, \cdot)$ with a tensor product smooth; it is often preferred over a simple multivariate smooth when coordinates have rather different scales. To be specific, we choose separate bases for the s-direction and the t-direction, and then consider

$$\beta^\tau(s, t) \approx \sum_{l=1}^{L} \sum_{d=1}^{D} \delta_{dl}^\tau \psi_l(t) \varphi_d(s).$$

Notice that we use the same basis for $\alpha^\tau(\cdot)$ and the t-direction in $\beta^\tau(\cdot,\cdot)$ although not strictly necessary. The extended version of (3.3.2) becomes

$$Q^{\text{spline},\tau}_{Y_{ij}|t_{ij},X_{ij},u_i^\tau}(\tau) = u_i^\tau + \sum_{l=1}^{L} a_l^\tau \psi_l(t_{ij}) + \sum_{l=1}^{L}\sum_{d=1}^{D} \delta_{dl}^\tau \psi_l(t_{ij}) Z_{d,ij}(s) \tag{3.3.8}$$

where $Z_{d,ij}(s) = \int_S \varphi_d(s) X_{ij}(s)\, ds$ as previously. A penalty term for the scalar coefficients δ_{dl}^τ is added to the loss function, accounting for a tradeoff of wiggliness in the two directions, see Wood (2017, Chapter 5) for details about tensor product smooths.

For the FPCA approach, recall that eigenfunctions $\phi_k(\cdot)$ and scores $\xi_{ij,k}$ are available, and $X_{ij}(s)$ is approximated by $\widehat{X}_{ij}^K(s) = \mu(s) + \sum_{k=1}^K \xi_{ij,k}\phi_k(s)$. For fixed K we use the eigenfunctions in the s-direction in a tensor smooth construction for $\beta^\tau(\cdot,\cdot)$, i.e. $\beta^\tau(s,t) \approx \sum_{l=1}^L \sum_{k=1}^K \theta_{kl}^\tau \psi_l(t)\phi_k(s)$. Then, the approximation of (3.3.7) becomes

$$Q^{\text{fpca},\tau}_{Y_{ij}|t_{ij},X_{ij},u_i^\tau}(\tau) = u_i^\tau + \sum_{l=1}^{L} \tilde{a}_l^\tau \psi_l(t_{ij}) + \sum_{k=1}^{K}\sum_{l=1}^{L} \theta_{kl}^\tau \psi_l(t_{ij}) \xi_{ij,k}. \tag{3.3.9}$$

In terms of the parametrisation used in (3.3.7) the intercept term should be interpreted as $\sum_{l=1}^L \tilde{a}_l^\tau \psi_l(t) = \sum_{l=1}^L a_l^\tau \psi_l(t) + \int_S \beta(s,t)\mu(s)\, ds$ and thus incorporates the mean $\mu(\cdot)$ of the covariate functions. The intercept function $\alpha^\tau(\cdot)$ can be estimated from estimates of $\tilde{a}_l^\tau$s and $\beta^\tau(\cdot,\cdot)$ in a straightforward way. The BIC or BIC^{adj} criterion for the associated marginal model is used for selection of K, with the modification that ELF loss and likelihood now includes penalty terms for $\{a_l^\tau\}_l$ and $\{\theta_{kl}^\tau\}_{kl}$ and that the sum of effective degrees of freedom associated to $\{a_l^\tau\}_l$ and $\{\theta_{kl}^\tau\}_{kl}$ is used instead of $K+1$ as in the simpler case.

3.3.4 Covariates observed with noise

In the above sections we assumed that the covariate functions were observed without measurement noise, but this happens rarely in practice. Therefore, assume instead that we observe

$$W_{ij,h} = X_{ij}(s_h) + \epsilon_{ij,h} \qquad h = 1,\dots,H, \tag{3.3.10}$$

where $\{\epsilon_{ij,h}\}_{ijh}$ are iid. random variables with mean zero, and mutually independent of the underlying functions. We propose to carry out a preliminary smoothing step and proceed with the analysis with the unobserved values $X_{ij}(s_h)$ replaced by their fitted/predicted values. There are many smoothing techniques available for functional data, e.g. kernel-based methods, smoothing splines, and smoothing with data-driven bases, see for example Ramsay and Silverman (2005).

We use FPCA and thus the truncated representation $\widehat{X}_{ij}^K(s_h) = \mu(s) + \sum_{k=1}^K \xi_{ij,k}\phi_k(s_h)$ where K is selected with the PVE criterion as the prediction of $X_{ij}(s_h)$, see Section 3.3.2. Estimation of the principal components and scores requires extra attention when the true underlying function values $X_{ij}(s_h)$ are not available. There is large variety of FPCA implementations devoted to different sampling patterns of the functional data (dense or sparse, same or different sampling locations, missing values), see e.g. Yao et al. (2003) and Xiao et al. (2018), and references therein. We use the fast covariance estimation (FACE) method from Xiao et al. (2016) in this work. It is based on a sandwich estimator for the covariance function of the true underlying functions and smoother matrices constructed by penalized splines and is particularly useful for very dense observation due to its efficiency. It allows for missing values which is relevant for our application, but ignores potential

dependence among functions (see the references in Section 3.3.2 for alternatives in this direction).

3.3.5 Bootstrap procedures for variance assessment and bias adjustment

In the application we are mainly interested in estimation and inference for quantiles and differences between quantiles in certain directions of the functional covariate. It is known from the literature on quantile regression for longitudinal data with scalar covariates, that estimators may be biased and that it is difficult to properly assess the sampling variability of the estimators without resampling methods (Kato et al., 2012; Galvao and Montes-Rojas, 2015; Battagliola et al., 2021). Therefore, it comes as no surprise that we experience the same problems in our simulation experiments for the more complicated framework with functional covariates. We propose to use bootstrap strategies for variance estimation and bias adjustment.

Recall the quantile model (3.3.7) with repeated measurements of functional covariates and responses for each subject. To be specific about the targets, consider a fixed time point t and a function $X(\cdot) \in L^2(S)$ and the corresponding linear predictor $Q^{\tau,0} = \alpha^\tau(t) + \int \beta(s,t)X(s)\,ds$. Notice that the linear predictor is computed without random effect and is therefore interpreted as the τth quantile for an average subject (with $u^\tau = 0$). The function $X(\cdot)$ may or may not be one of the functions in the dataset. Furthermore, consider two functional covariates $X_A(\cdot), X_B(\cdot) \in L^2(S)$ with pointwise difference, $\Delta X(s) = X_A(s) - X_B(s)$. For a fixed cluster, i.e. a fixed u^τ and a fixed measurement time t, the corresponding difference in the τth quantile is

$$\Delta Q^\tau = Q_{Y|X_A,t,u^\tau}(\tau) - Q_{Y|X_B,t,u^\tau}(\tau) = \int_S \beta^\tau(s,t)\Delta X(s)\,ds, \tag{3.3.11}$$

so ΔQ^τ is the difference in quantile for a fixed subject when $X(\cdot)$ is changed in direction $\Delta X(\cdot)$. In the following we talk about $Q^{\tau,0}$ or ΔQ^τ as targets T of interest and let $\hat{T}$ denote the corresponding estimate with $\hat{\beta}^\tau(\cdot,\cdot)$ and $\hat{\alpha}^\tau(\cdot)$ (in the case of $Q^{\tau,0}$) and inserted for $\beta^\tau(\cdot,\cdot)$ and $\alpha^\tau(\cdot)$.

The estimates of coefficients in our models, e.g. $\{\hat{a}_l^\tau\}_l$ and $\{\hat{\delta}_{dl}^\tau\}_{dl}$ in equation (3.3.8), are accompanied with a variance-covariance matrix which can be used for computation of a standard error for the estimated target $\hat{T}$. We refer to these standard errors as model-based standard errors. However, penalization of random effects is likely to cause underestimation of the true sampling variation. We follow the suggestion from Galvao and Montes-Rojas (2015) and use cross-sectional resampling (or block resampling), meaning that complete subject data are sampled with replacement. More specifically, $i_1^*, \ldots, i_N^*$ are sampled with replacement from $\{1, \ldots, N\}$, and the bootstrap dataset is $\{(Y_{i^*j}, X_{i^*j}(s_h), t_{i^*j})\}_{ijh}$. In this way, within-subject dependence is maintained. For a target of interest, T, we proceed as follows: Draw a bootstrap sample as just described, carry out estimation, and compute the estimated target. Repeat this B times, and compute the standard deviation $\mathrm{sd}_{\mathrm{boot}}(\hat{T}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^B (\hat{T}^b - \bar{T})^2}$ where $\hat{T}^b$ is the estimated target from iteration b, and $\bar{T} = \frac{1}{B}\sum_{b=1}^B \hat{T}^b$. The same method was used by Canay (2011) and Geraci and Bottai (2014).

As documented by Battagliola et al. (2021), bias can occur even for large samples, caused by a combination of the incidental parameter problem (the number of parameters increase with sample size, Neyman and Scott (1948); Lancaster (2000)), non-linearity

of quantiles, and penalization of the subject-specific intercepts. Unfortunately, cross-sectional resampling cannot be used for bias adjustment because the target parameter of interest is not computable under the bootstrap distribution. See also Karlsson (2009) who obtained little or no effect in an attempt to adjust for bias in a nonlinear quantile regression for longitudinal data. Instead we propose to use the technique developed by Battagliola et al. (2021).

The idea is to generate bootstrap dataset where $\hat{T}$ (the estimate obtained from the observed data) is the true value of T, such that the bias can be estimated from the bootstrap estimates. More specifically, a bootstrap dataset consists of $\{(Y_{ij}^*, X_{ij}(s_h), t_{ij})\}_{ijh}$ where

$$Y_{ij}^* = u_i^* + \hat{\alpha}(t_{ij}) + \int_S \hat{\beta}^\tau(s, t_{ij}) X_{ij}(s)\, ds + \varepsilon_{ij}^*. \tag{3.3.12}$$

The estimates $\hat{\alpha}^\tau(\cdot)$ and $\hat{\beta}^\tau(\cdot,\cdot)$ are those obtained from the observed data, and the integral is computed numerically. Notice that the values of $X_{ij}(s_h)$ and t_{ij} from the observed data are used unchanged. The subject-specific intercepts $u_1^*, \ldots, u_N^*$ are drawn with replacement from the estimates $\hat{u}_1, \ldots, \hat{u}_N$ obtained from the observed data, and the residual terms $\{\varepsilon_{ij}^*\}_{ij}$ are generated via wild bootstrap. This means that $\varepsilon_{ij}^* = w_{ij}|\varepsilon_{ij}|$, where $\varepsilon_{ij} = Y_{ij} - \hat{\alpha}(t_{ij}) - \int_S \hat{\beta}^\tau(s, t_{ij}) X_{ij}(s)\, ds - \hat{u}_i$ are residuals from the model, and w_{ij}s are drawn independently as

$$w_{ij} = \begin{cases} 2(1-\tau), & \text{with probability } 1-\tau \\ -2\tau, & \text{with probability } \tau \end{cases}.$$

Wild bootstrap was introduced by Wu (1986) and Liu (1988) for mean regression, and adapted to quantile regression by Feng et al. (2011). Results in Feng et al. (2011), Wang et al. (2018a) and Battagliola et al. (2021) indicate that wild bootstrap captures asymmetry and heteroskedasticity better than ordinary resampling of residuals. Data generated as in (3.3.12) satisfy equation (3.3.7), with parameters $\alpha^\tau(\cdot) = \hat{\alpha}^\tau(\cdot)$ and $\beta^\tau(\cdot,\cdot) = \hat{\beta}^\tau(\cdot,\cdot)$, and the true value (in the bootstrap data) of the target is therefore $\hat{T}$. Let $\tilde{T}_1, \ldots, \tilde{T}_B$ be estimated values of T for B bootstrap datasets, then bias is estimated as $\text{bias}(\hat{T}) = \frac{1}{B}\sum_{b=1}^B (\tilde{T}^b - \hat{T})$.

The two bootstrap sampling schemes differ in several ways. While the cross-sectional sampling methods is completely non-parametric, the wild bootstrap method relies on the model used for estimation. Another important difference is that the covariate functions are resampled (together with the responses) by the cross-sectional method whereas they are kept exactly as in the dataset for the wild bootstrap methods. As a consequence, the procedure based on wild bootstrap would underestimate the variance of the estimator $\hat{T}$. Our suggested solution is to combine the estimated bias and estimated standard deviation from the two bootstrap sampling methods, respectively, to construct confidence intervals for the target T. If the distribution of $\hat{T}$ is well approximated by a normal distribution, $\hat{T} \sim N(T + B_T, \sigma_T^2)$, then $\hat{T} - B_T \pm q_{1-\alpha/2}\sigma_T$, where $q_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $N(0,1)$, is an approximate $1-\alpha$ confidence interval for T. Estimating the bias and standard deviation as described above leads to the confidence interval

$$\hat{T} - \text{bias}(\hat{T}) \;\pm\; q_{1-\alpha/2}\; \text{sd}_{\text{boot}}(\hat{T}). \tag{3.3.13}$$

Battagliola et al. (2021) demonstrated in a wide variety of simulation settings with clustered data and scalar covariates that bias was removed or reduced with the above bootstrap sampling process combining resampled cluster-specific intercepts and wild bootstrap for the residuals, and Galvao and Montes-Rojas (2015) demonstrated that

sampling variation of estimators is measured appropriately with cross-sectional resampling. Nevertheless, we admit that the construction of confidence intervals in (3.3.13) is ad hoc and that its properties should be investigated theoretically or by simulation, but this is left for future research. In this paper we apply the techniques and study their effects in the application.

3.4 Implementation

We used the software environment `R` (R Core Team, 2020b) for the computations. The FACE method is implemented in the function `fpca.face`, which is part of package `refund` (Goldsmith et al., 2019). It can handle functional data observed on a dense or a sparse grid, and values can be missing. One specifies either the selected PVE (`pve`) or the number of principal components (`npc`) of choice. The resulting eigenfunctions, the functional mean, and the predicted/smoothed functions are evaluated and returned at a dense grid.

The method developed by Fasiolo et al. (2020) is implemented in the package `qgam`. It includes the `qgam` function, which is a wrapper of the function `gam` (Wood, 2017; Wood and Scheipl, 2020). The call to `qgam` has the following structure:

```
qgam(y ~ formula, qu=tau, data=data)
```

where `y` is the response and `formula` specifies any ordinary covariates, smooth effects, and random effects to include in the model. The quantile level of interest τ is passed to `qu`, and the entry `data` specifies the data frame of interest. To introduce more flexibility in the estimation, one can employ the following formulation of `qgam`:

```
qgam(list(y ~ formula, ~ formula), qu=tau, data=data)
```

which allows the learning rate σ to vary with the covariates.

For the framework presented in this paper, recall that M is the total number of observations in the dataset and H is the length of the dense grid of points over which functional covariates are observed. We consider the situation with measurement noise and let `Xhat` be the $M \times H$ matrix containing the predictions $\{\widehat{X}_{ij}^{K}(s_h)\}$ from FACE with $K = K^{\text{PVE}}$. If the values $X_{ij}(s_h)$ are observed without noise, then `Xhat` is replaced by the matrix with observed values in the following. Furthermore, let `sGrid` be the $M \times H$ matrix whose rows contain $(s_1, \ldots, s_H)$.

Consider first the situation from Sections 3.3.1 and 3.3.2 with a scalar intercept α^τ and a functional coefficient $\beta^\tau(\cdot)$. Minimization of the loss function (3.3.4), now with $X_{ij}(\cdot)$ replaced by $\widehat{X}_{ij}(\cdot)$, is carried out with

```
formula = s(sGrid, by=Xhat, bs='cr') + s(id, bs='re')
```

The option `bs='cr'` implies that a cubic spline basis is used for $\beta^\tau(\cdot)$, with a default of ten basis functions. The corresponding penalty matrix B penalizes the integrated squared second derivative, sometimes called "curvature", of $\hat{\beta}^\tau(\cdot)$. As default, if no smoothing basis `bs` is supplied, then `gam` and thus `qgam` use thin plate regression splines. Moreover, `gam` allows to choose cyclic cubic regression splines (`bs='cc'`), for example, when it is desirable for the smooth term to take the same values at the boundaries of its domain, and we apply this for the application. The cluster-specific intercepts u_i^τ are included in the model with the `s` function, too, in which we specify `bs='re'` (`'re'` for random effects) as well as `id`, the grouping level factor associated to clusters. We refer to Wood

(2017) for more details about possibilities with smooth terms in `gam`. When we use eigenfunctions as the basis for $\beta^\tau(\cdot)$, minimizing (3.3.6) with $K = 2$, for instance, then we use

```
formula = xi1 + xi2 + s(id, bs='re')
```

where `xi1` and `xi2` are vectors of length M containing the principal component scores $\{\xi_{ij,1}\}$ and $\{\xi_{ij,2}\}$.

With the introduction of longitudinal time as in (3.3.7), we need to include a smooth intercept in time as well as the bivariate functional coefficient. For the tensor smooth approximation in (3.3.8), we employ

```
formula = s(tgrid, bs='cr') + te(sGrid, tGrid, by=Xhat)
              + s(id, bs='re')
```

where `tgrid` is the vector of length M whose ijth entry is t_{ij}, and `tGrid` is the $M \times H$ matrix whose columns are copies of `tgrid`. By default, the chosen marginal basis functions for both directions are five penalized cubic regression splines (`bs="cr"`), but both the type and the size of the bases can be changed, and it is also possible to use different sets of basis functions in the s- and t-directions. For estimation of the coefficients $\{\theta^\tau_{kl}\}$ in equation (3.3.9), the syntax (for $K = 2$) is

```
formula = s(tgrid, bs='cr') + s(tgrid, by=xi1)
             + s(tgrid, by=xi2) + s(id, bs='re')
```

because the scores are multiplied onto the unknown coefficients.

The output of `qgam` is a `gamObject`, which stores several quantities related to the model and the estimation process, such as the twice the log-likelihood (`logLik`) and the estimated effective degrees of freedom (`edf2`). Both are used for the computation of $\text{BIC}^{\text{adj},0.5}(K)$ in the presence of smooth effects, i.e. for model (3.3.9), while only the log-likelihood is needed for the simpler model (3.3.5). The values are also used for computation of AIC values in the application. Moreover, `Vp`, the variance-covariance matrix of all estimated coefficients, is available, and can be used to compute model-based standard errors and confidence bands for functions of the parameters, such as the targets mentioned in Section 3.3.5. Predicted quantiles for new covariate functions can be computed with the function `predict.gam`, and the function allows to exclude one or more terms from the model, such as the subject-specific intercepts, in the prediction.

3.5 Simulations

In this section we are going to examine the performance of the estimation methods described in Section 3.3 by means of simulation studies. Firstly, we consider the case in which the functional coefficient solely depends on the functional coordinate $s \in S$. Afterwards, we consider longitudinal data where the functional coefficient also depends on longitudinal time.

We intended to compare our proposed methods to the estimation method proposed by Brockhaus et al. (2017), which allows to perform scalar-on-function quantile regression with computations based on boosting (Bühlmann and Hothorn, 2007; Schmid and Hothorn, 2008) and is available in the `R` package `FDboost`. However, despite many attempts with different options in the `R` functions we never managed to get reliable results, and we refrain from showing the results.

3.5.1 Data generation

We simulate data $\{(Y_{ij}, X_{ij}(\cdot), W_{ij,h}, u_i, t_{ij})\}_{ij}$, with $i = 1, \ldots, N$ denoting subject and $j = 1, \ldots, n$ denoting observation number within subject. For simplicity we use the same number of repeated measures for all subjects and furthermore let the longitudinal time stamps be equally spaced in $T = [0, 1]$ for all subjects.

Inspired by simulation studies in Goldsmith et al. (2012) and Kundu et al. (2016) the functional covariates are generated as

$$X_{ij}(s) = c_{0,ij} + c_{1,ij}\phi_1(s) + c_{2,ij}\phi_2(s) + c_{3,ij}\phi_3(s) + c_{4,ij}\phi_4(s) + c_{5,ij}s \tag{3.5.1}$$

where $\phi_1(s) = \sin(2\pi s)$, $\phi_2(s) = \cos(2\pi s)$, $\phi_3(s) = \sin(4\pi s)$, $\phi_4(s) = \cos(4\pi s)$, and $s \in S = (0, 1]$. We use coefficients $c_{0,ij}, c_{5,ij} \overset{iid}{\sim} N(3, 1)$, while $c_{1,ij} \overset{iid}{\sim} N(0, 2)$, $c_{2,ij} \overset{iid}{\sim} N(0, 1)$, $c_{3,ij} \overset{iid}{\sim} N(0, 0.5)$ and $c_{4,ij} \overset{iid}{\sim} N(0, 0.25)$. The covariate functions are observed with Gaussian noise on an equally-spaced dense grid of $H = 100$ points,

$$W_{ij,h} = X_{ij}(s_h) + \epsilon_{ij,h},$$

where $\epsilon_{ij,h} \overset{iid}{\sim} N(0, 0.25^2)$. The response Y_{ij} is constructed as

$$Y_{ij} = u_i + \alpha(t_{ij}) + \int_S X_{ij}(s)\beta(s, t_{ij})ds + \left(1 + \gamma \int_S X_{ij}(s)ds\right) e_{ij} \tag{3.5.2}$$

where $u_i \overset{iid}{\sim} N(0, \sigma_u^2)$ is a subject-specific intercept, $\alpha(t) = \log(5t + 1)$ is the smooth intercept with respect to the longitudinal time $t_{ij} \in T = [0, 1]$, $\gamma \geq 0$ is a heteroskedasticity parameter, and $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$, where $\sigma_e > 0$. The coefficient function $\beta(\cdot, \cdot)$ is specified later. Importantly, the response is generated by means of the true underlying $X(\cdot)$ rather than the observations $W_{ij,h}$, so a preliminary smoothing step is carried out as described in Section 3.3.

For $\tau \in (0, 1)$ the implied quantile model is

$$\begin{aligned} Q_{Y_{ij}|X_{ij},t_{ij},u_i}(\tau) &= u_i + \alpha(t_{ij}) + \sigma_e\Phi^{-1}(\tau) + \int_S \left(\beta(s, t_{ij}) + \gamma\sigma_e\Phi^{-1}(\tau)\right) X_{ij}(s)ds \\ &= u_i + \alpha^\tau(t_{ij}) + \int_S \beta^\tau(s, t_{ij})X_{ij}(s)ds \end{aligned} \tag{3.5.3}$$

where Φ is the CDF of the standard Gaussian distribution, and the definitions of the functions $\alpha^\tau(\cdot)$ and $\beta^\tau(\cdot, \cdot)$ appear from the formula. In particular, the expression for the quantile in (3.5.3) has the same form as in (3.3.7). Notice that $\alpha^\tau(\cdot)$ differs from $\alpha(\cdot)$ unless $\tau = 0.5$, and that $\beta^\tau(\cdot, \cdot)$ differs from $\beta(\cdot, \cdot)$ unless $\tau = 0.5$ and/or $\gamma = 0$. The random intercepts u_i are independent of τ.

We study four scenarios in most detail; here $N = 200$, $n_i = n = 10$, $\sigma_e^2 = 0.5$, $\sigma_u/\sigma_e = 1.5$, either $\gamma = 0$ (the homoskedastic case) or $\gamma = 0.5$ (the heteroskedastic case), and the quantile level of interest is either $\tau = 0.1$ or $\tau = 0.5$. The combinations of τ and γ give four scenarios: A ($\tau = 0.5$, $\gamma = 0$), B ($\tau = 0.5$, $\gamma = 0.5$), C ($\tau = 0.1$, $\gamma = 0$), and D ($\tau = 0.1$, $\gamma = 0.5$). Notice that $\alpha^\tau(\cdot) = \alpha(\cdot)$ in scenarios A and B, and $\beta^\tau(\cdot, \cdot) = \beta(\cdot, \cdot)$ in scenarios A–C. In the literature is it common to focus on homoskedastic data and/or median regression, but we will we pay special attention to scenario D since it is the most difficult one. Scenarios E–J are modifications of scenario D, all with $\gamma = 0.5$ and $\tau = 0.1$: see the overview in Table 3.1. For each scenario we consider 200 replications.

	τ	γ	N	n	σ_e^2	σ_u/σ_e	e_{ij}
A	0.5*	0*	200	10	0.5	1.5	Normal
B	0.5*	0.5	200	10	0.5	1.5	Normal
C	0.1	0*	200	10	0.5	1.5	Normal
D	0.1	0.5	200	10	0.5	1.5	Normal
E	0.1	0.5	300*	10	0.5	1.5	Normal
F	0.1	0.5	200	15*	0.5	1.5	Normal
G	0.1	0.5	200	10	0.5	1.5	ALD*
H	0.1	0.5	200	10	0.5	1.5	t_3^*
I	0.1	0.5	200	10	0.5	0*	Normal
J	0.1	0.5	200	10	0.5	2*	Normal

Table 3.1: Description of quantile level and parameter values used in the simulation scenarios. Asterisks (*) denote values or error distributions where the scenario differs from scenario D.

3.5.2 Time-invariant regression coefficient

We first analyze the case in which data is generated with $\beta(s,t) = \beta(s)$ in (3.5.2). Specifically, we will adopt two possible functional coefficients, namely $\beta_1(s) = \sqrt{2}\cos(2\pi s)$ and $\beta_2(s) = s$.

As a first step, we smooth the noisy observations $\{W_{ij,h}\}$ with FACE at a PVE of 0.9999 in order to get $\widehat{X}_{ij}(\cdot)$ as described in Section 3.3.2. We use this large PVE in order to avoid oversmoothing and to get relatively many terms that can potentially be included in the quantile regression driven by eigenfunctions. Figure 3.1 shows 50 realizations of observed curves and the corresponding estimated functions from a dataset generated from the benchmark scenario. Despite the large PVE, the random noise is removed because it is not common to the curves.

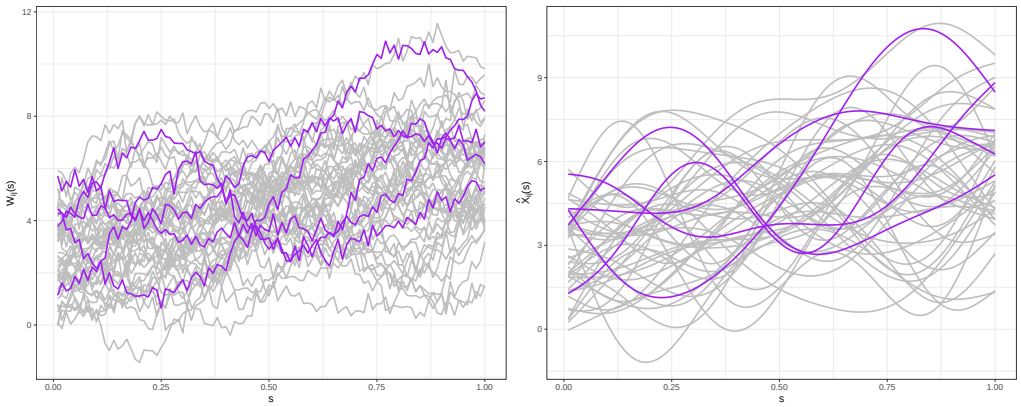

Figure 3.1: Fifty observed (left panel) and smoothed (right panel) functional covariates from a realization of functional data in our simulation setup. The curves are in grey and five of them are in purple to better show the features of the simulated functions.

Comparison of spline and eigenfunction approaches

We start a comparison of the results from the spline methods and the eigenfunction method. We use ten cubic regression splines for $\alpha^\tau(\cdot)$. Furthermore, we employ ten cubic regression splines for $\beta^\tau(\cdot)$ for the spline approach, while for the eigenfunction approach

we compare values of $\text{BIC}^{\text{adj},0.5}(K)$ for the number of included eigenfunctions, K, ranging from $K_1 = 2$ to $K_2 = K^{\text{PVE}}$. We also used the BIC criterion without adjustment for high-dimensional data and got similar results for predicted quantiles and the smooth intercept. However, a much larger K was usually selected resulting in severe overfitting of $\beta^\tau(\cdot)$. Furthermore, we experimented with variations of the BIC criterion for selection of K, relying on the asymmetric Laplace distribution as in Kato (2012), both with the classical formulation of BIC as well as its adjusted version for high dimensions (Lee et al., 2014). They both gave very similar results to those using $\text{BIC}^{\text{adj},0.5}$ so we will not discuss this any further.

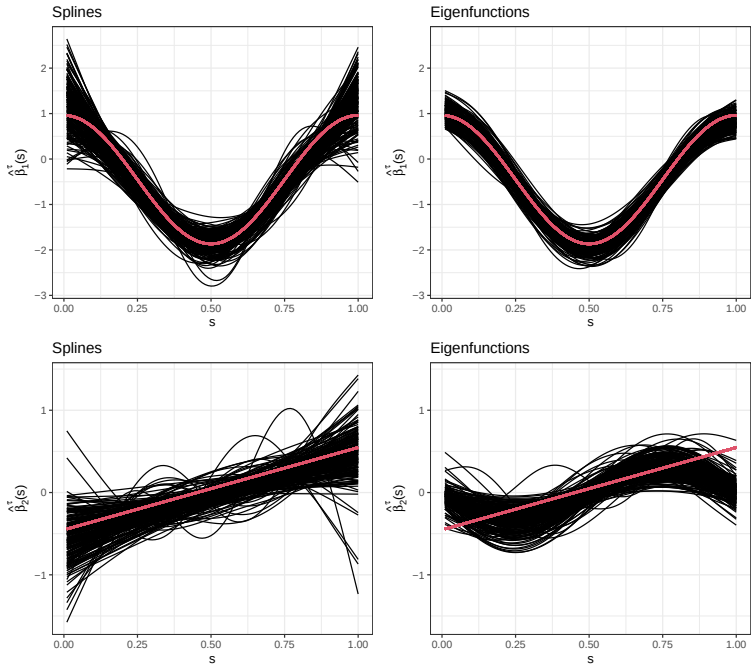

Figure 3.2: Estimated functional coefficients, $\hat{\beta}^\tau(\cdot)$, in scenario D when data are simulated with $\beta(s) = \beta_1(s)$ (top row) and $\beta(s) = \beta_2(s)$ (bottom row). We show the results for the approximation with splines (left column) and with eigenfunctions (right column). The red curves represent the true coefficient functions, $\beta^\tau(\cdot)$.

We start with a comparison of the spline and the eigenfunction approach for the estimated coefficient function in scenario D. Figure 3.2 shows the estimated coefficients (in black) and the true coefficients (in red) for both choices of $\beta(s)$. For $\beta_1(s)$, the estimates from the spline representation have higher variation than the ones from the representation by means of eigenfunctions, especially close to the borders of the functional domain. The spline estimates do not bend off like the true function, and this generates a bias for s around 0.15 and 0.85. The eigenfunction approximation captures the true shape of the coefficient for every replication. This is because the eigendecomposition essentially reconstructs sines and cosines in our simulation set-up. The representation via eigenfunctions is less suitable when the true function is $\beta_2(s)$, while the spline approximation reproduces a straight line for most of the replications. This is likely because non-linearity is penalized in the spline approach.

Estimation of $\alpha^\tau(\cdot)$ and $\beta^\tau(\cdot)$ is summarized for the four scenarios in the first two

columns in Figure 3.3 for $\beta_1(s)$ (top) and $\beta_2(s)$ (bottom). For each replicate we compute the integrated squared errors, $\text{ISE}(\hat{\alpha}^\tau) = \int(\hat{\alpha}^\tau(t) - \alpha^\tau(t))^2\,dt$, and $\text{ISE}(\hat{\beta}^\tau) = \int(\hat{\beta}^\tau(s) - \beta^\tau(s))^2\,ds$, and the panels show boxplot over the 200 replicates. The differences from Figure 3.2 between the spline and the eigenfunction representation for $\hat{\beta}^\tau(\cdot)$ in scenario D are clearly recognized as $\text{ISE}(\hat{\beta}^\tau)$ differ between methods, and in opposite directions for $\beta_1(s)$ and $\beta_2(s)$. The same conclusions hold for the scenarios A–C, but with smaller $\text{ISE}(\hat{\beta}^\tau)$. Notice that the variation in $\text{ISE}(\hat{\beta}^\tau)$ is larger for the spline approximation than for the eigenfunction approximation in all eight cases. For $\hat{\alpha}^\tau(\cdot)$, there are hardly any differences between the two estimation approaches.

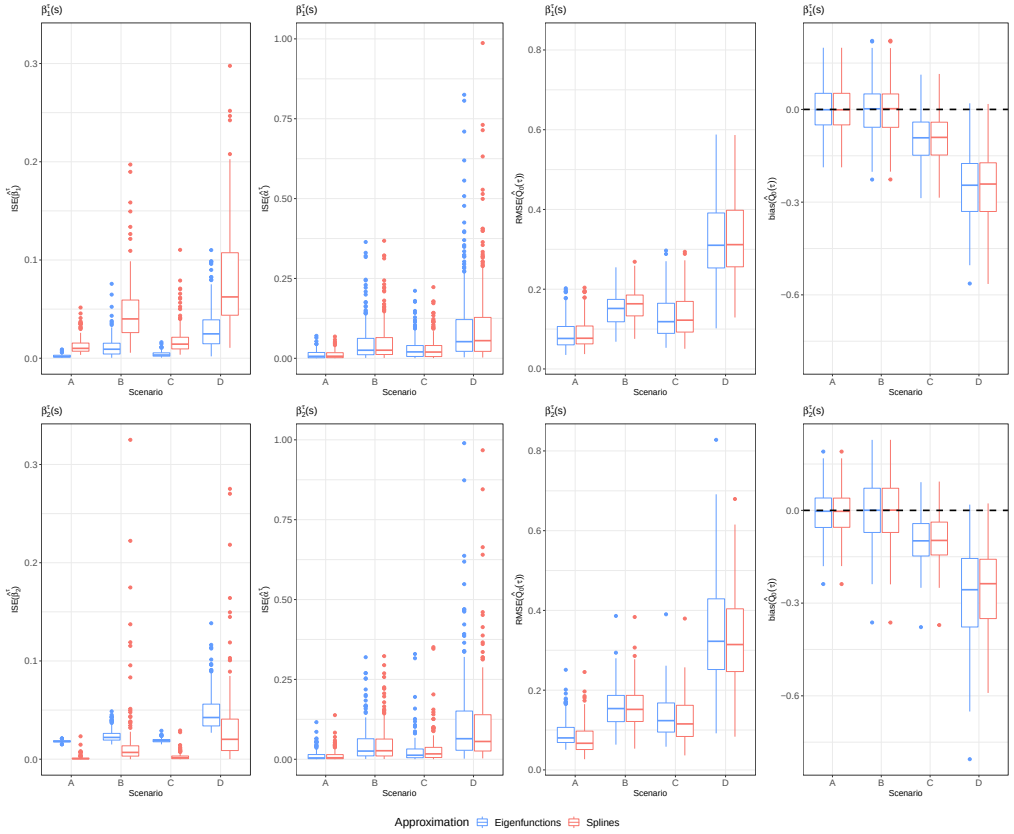

Figure 3.3: Comparison of results for the spline (blue) and the eigenfunction (red) approximation in four scenarios. Boxplots of the ISE of the estimated functional coefficient $\hat{\beta}^\tau(\cdot)$ and of the smooth intercept $\hat{\alpha}^\tau(\cdot)$ (first and second column, respectively) and boxplots of the RMSE and bias of the linear predictor $\hat{Q}^{\tau,0}$ (third and fourth column, respectively) when data are simulated with $\beta(s) = \beta_1(s)$ (first row) and $\beta(s) = \beta_2(s)$ (second row). See Table 3.1 for a detailed explanation of the scenarios.

All that being said about estimation of $\beta^\tau(\cdot)$, one should be careful to pay too much attention to differences in the estimated coefficient functions, since identifiability of coefficient functions is an issue in functional regression. The identifiable component is the integral $\int \beta^\tau(s)X(s)\,ds$ for X belonging to the space of observable functions, not necessarily $\beta^\tau(\cdot)$ itself. A fitted model may be good at reconstructing the integrals, and thus the quantiles, even if $\hat{\beta}^\tau(\cdot)$ does not reconstruct $\beta^\tau(\cdot)$ well.

We therefore turn our attention to the linear predictor $Q_{ij}^{\tau,0} = \alpha^\tau(t_{ij}) + \int \beta(s)X_{ij}(s)\,ds$ and the corresponding estimates $\hat{Q}_{ij}^{\tau,0} = \hat{\alpha}^\tau(t_{ij}) + \int \hat{\beta}(s)X_{ij}(s)\,ds$. Notice that quantiles are computed without random effects and is therefore interpreted as the τth quantile for an average subject (with $u_i^\tau = 0$). For each replicate we compute a root mean squared error, i.e. $\text{RMSE}(\hat{Q}^{\tau,0}) = \sqrt{\frac{1}{M}\sum_{i,j}(\hat{Q}_{ij}^{\tau,0} - Q_{ij}^{\tau,0})^2}$ where $M = nN$ is the total number of observations. Boxplots of the RMSE over the 200 simulated datasets are shown in the third column in Figure 3.3 for scenarios A–D, and for $\beta_1(s)$ (top) and $\beta_2(s)$ (bottom)., respectively. The last columns show boxplots over the average bias: $\text{bias}(\hat{Q}^{\tau,0}) = \frac{1}{M}\sum_{i,j}(\hat{Q}_{ij}^{\tau,0} - Q_{ij}^{\tau,0})$. The most striking observation is that the spline approach and the eigenfunction approach give extremely similar results when it comes to prediction of the linear predictors even if the estimated coefficients can be quite different as illustrated in Figure 3.2. Moreover, the complexity of the four scenarios is clearly reflected in the results: The linear predictor is estimated without bias when $\tau = 0.5$, but with non-negligible bias when $\tau = 0.1$, and the RMSE is larger when $\gamma = 0.5$ compared to $\gamma = 0$.

More details concerning estimation of the linear predictor is provided in Table 3.3 in the appendix for the spline approach for $\beta_1(s)$. In particular the average bias, average standard deviation and average RMSE (average over the 200 bootstrap replicates) are reported to give an indication of the contributions of mean and variation, respectively, to the RMSE. The table also reports the average RMSE of the linear predictor for out-of-sample prediction: For each of the 200 replicates, we simulated extra data (test data) consisting of 200 subjects with 10 repeated measures each, and compared the true and estimated linear predictor. The in-sample and out-of-sample RMSE do not differ. Finally, the table lists reports the RMSE for a marginal estimator and for more scenarios; these results are commented on below.

In summary, the two estimation methods behave similarly when it comes to estimation of the quantiles. The spline approach has larger flexibility for the functional form of the estimated coefficient function (but at the expense of larger variation) and is faster to run because the eigenfunction approach requires the preliminary step with selection of K. Therefore, we focus on spline approximations in the following. Moreover we focus on scenario D and variations of it since it is the most difficult and thus interesting case.

Effect of changing sample size or error distribution

We now study the effect of changing the number of subjects, the number of repeated measures, or the error distribution one at a time. Apart from those changes, the setting is as in scenario D; in particular we use $\tau = 0.1$ as the target quantile level. We use the coefficient function $\beta_1(s)$. Figure 3.4 is organized like each row in Figure 3.3 and shows the distributions of $\text{ISE}(\hat{\alpha}^\tau)$, $\text{ISE}(\hat{\beta}^\tau)$, $\text{RMSE}(\hat{Q}^{\tau,0})$, and $\text{bias}(\hat{Q}^{\tau,0})$.

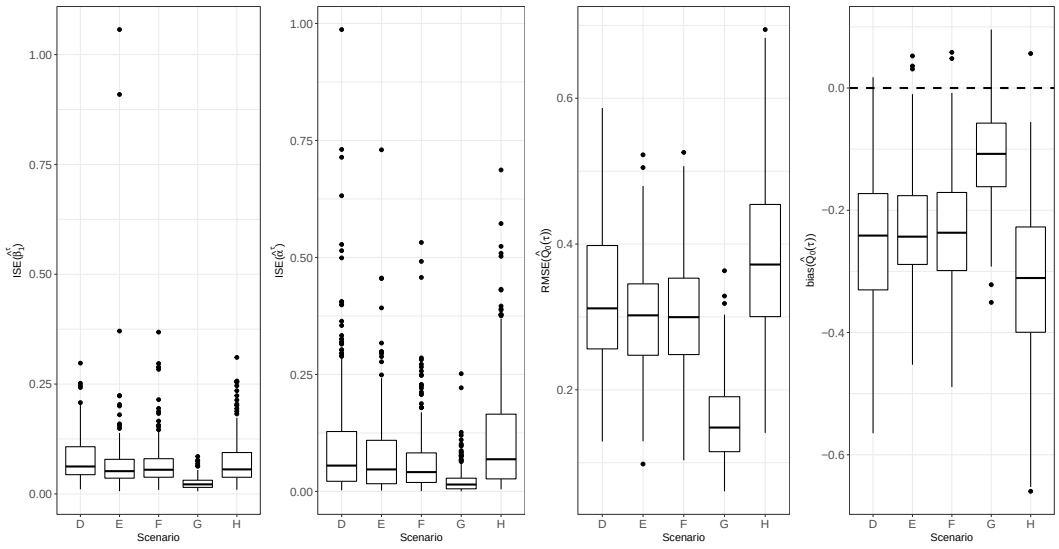

Figure 3.4: Boxplots of the ISE of the estimated functional coefficient $\hat{\beta}^\tau(\cdot)$ and of the smooth intercept $\hat{\alpha}^\tau(\cdot)$ (first and second column, respectively) and boxplots of the RMSE and bias of the linear predictor $\hat{Q}^{\tau,0}$ (third and fourth column, respectively) when data are simulated with $\beta(s) = \beta_1(s)$, and estimation is carried out with the spline approximation. See Table 3.1 for a detailed explanation of the scenarios.

For scenarios E and F, we increase the number of subjects from 200 to 300, and the number of repeated measurements for each subject from 10 to 15, respectively, so they share the number of total observations (3000, 50% more than in scenario D). Comparison of the first three boxplots in Figure 3.4 shows that this has little effect on the average level of the error terms, but the variation over datasets decreases slightly when the total sample size increases.

Next, we consider two more settings in which we either sample the residual terms e_{ij} in equation (3.5.2) independently from an ALD centred in 0 with skewness parameter $\tau = 0.1$ (equal to the target quantile) and scale parameter $\rho > 0$ (scenario G) or independently from a scaled Student t-distribution with three degrees of freedom (scenario H). In both cases we scale the distributions such that their variance is equal to 0.5 (σ_e^2 from the benchmark setting), such that relative level of variation between subject-specific effects and error terms is preserved. More specifically, for the ALD we imposed $\rho = \frac{\sigma_e(\tau(1-\tau))}{\sqrt{1-2\tau+2\tau^2}}$ (Yu and Zhang, 2005). Notice that the definitions of $\alpha^\tau(\cdot)$ and $\beta^\tau(\cdot)$ are also changed compared to scenario D with Gaussian residual terms.

The boxplots for scenario G in Figure 3.4 shows that estimation errors are smaller for ALD residuals compared to Gaussian errors. This is expected since the (unpenalized) criterion based on the check function corresponds to the log-likelihood function (with a minus) for data sampled from the ALD; therefore, our estimators based in the smooth ELF generalization of the check function can be considered approximate maximum likelihood estimates in case of ALD residuals. On the other hand, the boxplots for scenario H shows that estimation errors are larger when the error distribution is heavy-tailed. For more details, see Table 3.3 in the appendix.

Conditional vs. marginal quantile regression

We now illustrate the effect of ignoring the longitudinal structure in the data, i.e. of fitting a functional linear quantile regression without random effects. More specifically, as a supplement to the usual model, we also fit a linear marginal model (without subject-specific terms)

$$mQ^{\text{lin}}_{Y_{ij}|X_{ij},t_{ij}}(\tau) = \alpha^{\tau}_{m}(t_{ij}) + \int_{S} \beta^{\tau}_{m}(s,t_{ij})X_{ij}(s)ds \tag{3.5.4}$$

with unknown coefficients $\alpha^{\tau}_{m}(\cdot)$ and $\beta^{\tau}_{m}(\cdot,\cdot)$; subscript m is used to stress that these are coefficients in a marginal model for the quantiles. We use the same approach as for the conditional model, with the exception that there are no subject-specific parameters in the model (this, by the way, makes estimation much faster).

Recall the true association (3.2.2) between the covariate X_{ij} and the quantile in the marginal distribution. When $\tau = 0.5$ and/or $\sigma_u = 0$ then the marginal and the conditional models coincide, and we therefore expect the marginal and the conditional estimation approaches to give similar results. When $\gamma = 0$ then equation (3.5.4) has the correct functional form but $\alpha^{\tau}(\cdot)$ and $\alpha^{\tau}_{m}(\cdot)$ differ, so the two estimation methods have different targets. For other values of τ, σ_u, and γ, equation (3.5.4) does not even have the correct functional form. In other words, the model is misspecified, and there are no true values of $\alpha^{\tau}_{m}(\cdot)$ and $\beta^{\tau}_{m}(\cdot,\cdot)$.

Nevertheless, it is still possible to check the misspecified model's ability to estimate the quantiles and thus compare the estimated quantiles from the model fits corresponding to the conditional and the (misspecified) marginal model, respectively. More specifically, we compute the average bias, i.e. $\text{bias}(\hat{Q}^{\tau,0}_{m}) = \frac{1}{M}\sum_{i,j}(\hat{Q}^{\tau,0}_{ij,m} - Q^{\tau,0}_{ij})$, and the corresponding RMSE, i.e. $\text{RMSE}(\hat{Q}^{\tau,0}_{m}) = \sqrt{\frac{1}{M}\sum_{i,j}(\hat{Q}^{\tau,0}_{ij,m} - Q^{\tau,0}_{ij})^2}$, where $\hat{Q}^{\tau,0}_{ij,m} = \hat{\alpha}_m(t_{ij}) + \int \hat{\beta}_m(s)X_{ij}(s)\,ds$ using estimates from the marginal estimation.

Figure 3.5 shows boxplots over 200 replicates of the results for scenarios A–H as well as two extra scenarios, all with $\beta(s,t) = \beta(s) = \beta_1(s)$. The extra scenarios are identical to scenario D except that $\sigma_u = 0$ (scenario I) and $\sigma_u/\sigma_e = 2$ (scenario J). The average $\text{RMSE}(\hat{Q}^{\tau,0}_{m})$ is listed in Table 3.3 in the appendix. As expected the results are similar for the conditional and the marginal estimation approach when $\tau = 0$ (scenarios A and B) or $\sigma_u = 0$ (scenario I). In all other cases the estimates from the marginal approach have larger bias and RMSE compared to the estimates from the conditional approach. This is also to be expected since the conditional approach is targeted towards estimation of $Q^{\tau,0}_{ij} = \alpha^{\tau}(t_{ij}) + \int_S \beta^{\tau}(s)X_{ij}(s)\,ds$, whereas the marginal approach is targeted towards the quantiles in the marginal distribution. In summary, if one is interested in the quantiles $Q^{\tau,0}_{ij}$, then is is of great importance to take the within-subject dependence into account in the estimation procedure.

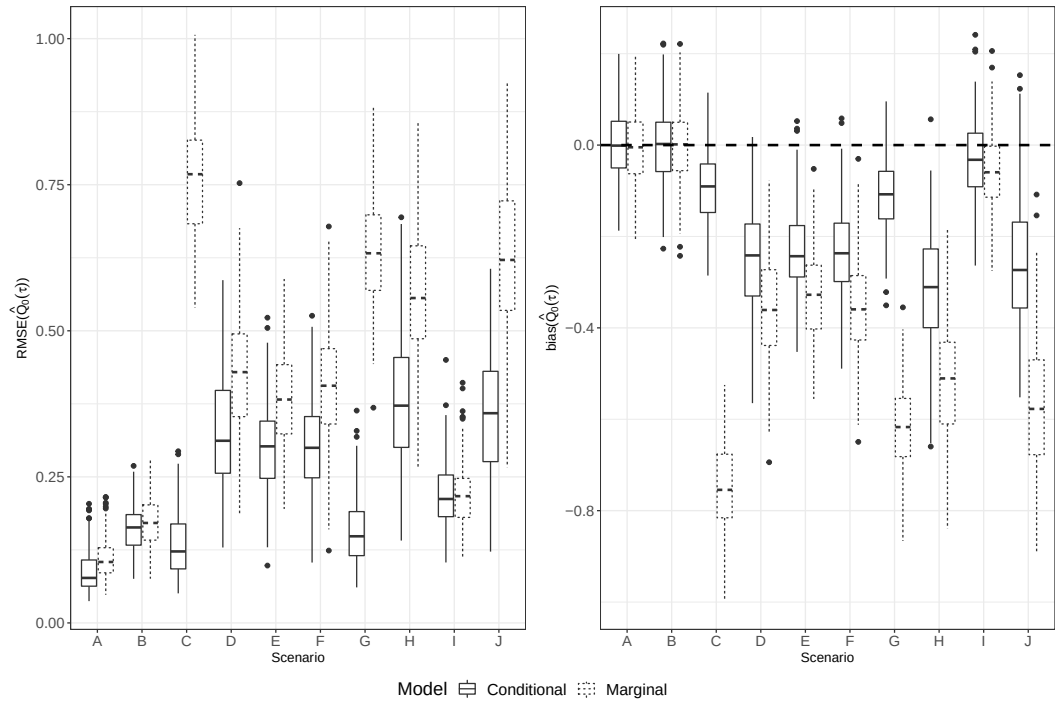

Figure 3.5: Boxplots of the RMSE (left) and the bias (right) of the linear predictor $\hat{Q}^{\tau,0}$ when data are simulated with $\beta(s) = \beta_1(s)$, and estimation is carried out with the spline approximation, either using the correct conditional model or a marginal model. See Table 3.1 for a detailed explanation of the scenarios.

Assessment of model-based standard errors

The output from `qgam()` includes an estimated variance-covariance matrix for the set of model parameters, which can be used to compute model-based standard errors (SEs) for quantities of interest such as the targets mentioned in Section 3.3.5. We now check the validity of these SEs in scenario D with $\beta(s) = \beta_1(s)$.

We consider targets ΔQ^τ for a variety of $\Delta X(\cdot)$ functions; in particular, $\Delta X(s) = \phi_1(s)$, $\Delta X(s) = \phi_2(s)$ and $\Delta X(s) = s$ corresponding to the primary directions in the generation of $X_{ij}(\cdot)$s, see equation (3.5.1), and $\Delta X = X_{80} - X_{20}$ where X_{20} and X_{80} are the pointwise 20% and 80% quantiles in the distribution of $X_{ij}(\cdot)$. The actual standard deviation over the 200 replicates, i.e. $\text{SD}(\widehat{\Delta Q}_r^\tau) = \sqrt{\sum_{r=1}^{200}\big(\widehat{\Delta Q}_r^\tau - \text{mean}(\widehat{\Delta Q}_r^\tau)\big)^2/199}$ is about a factor 1.3 larger than the average of the standard errors computed from the variance-covariance matrix from `qgam()` for all the tested choices of $\Delta X(\cdot)$, meaning that the standard errors underestimate the actual variation of the estimators. Moreover, we experience bias in the estimation of ΔQ^τ for some, but not all $\Delta X(\cdot)$ functions. This comes as no surprise considering the simulation results already reported. As a consequence, we implement the two bootstrap strategies from Section 3.3.5 in the application, one aiming at bias adjustment and one aiming at reliable estimation of the variability.

3.5.3 Time-varying regression coefficient

In this section we show the results from simulations where the functional coefficient is allowed to vary with longitudinal time. Specifically, we generate data as in equation (3.5.2) with either $\beta(s,t) = \beta_1(s,t) = \sqrt{2}\cos(2\pi s)t$ or $\beta(s,t) = \beta_2(s,t) = st$. In both cases the functional coefficients is zero at $t = 0$, increases linearly in t and match those considered in Section 3.5.2 at $t = 1$. We consider scenarios A–D from above, i.e the four combinations of $\gamma \in \{0, 0.5\}$ and $\tau \in \{0.5, 0.1\}$. For the representation of $\beta^\tau(\cdot,\cdot)$ we employ a tensor product smooth with five marginal cubic regression splines in the s as well as the t direction, and for the representation of $\alpha(\cdot)$ we use ten thin plate regression splines, which brought similar results compared to cubic regression splines; apart from this we proceed as in Section 3.5.2. Two-hundred replications are considered.

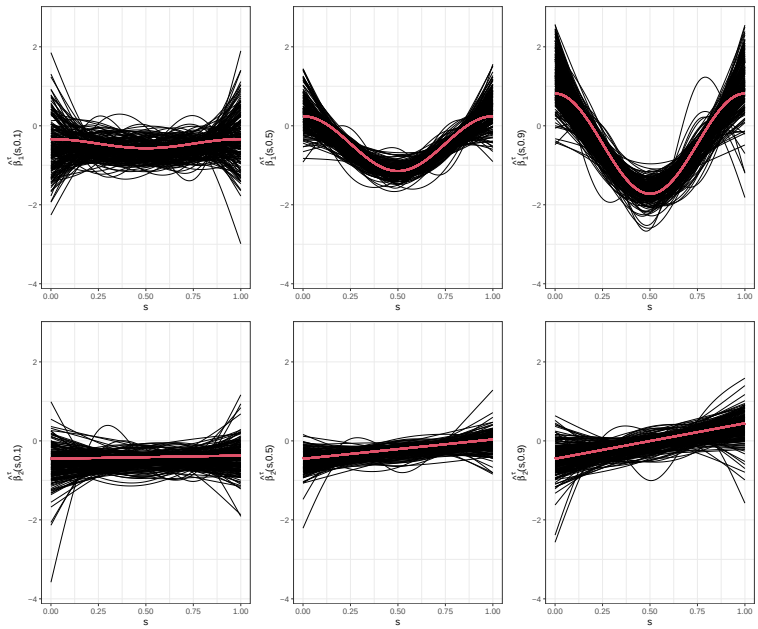

Figure 3.6: Estimated functional coefficients in scenario D when data are simulated with $\beta(s,t) = \beta_1(s,t)$ (top row) and when $\beta(s,t) = \beta_2(s,t)$ (bottom row). We show the results for fixed longitudinal time points, namely $t = 0.1$ (left column), $t = 0.5$ (central column) and $t = 0.9$ (right column). The red curves represent the true functional coefficients $\hat{\beta}^\tau(\cdot,t)$.

Figure 3.6 illustrates the estimated functional coefficients for scenario D ($\gamma = 0.5$, $\tau = 0.1$) when data was generated with $\beta_1(s,t)$ (top row) and $\beta_2(s,t)$ (bottom row) The estimated coeffcient functions take (s,t) as arguments; in order to compare to Figure 3.2 we show $s \mapsto \hat{\beta}^\tau(s,t)$ for $t = 0.1$ (left), $t = 0.5$ (middle) and $t = 0.9$ (right). The true functional coefficients are displayed in red. Estimation at the boundaries of T is inherently difficult; therefore we chose interior point of T, and we also see that variation is smaller at $t = 0.5$ compared to $t = 0.1$ and $t = 0.9$. Overall, the estimation reproduces that association between functional covariates and quantiles become stronger as t increases, and also roughly the correct shape of the coefficient.

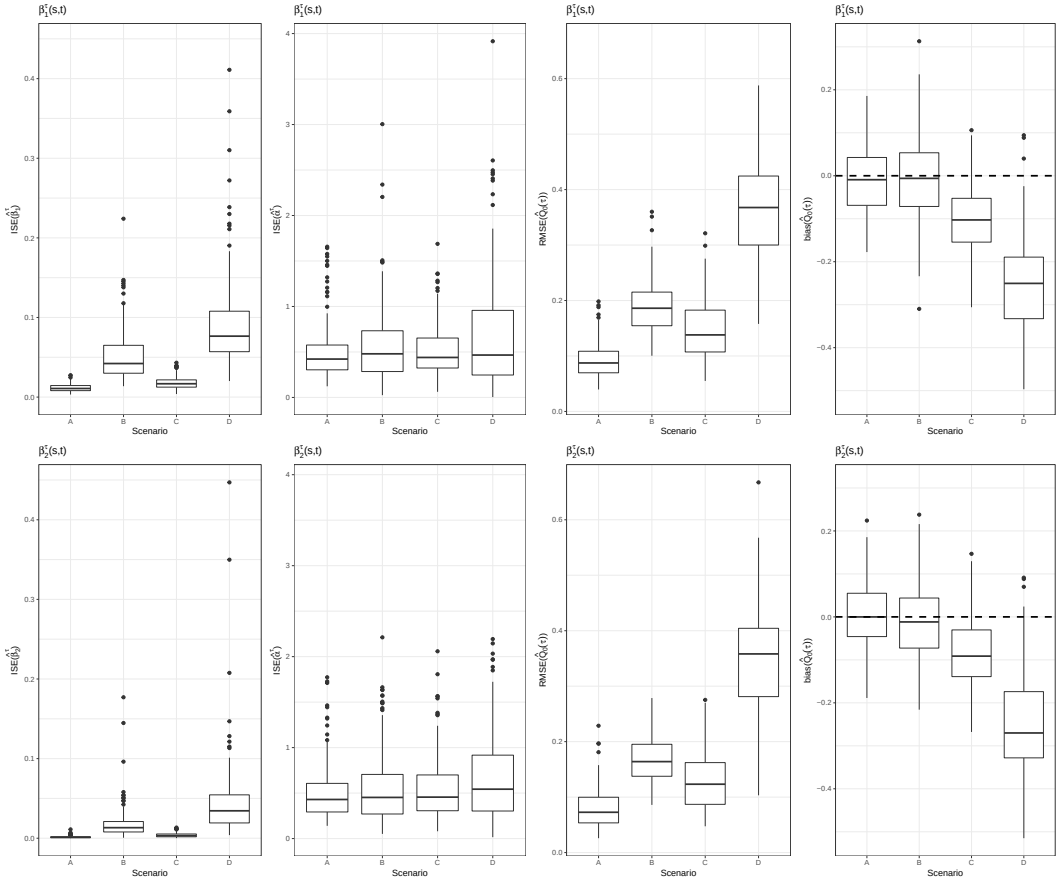

Figure 3.7: Boxplots of the ISE of the functional coefficient $\hat{\beta}^{\tau}(\cdot,\cdot)$ and of the smooth intercept $\hat{\alpha}^{\tau}(\cdot)$ (first and second columns respectively), as well as of the RMSE and the bias of the linear predictor $\hat{Q}^{\tau,0}$ (third and fourth column respectively) in the cases where data are simulated with $\beta(s,t) = \beta_1(s,t)$ (first row) and $\beta(s,t) = \beta_2(s,t)$ (second row). See Table 3.1 for a detailed explanation of the scenarios, but notice that data are now simulated with time-varying coefficients.

Similarly to Figure 3.3, in Figure 3.7 we compare the RMSE of the estimated linear predictor from (3.5.3) (first row), the ISE for the estimated smooth intercept $\hat{\alpha}^{\tau}(\cdot)$ (second row) and the ISE for the estimated functional coefficient at separate longitudinal time points (third row) for the four different scenarios corresponding to the combinations of $\gamma \in \{0, 0.5\}$ and $\tau \in \{0.1, 0.5\}$. As in Section 3.5.2 we see that scenario D with $\gamma = 0.5$ combined with estimation at $\tau = 0.1$ is the more difficult one, in particular when it comes to estimation the quantiles. As it was already visible in Figure 3.6, the variation in the distribution of the error is higher at the boundaries than in the middle point of the longitudinal time interval in all cases. Finally, notice how the distributions of the RMSE of the linear predictor are comparable with those in Figure 3.3, although with slightly increased variation, whereas the ISE of the smooth intercept and functional coefficients are generally much higher in the complex model compared to the simple model. We conclude that, even though the increased complexity of the model affects the estimation of the ingredients in the model, there is no dramatic change in the quality of the estimates of the quantiles themselves.

3.6 Application

Our ultimate goal is to study the impact of thermal conditions on the daily food intake for lactating sows, taking into account the progression of food intake over lactation days. A low food intake is an indicator of a high stress and has consequences for the sow itself (compromised reproductive system) as well as for the litter. In particular, low food intake may lead to increased body weight loss and reduced milk production implying slower and poorer weight gain of the litter (see Park et al. (2019), Staicu et al. (2020) and references therein). As a consequence, we are interested in the relation between quantiles of daily food intake at low quantiles levels and cell temperature along the hours, allowing for the association to vary over time.

The data comes from a commercial research unit in Oklahoma, where 480 sows were monitored from July to October 2013. The animals were divided into 21 groups approximately 5 days after giving birth and then assigned to cells, where they were kept under observations for the lactation period of up to 21 days. For each sow at each lactation day, the food intake (in kg) is available, as well as the cell temperature (in °C), measured every five minutes for 24 hours from 2.00 pm to 1.59 pm. Moreover, the parity of each sow is registered, i.e. the number of pregnancies the animal had before the current one. We will consider parity as a measure of age: a sow is "young" if it is at its first pregnancy and it is "old" otherwise. Previous studies have shown that younger and older sows behave differently (Staicu et al., 2020), so we analyze data from young and older sows separately. Finally, in our analysis the total number of animals is 475, 237 young sows and 238 old ones as a few unreliable observations were discarded by the experimenters.

The data are illustrated in Figure 3.8. Feed intake curves with three randomly selected animals from each age group are plotted in the left part of the figure. Although there is large within-sow variation over lactation days, it is also clear that some sows tend be have low (or high) feed intake throughout, calling for a subject-specific level of each sow. In a preprocessing step we smoothed the temperature curves with FACE (see Section 3.2), using a PVE as high as 0.9999; then the features of the curves are maintained and missing values can be replaced by the smoothed function values, while variation on small time-scale is partly smoothed away. The smoothed temperature curves are illustrated in the right part of Figure 3.8, and the corresponding pointwise 20% and 80% quantiles are shown in blue and red, respectively. We use these curves, denoted Temp_{20} and Temp_{80} respectively, when we report the results from our analyses below. In the same plot we show pointwise 50% quantile (in green) as well.

Staicu et al. (2020) used a longitudinal dynamic functional regression framework for mean regression for the same data with emphasis on prediction of response trajectories. Park et al. (2019) carried out separate quantile regression analyses for a derived variable at three selected lactation days. For each day, the CDF was first estimated and then inverted, hence estimated quantiles were extracted. In contrast, we carry out quantile regression for all lactation days simultaneously with the framework and methods introduced in Sections 3.2 and 3.3, and with focus on estimation and inference for the temperature effect on quantiles of feed intake.

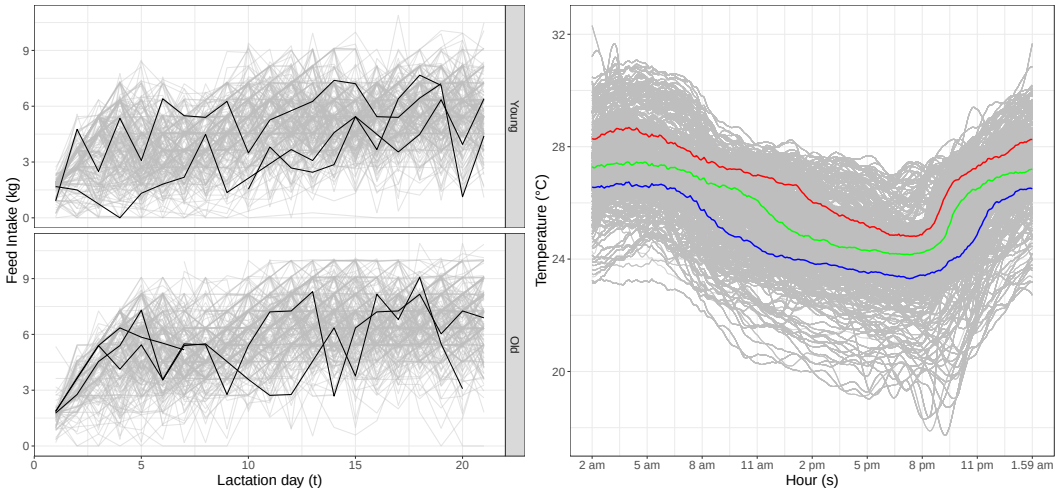

Figure 3.8: Descriptive plots of data. To the left, daily feed intake profiles over lactation days of young sows (upper panel) and of old sows (lower panel) with three randomly selected profiles (black) in each group. On the right, smoothed temperature curves (grey), as well as the pointwise temperature quantiles curves at quantile levels 0.2 (blue), 0.5 (green) and 0.8 (red) based on the whole dataset.

3.6.1 Estimated quantiles of feed intake

We consider data $\{(\mathrm{FI}_{ij}, \mathrm{Temp}_{ij}(\cdot), t_{ij})\}_{ij}$. For each sow $i = 1, \ldots, N$ ($N = 475$) and repeated measurement $j = 1, \ldots, n_i$ (n_j ranging from 7 to 21), FI_{ij} refers to the daily feed intake expressed in kg, $\mathrm{Temp}_{ij}(\cdot)$ to the smoothed temperature function in °C recorded over a day and t_{ij} to the lactation day. We allow for a subject specific intercept u_i^τ to account for the correlation of observations from the same sow. For each age group, we consider the model

$$Q_{\mathrm{FI}_{ij}|\mathrm{Temp}_{ij},t_{ij},u_i^\tau}(\tau) = u_i^\tau + \alpha^\tau(t_{ij}) + \int_S \beta^\tau(s, t_{ij})\mathrm{Temp}_{ij}(s)ds, \quad i = 1, \ldots, N,\ j = 1, \ldots, n_i \tag{3.6.1}$$

where S represents a whole day from 2.00 pm to 1.59 pm. We approximate the smooth intercept $\alpha^\tau(\cdot)$ by means of ten cubic splines, and for coefficient function $\beta^\tau(\cdot, \cdot)$, we compare representations in terms of splines and eigenfunctions. For the spline approach we employ ten cyclic cubic splines and ten cubic splines for the s- and t- directions of $\beta^\tau(\cdot, \cdot)$ respectively. We choose a cyclic basis in the s-direction since both end-points in S represent the same time of day (except for five minutes). We use Algorithm 2 for the eigenfunction approach and the $\mathrm{BIC}^{\mathrm{adj}}$ criterion selects 2 and 9 eigenfunctions for the sub-datasets of the young and old sows, respectively.

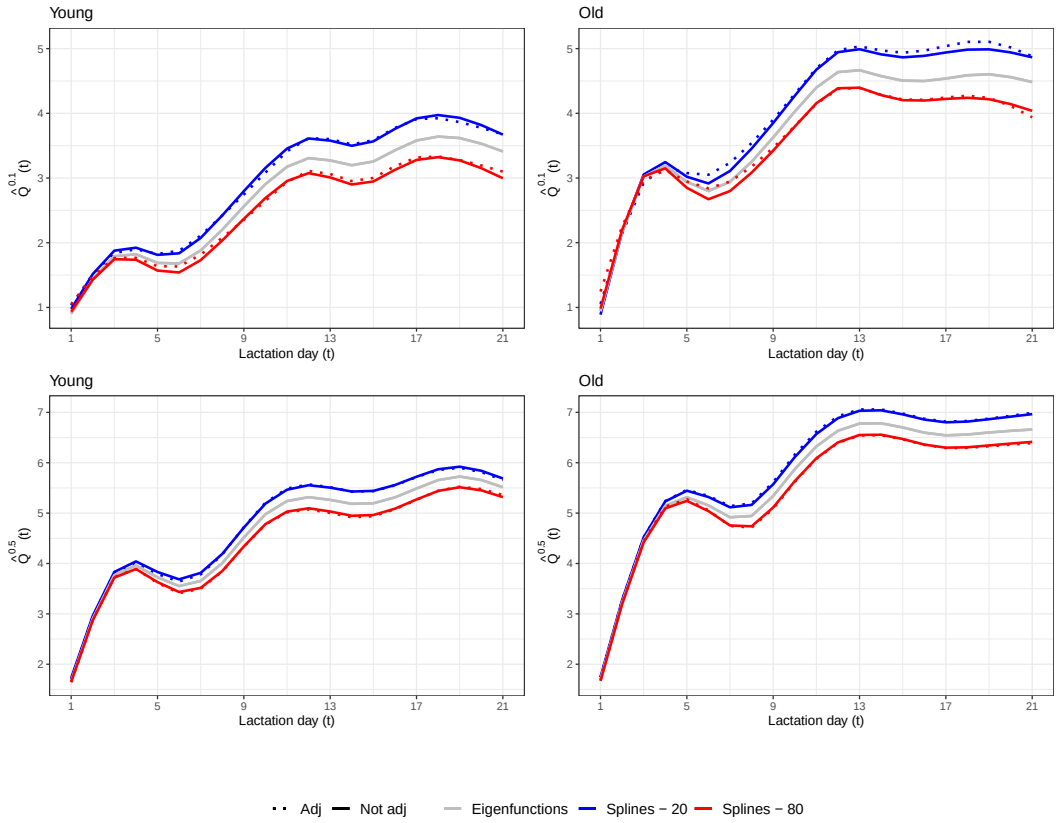

Figure 3.9: Predicted quantiles corresponding to the 20% and 80% pointwise temperature profiles when using splines and eigenfunctions approximations (solid lines in blue and red for the former and grey for the latter). Bootstrap-adjusted estimates are shown with dotted curves for spline approximations. The left column refers to sows at their first pregnancy, while right one refers to the older sows. Results at quantile levels $\tau = 0.1$ and $\tau = 0.5$ are shown in the top and bottom row, respectively. Notice that predicted quantiles at different levels are plotted on different scales.

Figure 3.9 shows estimated quantile profiles for young/old sows (left/right), at quantile levels 0.1/0.5 (top/bottom), and for the pointwise 20% and 80% temperature curves Temp_{20} and Temp_{80} (colours are explained below). More specifically, the graphs show

$$\hat{Q}^{\tau}_{20}(t) = \hat{\alpha}^{\tau}(t) + \int_S \text{Temp}_{20}(s)\hat{\beta}^{\tau}(s,t)ds, \quad \hat{Q}^{\tau}_{80}(t) = \hat{\alpha}^{\tau}(t) + \int_S \text{Temp}_{80}(s)\hat{\beta}^{\tau}(s,t)ds,$$

plotted over t, for each age group and for $\tau = 0.1, 0.5$ where $\hat{\alpha}^{\tau}(\cdot)$ and $\hat{\beta}^{\tau}(\cdot,\cdot)$ are estimated with either the spline or eigenfunction approach. Notice that no random effects are included in the predictions such that their interpretation is for an "average sow" (with $u_i^{\tau} = 0$).

The solid blue and red curves are estimated profiles obtained by the spline method, and we see a clear distinction between low (blue) and high (red) temperatures, at least from around lactation day five. High temperatures negatively influence the appetite of the animals—they tend to eat more in cooler conditions—and this difference increases

over time, particularly at the 0.1 level. The group of young and old sows have similar 10% quantiles of feed intake at the beginning of their stay in the cells, but develop slightly different eating habits over lactation days. The 10% quantile stabilises around 4.5 kg for older sows when the environment is warm and around 5 kg when it is cold, whereas the quantile for sows at their first pregnancy has a lower plateau, which is reached later in time, and takes values of approximately 3.5 kg and 4 kg for high and low temperature profile, respectively. At quantile level 0.5 these trends are preserved, with plateaus at 6 kg and 7 kg for young and old animals, respectively, in the lower temperature profile, and at 5.5 kg and 6.5 kg for the higher temperature profile. The dashed curves show the bias-adjusted estimates. Although the bias-adjustment is hardly visible, it is actually significantly different from zero at many instances at a 5% significance level (based on pointwise one-sample t-tests).

Turning to the eigenfunction approach, the difference between $\hat{Q}^{\tau}_{20}(t)$ and $\hat{Q}^{\tau}_{80}(t)$ is negligible meaning that no temperature effect has been identified. The profiles, which cannot the distinguished from one another, are plotted in grey in Figure 3.9, and fall in between the estimates from the spline method; as an average over temperature curves. It is interesting that only one of the approximation methods is able to identify a temperature effect, and not even inclusion of more eigenfunctions than suggested by the BIC^{adj} criterion makes the estimated temperature effect non-negligible. We only use the spline approach in the following, and compare results for model (3.6.1) to results for simpler models without temperature effects (Section 3.6.2).

In order to illustrate the estimated temperature effects from the spline approach more clearly, Figure 3.10 plots differences between the estimated feed intake quantile profiles for low and high temperatures, i.e. $\hat{D}^{\tau}(t) = \hat{Q}^{\tau}_{20}(t) - \hat{Q}^{\tau}_{80}(t)$. The black curve and confidence bands show the estimates without adjustment and the corresponding model-based 95% pointwise confidence interval based on the variance-covariance matrix reported from the analysis, and the orange curve and confidence bands show the bias adjusted estimates and confidence bands obtained by bootstrap, cf. equation (3.3.13). We used 100 bootstrap samples for each of the two bootstrap procedures. Bias adjustment is most prominent for young sows at the 0.1 quantile, with all pointwise P-values smaller than 0.00015. For the other three cases, P-values are below 0.01 for all except 3–4 lactation days. The bootstrap generated confidence bands are always wider than the model-based. Since the simulation results indicated that the model based standard errors underestimate the actual variation, we are in favour of the bootstrap generated confidence intervals. The profiles obtained from the 100 bootstrap datasets are shown in Figure 3.12.

No matter which methods we use for construction of confidence bands and no matter whether we adjust for bias or not, the overall conclusion is the same: No temperature effect is found early in the lactation period (up to around day five), but at later days quantiles of feed intake are negatively affected by high temperature at 0.1 and 0.5 quantile levels. In general, the influence of temperature on the quantiles becomes more prominent along the lactation period, but there are certain differences between age groups and between quantile levels. At quantile level 0.1 the difference in predictions has an increasing trend along lactation days for both groups of sows, while at the median the difference in predictions reaches a maximum of approximately 0.5 around lactation day 10–13 and then flattens out (or even decreases sligthly). This might indicate that the sows that eat less are those particularly sensible to the environmental temperature.

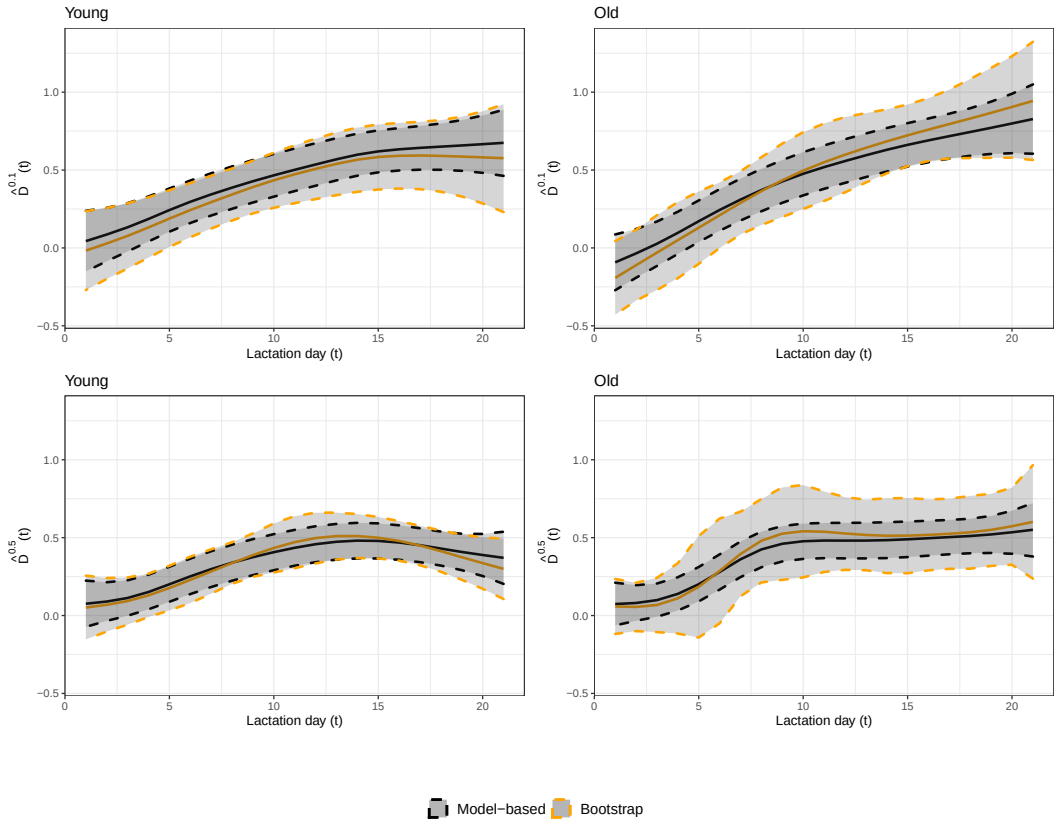

Figure 3.10: Estimated differences in quantiles between the pointwise 20% and 80% temperature curves, both without (solid black) and with (solid orange) bias adjustment. The corresponding pointwise confidence intervals, based on the model solely or on bootstrap, are illustrated with dashed curves. The left column refers to sows at their first pregnancy, while the right column refers to the older sows. Results at levels $\tau = 0, 1$ and $\tau = 0.5$ are shown in the top and bottom row respectively.

3.6.2 Comparison with simpler models

Now, let us turn to a comparison of the model (3.6.1), with three modifications which all have simpler specifications of the effect of temperature. The first modification has $\beta(s,t) = \beta_A(s)$ such that the temperature curves still have functional effects, but with same effect across lactation days; this would correspond to profiles of differences in Figure 3.10 being constant. The second modification has $\beta(s,t) = \beta_B(t)$. Then the model (3.6.1) becomes

$$Q_{\mathrm{FI}_{ij}|\mathrm{Temp}_{ij},t_{ij},u_i^\tau}(\tau) = u_i^\tau + \alpha^\tau(t_{ij}) + \beta_B^\tau(t_{ij}) \int \mathrm{Temp}_{ij}(s)\, ds \tag{3.6.2}$$

such that the quantile depends on the temperature curve only through its integral or, equivalently, the average temperature over the day, and it is no longer a functional quantile regression model. The third modification combines the two previous sub-models; it has

$\beta(s,t) = \beta_C$, such that

$$Q_{\mathrm{FI}_{ij}|\mathrm{Temp}_{ij},t_{ij},u_i^\tau}(\tau) = u_i^\tau + \alpha^\tau(t_{ij}) + \beta_C \int \mathrm{Temp}_{ij}(s)\, ds \tag{3.6.3}$$

and the temperature effect is the same across days and depends on the average temperature over the day only.

We measure goodness-of-fit with the AIC values based on the log-likelihood corresponding to the ELF distribution (Fasiolo et al., 2020) and the effective degrees of freedom (EDF) known from additive models (Wood, 2017), but it would also be possible to rely on other criteria, for instance the BIC, which penalizes a model more for its complexity. The EDF is partitioned into two parts; the effective degrees of freedom for the smooth coefficients ($\alpha^\tau(\cdot)$ and $\beta^\tau(\cdot,\cdot)$), denoted $\mathrm{EDF}_{\alpha,\beta}$, and the degrees of freedom corresponding to the subject-specific intercepts (u_i^τ), denoted EDF_u. Table 3.2 displays the AIC values and also the EDFs for the four models, with the smallest AIC value emphasized in each line.

		$\beta(s,t)$			$\beta_A(s)$			$\beta_B(t)$			β_C		
	τ	AIC	$\mathrm{EDF}_{\alpha,\beta}$	EDF_u	AIC	$\mathrm{EDF}_{\alpha,\beta}$	EDF_u	AIC	$\mathrm{EDF}_{\alpha,\beta}$	EDF_u	AIC	$\mathrm{EDF}_{\alpha,\beta}$	EDF_u
Young	0.1	*18527*	19	191	18769	13	188	18724	13	189	18790	10	188
	0.5	15937	19	206	15947	14	206	*15934*	14	206	15946	11	206
Old	0.1	*18483*	24	201	18769	15	198	18633	13	199	18802	11	198
	0.5	*16044*	26	206	16048	15	207	16047	14	206	16060	11	207

Table 3.2: AIC and sum of effective degrees of freedom for young and old animals when adopting model (3.6.1) (first column), model (3.6.1) with $\beta(s,t) = \beta_A(s)$ (second column), model (3.6.2) (third column) and model (3.6.3) (fourth column).

For both groups the AIC values corresponding to the most complex model are the smallest by a reasonably wide margin when for quantile level $\tau = 0.1$, while the values are closer among the models at the median. This indicates that it is particularly important to allow for time-varying coefficients and functional effects at lower quantiles. At level $\tau = 0.5$ the most complex model is still selected for old sows, but for younger animals the smallest AIC is the one from model (3.6.2). Furthermore, in all cases, the AIC values from the model with $\beta_B(t)$ are smaller than the AIC value from the model with $\beta_A(s)$, indicating that it is more important to account for the temperature variation in the development along the lactation period than over the day.

For the EDFs, we notice that both EDF_p, as expected, is the highest for the model (3.6.1) since it describes variation in both the s and t direction. Both $\mathrm{EDF}_{\alpha,\beta}$ and EDF_u are larger when estimation is carried out at the median rather than at the 10% level; most likely because there is more information in the data about the median and thus more room for flexibility in the estimation. Finally, notice that EDF_u is always between 188 and 207, and thus smaller than 237 and 238, the number of young and old sows, respectively, so random effects are penalized effectively.

3.6.3 The estimated effect of temperature

Finally, we turn the attention to the estimated coefficient function $\hat{\beta}^\tau(\cdot,\cdot)$ in model (3.6.1), while having in mind that the function is not identified in the full space of functions.

Figure 3.11 shows $\hat{\beta}^{\tau}(\cdot, \cdot)$ for each age group and at quantile levels 0.1 and 0.5, respectively. In each panel, $s \mapsto \hat{\beta}^{\tau}(\cdot, t)$ is plotted for t fixed at each lactation day, on a colour scale that ranges from orange to green as the longitudinal time goes by. Recall that, by construction, the development in both s and t direction is smooth, and the functions are cyclic over day.

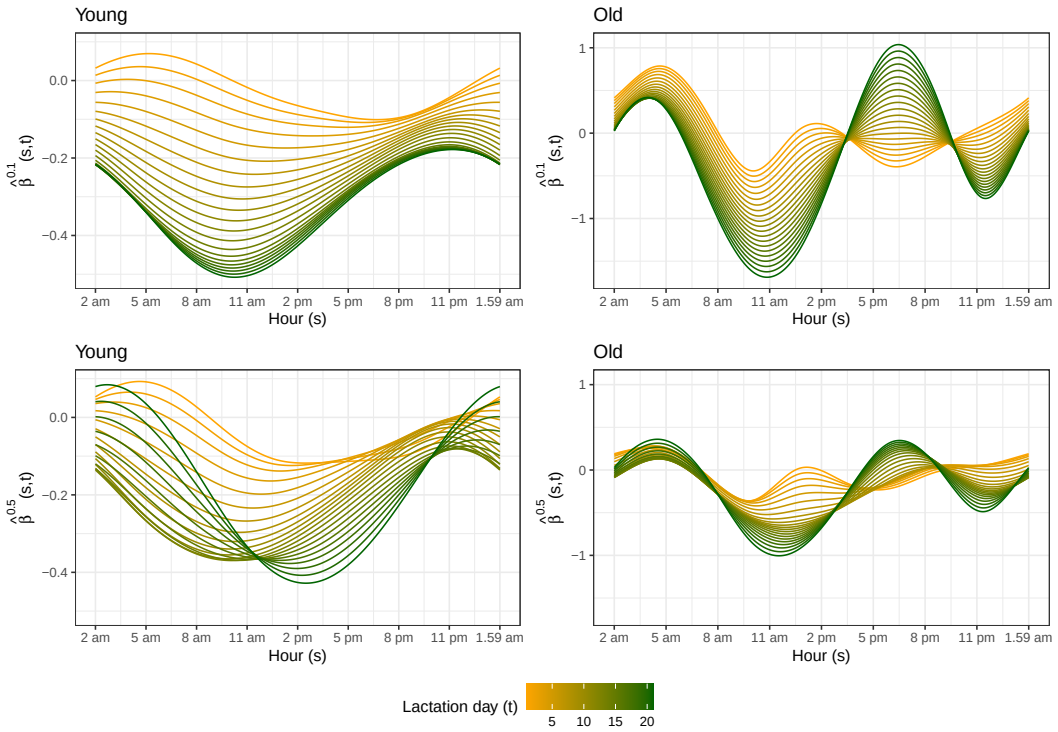

Figure 3.11: Illustration of the estimated coefficient function $\hat{\beta}^{\tau}(\cdot, \cdot)$ estimated by means of a tensor smooth for level $\tau = 0.1$ (top row) and $\tau = 0.5$ (bottom row), for both young (left column) and older sows (right column). Curves show $s \mapsto \hat{\beta}^{\tau}(\cdot, t)$ for each lactation days t.

The estimated coefficient functions are predominantly, although not everywhere, negative corresponding to the overall negative effect of temperature illustrated in Figure 3.8. With the risk of overinterpretation, we see that the impact of temperature on feed intake is most prominent in the morning hours (from about 8 am to about 12, a bit later at late lactation days for young sows at the median). This is also the time of the day with the largest differences in temperature effects between lactation days. Sensitivity against temperature appears to increase over lactation days, but stabilizes earlier for young compared to older sows.

3.7 Discussion

Our ultimate aim was to study how heat stress affects the health, monitored in terms of daily food intake, of lactating sows. To achieve this, we proposed a model and estimation framework for scalar-on-function quantile regression for clustered or longitudinal data. In particular, dependence within cluster/subject is taken into account by including cluster- or subject-specific intercept parameters. Estimation relies on basis expansions, using either

penalized splines or eigenfunctions for the functional covariates' covariance operator, and existing software can be applied to the new framework. We compared the estimation methods with simulations, and although the results were similar, we prefer the spline approach for its flexibility and because the eigenfunction approach requires as extra step where the level of truncation is selected. Simulations also revealed the importance of taking the within-subject dependence into account in the estimation process, and that estimates of quantiles can be biased and standard errors underestimated. The last observation inspired us to also consider bootstrap-based methods for bias adjustment and estimation of uncertainty.

Regarding the data application, our analysis offered some interesting insights. First, quantiles are similar for younger and older sows close to giving birth, but increase faster and to a higher level for older than younger sows, suggesting that sows at their second or later pregnancy acclimatise faster to the environment. Second, a high temperature in the stable affects feed intake negatively except for the early days in the lactation period; this is the case for both younger and older sows, and both at the median and at the 0.1 quantile levels. At early lactation days, the temperature effect is not significant, and also similar for both groups of sows. Third, the estimated temperature effect is generally larger at the 0.1 level compared to the 0.5 level, suggesting that sows that eat less are more sensible to temperature changes than the majority of sows; however this should be investigated further. Fourth, there is an increasing trend of the temperature effect throughout the lactation period at the 0.1 quantile level, and steeper for the older sows. This is confirmed by model comparisons where models with time-varying temperature effect are preferred over models with constant temperature effects. Fifth, for both groups and at the 0.1% quantile level, the model with functional effect of temperature over the day is preferred over a model which includes the average temperature only.

We have focused on models with a single functional covariate, but they could be extended to include more than one functional covariates or a mixture of functional and scalar covariates in a straight-forward way. Moreover, several grouping levels could be included as random effects; this could be relevant in the application because sows were kept together in the stables. It remains to study the robustness of estimates in such more complex models. The proposed models have similar flavor as models from Brockhaus et al. (2017), but we were not able to get reliable estimates with the software.

We adjusted our estimates and the corresponding standard errors with bootstrap methods. The sampling schemes have been used and studied in simpler models (Galvao and Montes-Rojas, 2015; Battagliola et al., 2021); yet the approach is ad hoc and further examination would be useful in future studies. Another topic for future research is hypothesis testing for functional effects. With inspiration from Abramowicz et al. (2018) and Pini et al. (2021), our preliminary ideas are based on permutation tests, where test statistics are computed as integrals over the domain S of pointwise test statistics, and their null distribution evaluated by bootstrap. The main challenge lies in designing appropriate permutation schemes that comply with the dependence structures in the data.

3.8 Acknowledgements

The project was partly funded by the Danish Research Council (DFF grant 7014-00221). The authors would like to thank Santa Maria Mendoza Benavides and Erik van Heugten for the data used. The data originated from work supported in part by the North Carolina Agricultural Foundation, Raleigh, NC.

3.9 Appendix

Extra results from simulation studies

Table 3.3 shows results from simulations in Section 3.5.2 with time-invariant coefficient function. Data are simulated with $\beta(s) = \beta_1(s)$, and estimation is carried out with the spline approximation. The numbers reported in the table are averages over 200 simulated datasets.

Scenario	bias($\hat{Q}^{0,\tau}$)	SD($\hat{Q}^{0,\tau}$)	RMSE($\hat{Q}^{0,\tau}$)	RMSE($\hat{Q}^{0,\tau}_{oos}$)	RMSE($\hat{Q}^{0,\tau}_{marg}$)
A	-0.00	0.06	0.09	0.09	0.11
B	0.00	0.14	0.16	0.16	0.17
C	-0.09	0.08	0.13	0.14	0.76
D	-0.24	0.21	0.33	0.33	0.42
E	-0.23	0.18	0.30	0.30	0.38
F	-0.24	0.18	0.30	0.31	0.41
G	-0.11	0.09	0.16	0.16	0.64
H	-0.32	0.19	0.38	0.38	0.56
I	-0.03	0.20	0.22	0.22	0.22
J	-0.26	0.22	0.36	0.36	0.63

Table 3.3: Mean values (over the 200 replications) of bias, standard deviation and RMSE of linear predictor computed in-sample (second, third, and fourth columns, respectively) as well as the RMSE of the linear predictor computed out-of-sample (fifth column) and the RMSE of the linear predictor obtained with the marginal model (sixth column).

Extra results from the application

Recall that $\hat{D}^{\tau}(\cdot)$ is the estimated difference in quantile between the pointwise 20% and 80% temperature curves. Figure 3.12 shows estimated differences $\hat{D}^{\tau,*}(\cdot)$ obtained from bootstrap data, for the young sows and at quantile level $\tau = 0.1$. For the left plot bootstrap data are sampled by cross-sectional resampling (block resampling), resampling sows, whereas for the right plot data are resampled by the wild bootstrap method explained in Section 3.3.5. One-hundred datasets are generated in both cases. The estimate from the observed data, $\hat{D}^{\tau}(\cdot)$, is shown in black.

The bootstrap estimates vary around the data estimate for the cross-sectional bootstrap, whereas the bootstrap estimates are on average larger than the data estimate for the wild bootstrap. Since the wild bootstrap data are generated such that $\hat{D}^{\tau}(\cdot)$ is the true value, the deviation measures the bias, and we use it for bias correction, see the upper left part of Figure 3.10. It is also clear from the graphs that the variation of estimates is larger for cross-sectional bootstrap than for wild bootstrap.

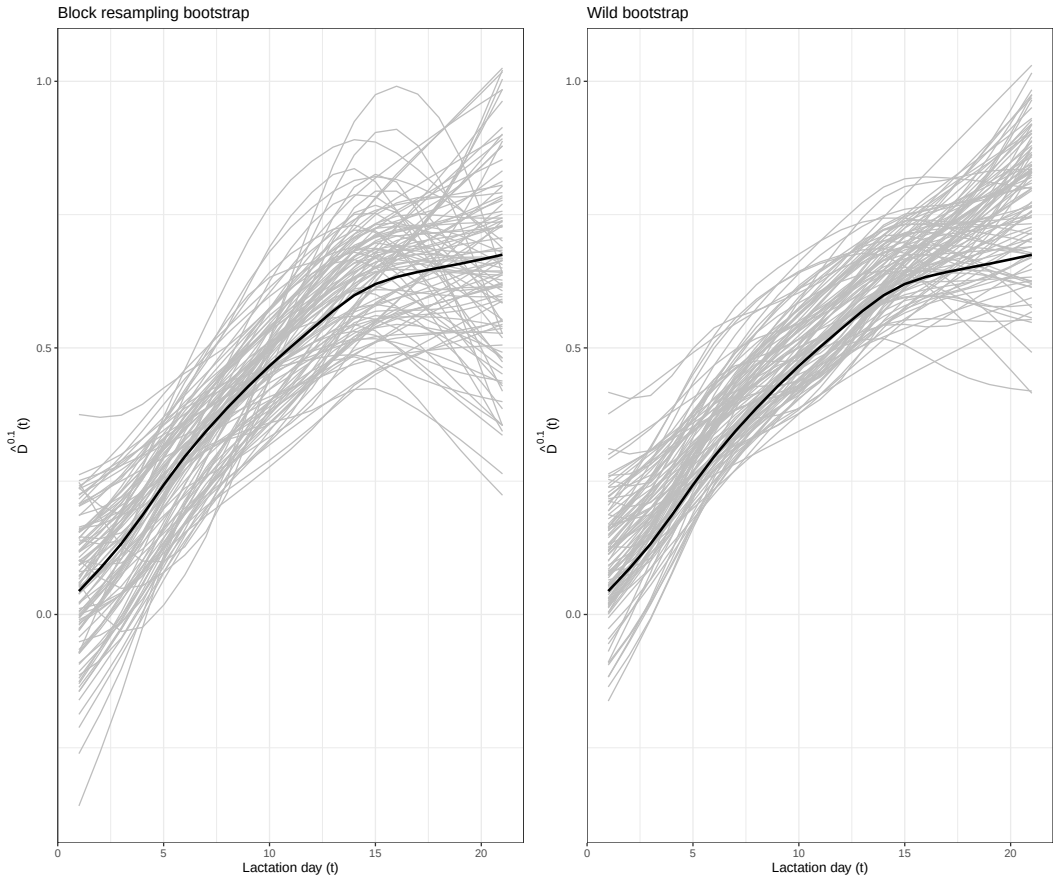

Figure 3.12: Estimated differences in quantiles between the pointwise 20% and 80% temperature curves, obtained from 100 bootstrap datasets generated by cross-sectional (block) resampling (left) and wild bootstrap (right) for young sows at quantile level 0.1. The black curves represent the estimate obtained from the observed data.

Chapter 4

Analysis of learning curves of mice by phase-amplitude separation

Maria Laura Battagliola, Laura J. Benoit, Sarah Canetta & R. Todd Ogden

In progress

Abstract

The manuscript presents an analysis of learning curves of mice recorded when the animals were trained and then tested on how to successfully complete a working memory-challenging task multiple times. In particular, we are interested in the comparison between the behavior of mice in a control group and mice with an induced brain lesion that resembles one that is common in patients affected by psychiatric disorders. We rely on existing methods in order to analyze bivariate functional objects composed of amplitude and phase components arising from registration of the learning curves. We compare the results corresponding to the two groups in different scenarios, each characterized by a different level of difficulty of the task.

Keywords: Learning curves; Registration; Square root velocity functions; Multivariate functional principal component analysis

4.1 Introduction

This work is motivated by the study of working memory impairment that might affect patients with psychiatric disorders. In particular, our analysis is based on the data presented by Benoit et al. (2020). In their paper they studied the behavior of mice when undergoing a task designed to challenge their memory. In particular, they considered a control group of mice and a group characterized by mice with an induced brain lesion that resembles one that can be detected in psychiatric patients affected by, for instance, schizophrenia. Every animal belonging to both groups was first trained to complete the task and then tested on completing the same task, but when delays were introduced. In particular, the task was considered successfully completed if the animal was able to remember in the second half of the task what happened in the first half. In the acquisition, when mice were trained, there was no delay between the first and second parts of the task. Afterwards, a delay of either 2, 4, 8 or 16 seconds was randomly chosen and introduced between the two halves of the task. During both acquisition and test, every animal undertook trials of such task multiple times a day over several days, and successes and failures were recorded.

Benoit et al. (2020) analyzed multivariate data arising from averaging the binary results of the trials at each day of the experiment. We decided to work instead with learning curves, with the idea that such framework could bring a more nuanced analysis of the data. In such context, we relied on tools from functional data analysis (see for instance Ramsay and Silverman (2005) for an overview on the topic). When a collection

of curves is available, it can be of interest to perform phase-amplitude separation on them. Such analysis allows to estimate the amplitude and phase components that play a key role in the registration, or alignment, of the curves. Registration is sometimes regarded as a pre-processing analysis, and several methods are available in the literature (see for instance Ramsay and Silverman (2005, Chapter 7), Gervini and Gasser (2004), Sangalli et al. (2010), Kneip and Ramsay (2008), Tang and Müller (2008), and Marron et al. (2015) for an overview). However, amplitude and phase components can bring precious information about the underlying trends of the observed curves. In our data application, studying the amplitude variability can bring an insight on how good the performance of the mice can be in terms of probability of completing a trial with success, while the phase variability gives an idea on how fast the animals make progress. Happ et al. (2019) proposed a multivariate functional principal component analysis (MFPCA) performed on the bivariate functional objects whose univariate elements are amplitude and phase components, so as to account for the simultaneous sources of variation between the these two functions. In this manuscript we rely on the findings and methods outlined by Happ et al. (2019) to give an insight into our data application.

In what follows, Section 4.2 outlines the mathematical framework we use, and Section 4.3 presents our analysis of the learning curves of the two groups of mice. Finally, in Section 4.4 we summarize our findings and discuss possible further developments.

4.2 Methods

This section is dedicated to the description of the mathematical framework we use for our data analysis. As mentioned in Section 4.1, our work is based on the methods suggested by Happ et al. (2019).

4.2.1 Registration via SRVF

Registration is an important tool in functional data analysis when it comes to the study of a collection of curves. In general, such analysis allows to quantify the warpings, or phase components, which deform functions so as to make them "aligned". The resulting aligned curves are regarded as amplitude components and such curves in general have features such as peaks and valleys occurring at similar coordinate point. Several approaches for phase-amplitude separation are available in the literature, from landmark to model-based registration methods. We rely on the work presented by Srivastava and Klassen (2016, Chapters 4, 7, 8) (see Wu and Srivastava (2014) for a concise yet complete description of the framework), and we briefly outline it in what follows.

Without loss of generality, consider time domain $T = [0, 1]$ and function $f \in \mathcal{F}(T)$, where $\mathcal{F}(T) \subset L^2(T)$ is the set of almost everywhere differentiable curves on T such that $f(t) = f(0) + \int_0^t \frac{df(s)}{ds} ds$, with $t \in T$. The square root velocity function (SRVF) corresponding to f is $q : L^2(T) \to \mathbb{R}$ and it is defined as

$$q(t) = \text{sign}\left(\frac{df(t)}{dt}\right)\sqrt{\left|\frac{df(t)}{dt}\right|}.$$

Notice that $q \in L^2(T)$ and we can always recover any f from its SRVF transformation q and $f(0)$ by

$$f(t) = f(0) + \int_0^t q(s)|q(s)|ds, \tag{4.2.1}$$

with $t \in T$. In particular, the Fisher-Rao (FR) metric for $f \in \mathcal{F}(T)$ becomes the L^2 metric for the corresponding SRVF q. Furthermore, consider

$$\Gamma_T = \{\gamma : T \to T \mid \gamma(0) = 0, \gamma(1) = 1, \gamma \text{ diffeomorphism}\},$$

the set of boundary-preserving diffeomorphisms of T. Such functions are smooth and invertible, and their inverse is smooth as well. We call Γ_T the set of the warping functions, which deform any $f \in \mathcal{F}(T)$ by right composition. We call such composition the warped version of f, and we denote it as $f \circ \gamma$. Moreover, the corresponding SRVF transformation of the warped function is $(q, \gamma) = (q \circ \gamma)\sqrt{\frac{d\gamma}{dt}}$, where q is the SRVF of f. Broadly speaking, registration is an analysis that allows to estimate the warping functions that deform curves so as to make them aligned. From a mathematical point of view, in order to be able to estimate the warping functions, it is important to adopt a proper criterion to be optimized. A possible choice could be taking the L^2 norm of the difference between two functions in order to measure the distance between them. However, taking $f_1, f_2 \in \mathcal{F}(T)$ and any $\gamma \in \Gamma_T$, such criterion would lack isometry under warping, namely $||f_1 \circ \gamma - f_2 \circ \gamma||_2 \neq ||f_1 - f_2||_2$. This entails that the L^2 norm is not suitable in this framework, since it does not preserve the distance between two functions when deformed in the same way. The main strength of working with the SRVF transformations $q_1, q_2 \in L^2(T)$ of the original functions is that they guarantee that the isometry under warping hold true, namely $||(q_1, \gamma) - (q_2, \gamma)||_2 = ||q_1 - q_2||_2$. This property is important when defining a metric for more complex spaces.

Another desirable property of the registration setting is that the amplitude of a function is left unchanged when it undergoes different warpings. Regarding amplitude as an "absolute" characteristic of a function, Srivastava and Klassen (2016, Chapters 4, 7, 8) considered it as an equivalence class, or orbit, and combined such assumption with the SRVF transformation framework. In particular, the amplitude of function $f \in \mathcal{F}(T)$ corresponds to all its warpings given by set Γ_T and their limit points:

$$[f] = closure\{f \circ \gamma \mid \gamma \in \Gamma_T\}.$$

A similar definition of amplitude holds for q, the SRVF transform of f, namely

$$[q] = closure\{(q, \gamma) \mid \gamma \in \Gamma_T\}$$

with $[q] \subset L^2$, and S the set of all orbits $[q]$. When one is interested in the pairwise alignment of function f_2 to function f_1, with $f_1, f_2 \in \mathcal{F}(T)$ whose corresponding SRVF transforms are q_1 and q_2 respectively, then the amplitude distance is defined as

$$d_a([q_1], [q_2]) = \min_{\gamma \in \Gamma_T} ||q_1 - (q_2, \gamma)||_2. \tag{4.2.2}$$

The phase and amplitude components of the phase-amplitude separation are the warping function γ that achieves the minimum in (4.2.2) and the warped f_2 that is aligned to f_1, namely

$$\tilde{f_2} = f_2 \circ \gamma,$$

respectively.

When it comes to the registration of a group of functions $f_1, \ldots, f_N \in \mathcal{F}(T)$ the procedure is similar, but instead of choosing one function from the collection to align all the others to, one usually aligns $f_1, \ldots, f_N$ to a template that captures the characteristics of all the curves. Consider the corresponding SRVF transforms $q_1, \ldots, q_N$ of the collection

of curves. Srivastava and Klassen (2016, Chapter 8) suggested taking the Karcher mean, namely

$$[\mu] = \underset{[q]\in S}{\arg\min} \sum_{i=1}^{N} d_a([q],[q_i])^2,$$

which corresponds to a generalization of mean in a metric space. Then, the template to align the functions to is μ_q, the center of orbit $[\mu]$. In particular, the warping functions $\gamma_1,\ldots,\gamma_N$ that align $q_1,\ldots,q_N$ to μ_q have the identity $\gamma_{id}(t) = t$ as the sample Karcher mean under the FR metric (see Srivastava and Klassen (2016, Chapter 7)). Similarly as in the pairwise alignment, the phase components of the phase-amplitude separation are $\gamma_1,\ldots,\gamma_N$, while the amplitude components are $\tilde{f}_1,\ldots,\tilde{f}_N$, with $\tilde{f}_i = f_i \circ \gamma_i$ for $i = 1,\ldots,N$. The two collections of phase and amplitude representatives are the first building block of the analysis carried our by Happ et al. (2019) and used in our application section.

4.2.2 MFPCA on amplitude and transformed phase components

Once the amplitude and phase components from the registration of $f_1,\ldots,f_2$ are available, it would be desirable to study them with classical functional data analysis tools, such as functional principal components analysis (FPCA). However, due to the non-convexity of Γ_T (Lee and Jung, 2016; Srivastava and Klassen, 2016, Chapter 4) it is not possible to carry out FPCA directly on the collection of warping functions $\gamma_1,\ldots,\gamma_N$. To overcome this drawback one needs to transform the phase components by some map ψ such that

$$\psi : \Gamma_T \to S^2(T), \tag{4.2.3}$$

where $S^2(T)$ is some convex subspace of $L^2(T)$, and hence conduct FPCA on $\psi(\gamma_1),\ldots,\psi(\gamma_N)$. An important feature that map ψ should have is to be a bijection, since we wish to transform the findings of any analysis back to Γ_T using $\psi^{-1} : S^2(T) \to \Gamma_T$. Happ et al. (2019) reviewed several possible choices for ψ and, out of preservation of geometric structure and computational stability arguments, they recommend relying on the centred log-ratio transformation (Egozcue et al., 2006; Hron et al., 2016), that we explain in what follows.

Given the characteristics of the elements of Γ_T, warping functions $\gamma_1,\ldots,\gamma_N$ can be interpreted as cumulative distribution functions of continuous random variables that take values in T. In such framework, their first derivatives $D(\gamma_1) = \gamma_1',\ldots,D(\gamma_N) = \gamma_N'$, where D is the differential operator, are then the corresponding probability density functions. The Hilbert Bayes space $\mathcal{B}^2(T)$ is the vector space formed by the equivalence classes of such functions, and both a definition of norm as well as operations are defined in $\mathcal{B}^2(T)$ (see Happ et al. (2019) and references within). It is possible to transform functions $\gamma_1',\ldots,\gamma_N' \in \mathcal{B}^2(T)$ by means of the centred log-ratio transformation $\psi_{\mathcal{B}} : \mathcal{B}^2(T) \to S^2_{\mathcal{B}}(T)$, bijective isometric isomorphism defined as

$$\psi_{\mathcal{B}}(\gamma')(t) = \log(\gamma'(t)) - \int_T \log(\gamma'(s))ds \tag{4.2.4}$$

where $S^2_{\mathcal{B}}(T) = \{g \in L^2(T) : \int_T g(s)ds = 0\}$. The corresponding inverse transformation is then

$$\psi_{\mathcal{B}}^{-1}(g)(t) = \frac{\exp(g(t))}{\int_T \exp(g(s))ds}. \tag{4.2.5}$$

As recommended by Happ et al. (2019), we choose (4.2.3) as

$$\psi = \psi_{\mathcal{B}} \circ D, \tag{4.2.6}$$

namely a composition of bijective maps.

Having now established a suitable transformation for the phase component that allows to easily compute mean and covariance functions, we come back to the analysis of the curves arising from phase-amplitude separation. As we mentioned earlier, it is possible to run FPCA separately on amplitude and adequately transformed phase components. However, such approach does not take into account the sources of simultaneous variation that amplitude and phase representatives naturally inherited from the fact that they are generated from the same functional observations. A joint FPCA approach was proposed by Lee and Jung (2016), who in the same paper suggested that the use of MFPCA might be more appropriate to study the joint variation of phase and amplitude representatives.

Such analysis was studied and implemented by Happ et al. (2019), who relied on the findings on MFPCA presented by Happ and Greven (2018). Specifically, consider $\tilde{f}_1, \dots, \tilde{f}_N$ and $\gamma_1, \dots, \gamma_N$, the phase and amplitude components arising from the registration of $f_1, \dots, f_N$ respectively. Happ et al. (2019) worked with bivariate functional objects $h_1, \dots, h_N \in \mathcal{H} = \mathcal{F}(T) \times S^2_{\mathcal{B}}(T)$, where $h_i \in \mathcal{H}$, with $i = 1, \dots, N$, is such that

$$h_i(t) = (\tilde{f}_i(t), \psi(\gamma_i)(t)),$$

with phase transformation $\psi(\cdot)$ as in (4.2.6). Taking for instance $h_1, h_2 \in \mathcal{H}$, the inner product we endow $\mathcal{H}$ with is

$$\langle\langle h_1, h_2 \rangle\rangle = \langle \tilde{f}_1, \tilde{f}_2 \rangle_2 + \langle \psi(\gamma_1), \psi(\gamma_2) \rangle_2, \tag{4.2.7}$$

where $\langle \cdot, \cdot \rangle_2$ is the inner product of L^2. It is also possible to weigh the two L^2 inner products in (4.2.7), especially if the variability of the two components is very different (Happ and Greven, 2018; Happ et al., 2019). In order to proceed with MFPCA analysis, one first has to find equivalence between the mean and variance in $\mathcal{H}$ and those in $\mathcal{G} = \mathcal{F}(T) \times \Gamma_T$, namely the space of the bivariate objects where the phase component has not been transformed. Happ et al. (2019) suggested a valid metric to endow $\mathcal{G}$ with, and this allows to establish equivalence results between mean and variance notions in $\mathcal{H}$ and $\mathcal{G}$. In such way, no information is lost in the transformation of the phase components, and that one can hence work in $\mathcal{H}$, which has a geometric structure that can be handled with the tools of functional data analysis.

It is possible now to perform MFPCA on a collection of bivariate functional objects $h_1, \dots, h_N \in \mathcal{H}$, with $h_i = (\tilde{f}_i, \psi(\gamma_i))$, $i = 1, \dots, N$. The work developed by Happ and Greven (2018) is based on theoretical results that allow to find en equivalence between MFPCA of $h_1, \dots, h_N$ and separate FPCA of $\tilde{f}_1, \dots, \tilde{f}_N$ and $\psi(\gamma_1), \dots, \psi(\gamma_N)$. In particular, they presented an estimation method based on merging the scores arising from the univariate FPCA of the components of the multivariate objects. The resulting quantities from MFPCA of $h_1, \dots, h_N$ are the eigenvalues $\nu_1 \geq \nu_2 \geq \dots \geq 0$ and the multivariate eigenfunctions $\varphi_1, \varphi_2, \dots \in \mathcal{H}$, with $\varphi_j = (\varphi_j^A, \varphi_j^P)$, $j = 1, 2, \dots$, where the superscripts indicate either the amplitude (A) or the phase (P) univariate components respectively. Scores are computed as

$$\xi_{i,j} = \langle\langle h_i, \varphi_j \rangle\rangle \qquad i = 1, \dots, N,\ j = 1, 2, \dots \tag{4.2.8}$$

In practice, the total number of principal components is truncated at a number K large enough such that most of the sources of variation are included in the analysis while

discarding the possible measurement noise. A common way of choosing K is by imposing that the Percentage of Variance Explained (PVE) $\sum_{j=1}^{K} \nu_j / \sum_{j=1}^{\infty} \nu_j$ is as high as 0.95 or 0.99.

In order to interpret the results, it is possible to transform the findings back to the space of the original functions $f_1, \ldots, f_N$. For instance, take the estimated average of the phase components $\overline{\gamma} = \psi^{-1}(\overline{\psi(\gamma)})$, where $\overline{\psi(\gamma)}$ is the point-wise mean of $\psi(\gamma_1), \ldots, \psi(\gamma_N)$, and the estimated point-wise mean of the amplitude components, namely $\bar{\tilde{f}}$. One can then compute the estimated mean of the original curves as

$$\overline{f} = \bar{\tilde{f}} \circ \overline{\gamma}^{-1}. \tag{4.2.9}$$

In a similar fashion, it is also possible to compute the perturbations of the mean in terms of amplitude and phase variations characterizing a selected principal component. For instance, consider the jth eigenvalue ν_j and eigenfunction φ_j and define $\gamma_{V_j} = \psi^{-1}(\overline{\psi(\gamma)} + C\sqrt{\nu_j}\varphi_j^P)$ and $\tilde{f}_{V_j} = \bar{\tilde{f}} + C\sqrt{\nu_j}\varphi_j^A$, where $C \in \{-1, 1\}$. Such elements define the perturbation of the mean amplitude component and the perturbation of the mean phase component respectively. We can then consider deforming $\tilde{f}_{V_j}$ with $\gamma_{V_j}^{-1}$, accounting for both amplitude and phase sources of variation, namely

$$f_{DV_j} = \tilde{f}_{V_j} \circ \gamma_{V_j}^{-1}, \tag{4.2.10}$$

where the subscript indicates the double source of variation. Function f_{DV_j} can be then directly compared to $f_1, \ldots, f_N$. In a similar fashion, one can also consider the perturbations of the mean function with only one of the two sources of variation. For the amplitude variation we deform $\tilde{f}_{V_j}$ with $\overline{\gamma}^{-1}$:

$$f_{AV_j} = \tilde{f}_{V_j} \circ \overline{\gamma}^{-1}, \tag{4.2.11}$$

while for the phase variation we deform $\bar{\tilde{f}}$ with $\gamma_{V_j}^{-1}$:

$$f_{PV_j} = \bar{\tilde{f}} \circ \gamma_{V_j}^{-1}. \tag{4.2.12}$$

Notice how, for $C = 0$, the variations (4.2.10), (4.2.11) and (4.2.12) all correspond to mean (4.2.9). In Section 4.3.2 we let C take values in $[-1, 1]$ so as to illustrate how the mean function reaches its perturbations while the standard deviation gradually increases.

4.3 Analysis of learning curves

As mentioned in Section 4.1, the motivating application of this work regards the consequences of memory loss in mice (Benoit et al., 2020). In the experimental set-up animals were first divided into a control group and a group in whose brain was induced a lesion that should cause a similar memory loss as the one witnessed in patients affected by psychiatric disorders such as schizophrenia. Our aim is to compare the behavior of 17 lesioned mice and 16 mice injected with placebo in the same memory-involving task. Specifically, after an acquisition phase when they had been trained on how to successfully complete the task, both groups of mice undertook several sequential trials a day of a test that required them to remember a previous action for different periods of time. In particular, to each trial a delay of time of either 2, 4, 8 or 16 seconds was assigned randomly. During the acquisition, mice were train for 18 or 19 consecutive days, with a

number of trials per day that went from 70 to 158, and an overall median of 118 trials. Then, when delays were introduced, both groups of mice were tested with 107 to 160 daily trials, with median 155, over either 5 or 8 consecutive days. During acquisition around 30 seconds would pass between the beginning of the two trials, while when delays were introduced that time span would be slightly longer. The results of trials were deemed inconclusive, and hence removed from the dataset, if mice were inactive for some time during the task.

4.3.1 Preprocessing and registration

First of all, we converted the collection of successes and failures of the trials into smooth learning curves, so that the functions take values corresponding to the probabilities of success over the experiment. We achieved that by means of logistic regression, taking for each subject the binary outcomes as responses and the time of experiment as smooth predictor approximated by means of penalised cubic splines. Computationally, we relied on function `gam()` available in the R-package `mgcv` (Wood et al., 2016) and we used six penalised cubic splines to represent a smooth effect in time. Since our ultimate goal is the comparison of the curves from the two groups in the different scenarios, we considered the binary outcomes as if recorded on a regular grid of points, and then smoothed them to take values on the same dense grid, constituted of 100 points distributed over T. Time interval T was chosen for interpretation purposes, so as to its extremes correspond to the start and the end of the overall experiment, which was conducted over several days.

Figure 4.1 shows the smoothed learning curves of the 33 animals in both the acquisition and test scenarios. The latter are displayed depending on the length of delay introduced in the test. We take probability 0.5 (horizontal black line) as reference of mice acting out of chance. Firstly, notice that the curves are not always nondecreasing. This might be due to the fact that the learning processes were recorded over several days, so there is a chance that the animals needed some adaptation after some time off of the experiment. However, in general the trend of these curves is increasing over time. Secondly, it is possible to notice some differences in the learning curves of the acquisition and of the test parts. In the former, there is no striking difference between the curves of the two groups, both in terms of variation as well as of learning achievements at the end of the experiment. On the other hand, when it comes to the test scenarios, one can see that the behaviors of the two groups differs for different delay lengths. In particular, when it comes to the 2 and 4 seconds delays, there seems to be a larger variation in the learning curves of the lesioned mice, who achieve rather different results by the end of the experiment. Mice in the control group seem to reach final high learning probabilities, ranging from around 0.7 to 1 for both types of delay. Regarding the test with 8 and 16 seconds delays, there is no evident distinction between the two groups of functions. Even though the final learning probabilities are generally lower than in the cases of shorter delay, the amount of variation for the two groups of mice is comparable in both pictures. In particular, notice that for the 16 seconds delays the learning curves belonging to both groups are somewhat flatter than in the other cases, as well as closer to 0.5 for the first half of the experiment. This might indicate the fact that 16 seconds is too long a delay for the animals to remember something, regardless of the group they belong to.

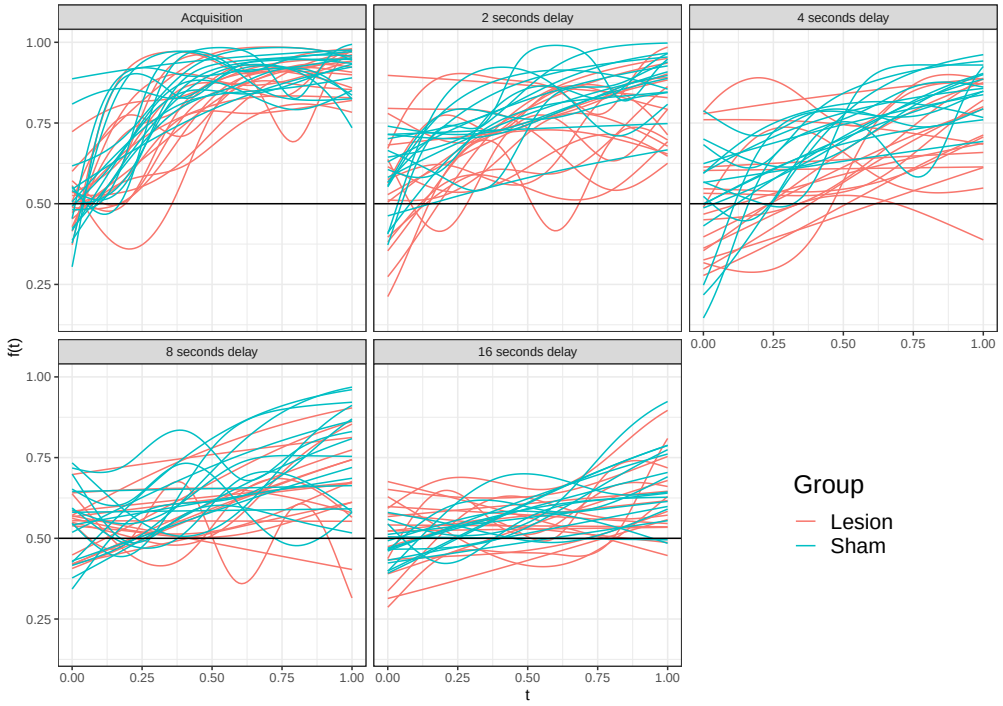

Figure 4.1: Learning curves of lesioned (orange) and sham (green) mice groups for the acquisition and test parts of the experiment. For the latter, we report the learning curves related to different delays introduced in the task, either 2, 4, 8 or 16 seconds long.

In light of the results in Figure 4.1, in the following analysis we decide to study three scenarios: the acquisition, and the test with 2 and 16 seconds delays. The reason for this choice is that curves related to the 4 and 8 seconds delays resemble those arising from the 2 and 16 seconds delays respectively. Moreover, we consider the chosen scenarios as representatives of three different activities the mice go through, namely learning how to successfully complete a trial, and remembering what they shall do both after a short and after a rather long time lag respectively.

In the following sections we rely on the results from the phase-amplitude separation described in Section 4.2.1. We used function `time_warping()` function of package `fdasrvf` (Tucker, 2020) with default setting, except for `MaxItr=500`, the parameter controlling the maximum iterations of the algorithm. Moreover, function `MFPCA()` is available in the R-package `MFPCA` (Happ-Kurz, 2020) to carry out the MFPCA procedure from Happ and Greven (2018).

4.3.2 Perturbation of the mean by amplitude and phase components on separate data

As mentioned in Section 4.2, the advantage of analysing the amplitude and phase components of a collection of curves with MFPCA is to gain knowledge on the main sources of variation in the data, while accounting for the simultaneous variation between the two components. In this section we consider the results from the separate registration of the learning curves of lesioned mice and mice injected with placebo in the three scenarios

we are interested in, namely acquisition, test with 2 seconds delay and with 16 seconds delay. This corresponds to performing phase-amplitude separation on six collections of curves, namely two groups in three scenarios, and then carrying out MFPCA on each one of them using the corresponding amplitude and phase components. In all of these cases we used $K = 15$ principal components in total. In Table 4.1 we show the eigenvalues ν_1, ν_2, ν_3 corresponding to the first three principal components, as well as the corresponding percentages of variance explained, computed as

$$PVE_j = \frac{\nu_j}{\sum_{j=1}^{K} \nu_j}, \qquad j = 1, 2, 3.$$

The first principal component explains a high percentage of the data in all three scenarios and for both groups of mice. The highest percentages of variance explained are those corresponding to the MFPCA carried out on the 16 seconds delay data, probably due to a simpler structure of the curves.

		PVE_1	PVE_2	PVE_3	ν_1	ν_2	ν_3
Acquisition	Sham	0.51	0.40	0.06	0.53	0.41	0.06
	Lesion	0.58	0.25	0.11	0.46	0.19	0.9
2 seconds delay	Sham	0.44	0.33	0.17	0.19	0.14	0.07
	Lesion	0.61	0.26	0.09	0.90	0.39	0.14
16 seconds delay	Sham	0.63	0.24	0.07	0.37	0.15	0.04
	Lesion	0.89	0.10	0.01	2.38	0.27	0.03

Table 4.1: Percentage of variance explained and eigenvalues by the first three principal components for both groups of animals in the scenarios considered.

For each of the three scenarios, we are interested in comparing the two groups of mice when it comes to the variations of the mean in both amplitude and phase, only amplitude and only phase, as in equations (4.2.10), (4.2.11) and (4.2.12) respectively. As mentioned in Section 4.2, for all of the three types of variation we take $C \in [-1, 1]$ (the extremes of the interval are coloured in blue and red respectively), where $C = 0$ (in grey) corresponds to the mean function (4.2.9). By showing all the three types of variations we aim at understanding how phase and amplitude play a role in the overall perturbations of the mean. Since the first principal component explains at least half of the overall variation in almost all the scenarios as shown in Table 4.1, we only show the results concerning the first principal component in this section, while the ones regarding the second are in the appendix.

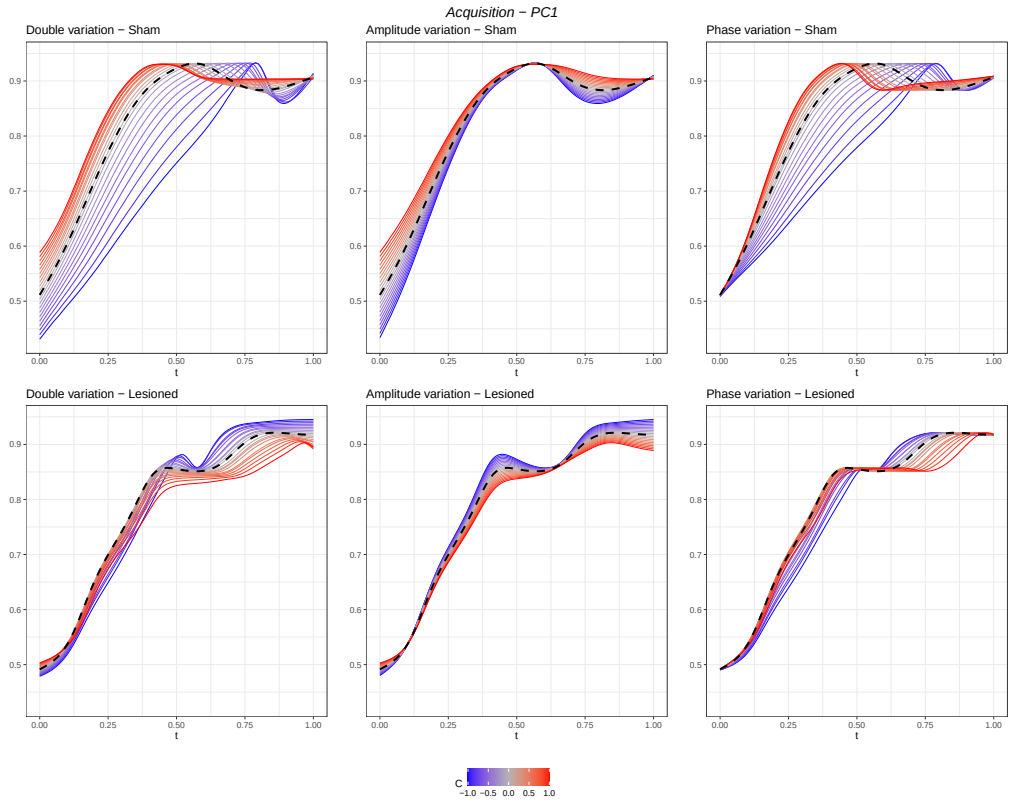

Figure 4.2: Double variation (4.2.10) (first column), amplitude variation (4.2.11) (second column) and phase variation (4.2.12) (third column) relative to the first principal component of MFPCA on sham and lesioned mice acquisition data (first and second row respectively). The mean functions, corresponding to $C = 0$, are marked with a black dashed line.

In Figure 4.2 we show the results for the first principal component of the acquisition data for both groups of animals. We can first of all notice that for two groups of animals the mean function is rather similar, but the variability in amplitude and phase are different. Mice in the sham group have more variability in amplitude in the beginning of the experiment, while lesioned mice show larger amplitude variability towards the end of the experiment. This might indicate that the animals in the sham group manage to reach high scores no matter how well they start the experiment, while mice in the lesioned group actually achieve different results at the end of the experiment. There is also a difference in the phase variation: for animals in the sham group there is variability throughout the experiment, while for lesioned it is from around 0.6 until the end. In the overall variation, we can see that sham mice either start the learning curve at a probability of success larger than 0.5 and plateau at high scores from the middle of the experiment (in red) or their initial score is lower than 0.5 and they reach their final score later (in blues). On the other hand, for lesioned mice there is very low variability both in phase and amplitude in the first half of the experiment, and then they either reach a higher score faster than the mean (in blue) or a lower score more slowly than the mean (in red).

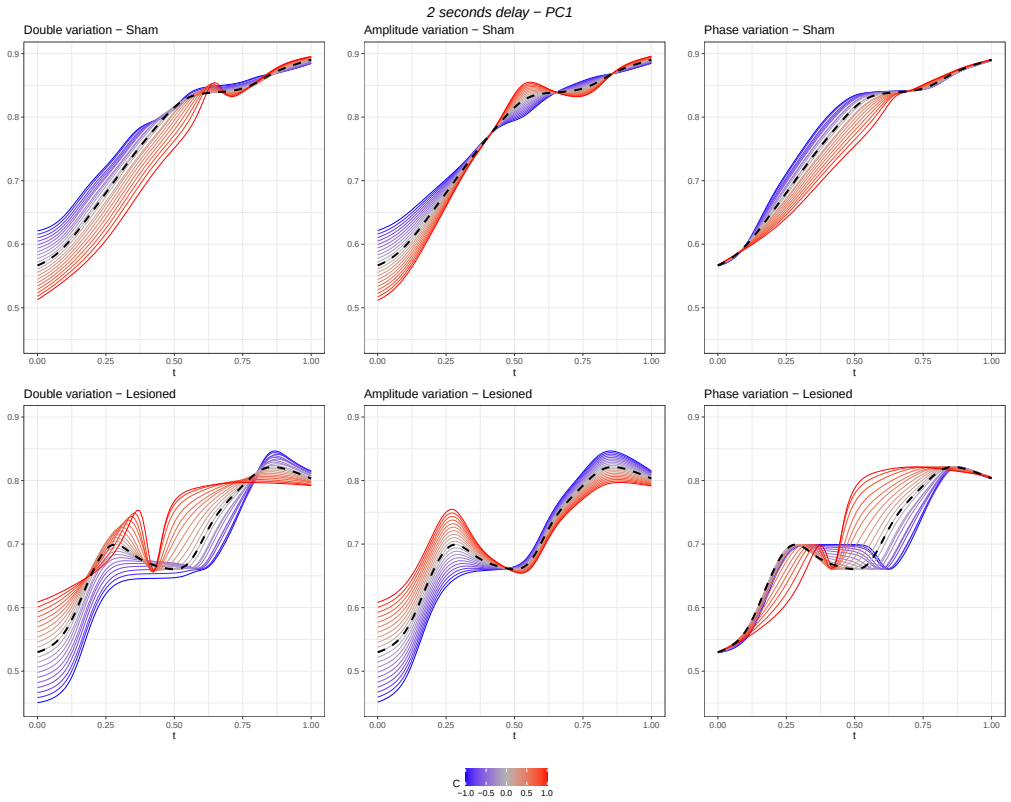

Figure 4.3: Double variation (4.2.10) (first column), amplitude variation (4.2.11)(second column) and phase variation (4.2.12) (third column) relative to the first PC of MFPCA on sham and lesioned mice 2 seconds delay data (first and second row respectively). The mean functions, corresponding to $C = 0$, are marked with a black dashed line.

The results for delays of short length are shown in Figure 4.3. The mean function of the sham group is smooth and it reaches 0.9, while the one for the lesioned mice in piecewise increasing, reaching 0.8. For the mice in the sham group there is larger variation both in amplitude and phase in the beginning of the experiment. This brings conclusions similar to those drawn for the acquisition in the double variation case. On the other hand, for the lesioned mice there is variability in amplitude especially in the first part of the experiment, while there is more variability in phase in the second half. This results in profiles of the overall variation that either progress more slowly and then reach a higher probability of success (in blue) or that progress seemingly faster, then have an abrupt drop in the performance in the middle of the experiment, then get back on track and stabilize a bit below 0.8 (in red).

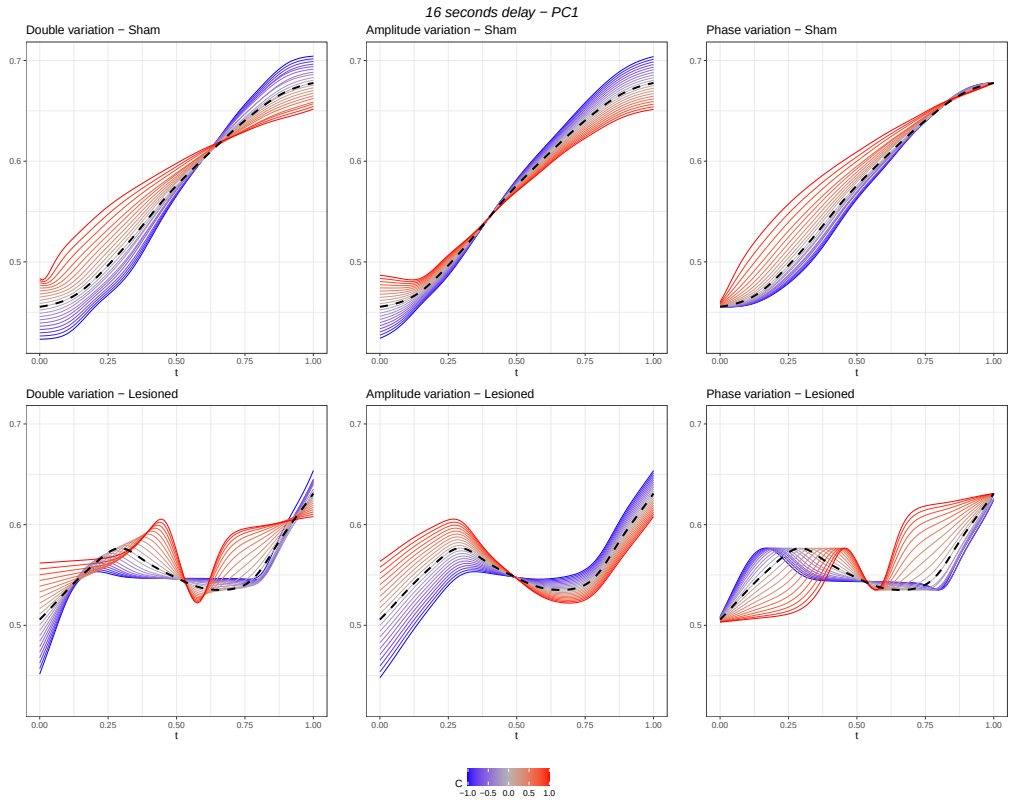

Figure 4.4: Double variation (4.2.10) (first column), amplitude variation (4.2.11) (second column) and phase variation (4.2.12) (third column) relative to the first PC of MFPCA on sham and lesioned mice 16 seconds delay data (first and second row respectively). The mean functions, corresponding to $C = 0$, are marked with a black dashed line.

When it comes to 16 seconds delay, the results are reported in Figure 4.4. On average, the final learning achievements at the end of the experiment of both groups are lower than for short delays, being a bit above and below 0.65 for the sham and lesioned groups respectively. For the mice in the sham group the variation in phase is mostly in the beginning of the experiment, while for the amplitude it is either in the very beginning or in the very end of it. The double variation plot shows two behaviors: mice either start at a lower probability of success and they slowly reach the end the experiment at higher results (blue), or they start their learning curve at higher probabilities and end up faster at poorer results. For lesioned mice the amplitude variation is predominantly at the extremes of the time interval too, but the phase variation is in the first and second half of the experiment. The results for the double variation are somewhat similar to those related to short delays: one profile is piecewise increasing, starting from a lower probability of success and reaching higher results than the mean function (in blue), and the other starts from higher probability of success and ends in poorer results than the mean function, with a drop in the middle of the experiment (in red).

We point out that, in the analysis carried out in this section, the sizes of the variations are not directly comparable between the two groups since they are the result of two separate

phase-amplitude separations and MFPCAs. However, it is still possible investigate in what way the behavior of the two groups differs. From this analysis we see that, while for the acquisition the mean function is comparable between the two groups, in those scenarios with delays the average learning curve for mice in the sham group is generally smooth, while the one for the lesioned animals is piecewise increasing, with a general flat trend in the middle of the experiment. This might indicate that lesioned mice seem to have problems at remembering the task from one session to the other, and they need some re-adjustment period before starting to improve again, while the sham group is better at remembering from one day to the other. Moreover, the modes of overall variation of acquisition and test with 2 seconds delay for mice in the sham group are more similar than for lesioned mice. A possible conclusion is that even short time delays are more challenging for animals on the lesioned group rather than in the sham group. Finally, even though it seems like the learning behaviors of the two groups are still different, in the test with 16 seconds delay both groups of mice reach on average comparable results at the end of the experiment.

4.3.3 Analysis of scores on joint data

With the analysis in the previous section we showed that there are some differences in the way mice in the sham and lesioned groups behave, and the conclusions are based on results arising from separate registrations and MFPCAs. In this section, for each of the three scenarios, we align the learning curves together, regardless if they are from the sham and the lesioned group of mice, and then we perform the MFPCA on the resulting amplitude and phase components. Notice that the collections of curves we consider are not independent, since we are ignoring the underlying grouped structure of the samples. This entails that we do not carry out a proper principal component analysis, and that the resulting sources of variations should not be addressed as eigenfunctions. However, the analysis is still useful to extract basis functions common to the two groups. With an abuse of notation, we refer to the elements resulting from the analysis withe the same names of those arising from MFPCA. We ran k-means algorithm (Hartigan and Wong, 1979) on the scores, computed as in (4.2.8), of the two first principal components for each scenario, in order to find four clusters for each one of them. Again, we chose $K = 15$ principal components of the MFPCA.

In the first column of Figure 4.5 we display the scores with the different colours denoting the four different clusters. In Table 4.2 we report the percentage of variance explained by the first two principal components, whose sum is bigger than 75% in all scenarios. Moreover, even though we processed the learning curves jointly, we distinguish the scores by treatment group by using different point shapes. There does not seem to be distinct pattern in the composition of the clusters in terms of observations from either the two groups of mice. This might indicate that the behavior of the animals cannot be solely explained by the treatment they underwent.

In order to further look into the difference between the clusters detected, in Figure 4.5 we also plot the warping functions (central column) as well as the original learning curves (last column) coloured according to the cluster they belong to. First of all, one needs to bear in mind that the clusters are not directly comparable across the three scenarios, since we performed three separate MFPCAs. Moreover, remember that the scores account for a combination of amplitude and phase characteristics, so detecting specific cluster trend for either solely amplitude or phase might be challenging. Nevertheless, it is possible to see some interesting patterns. As far as the acquisition is concerned, the clusters of

scores are quite distinct. This results in particular in four different trends of the warping functions: warping functions above (in light green) and below (in dark green) the identity line, corresponding to learning curves that reach the plateau after and before the template respectively, as well as warping functions below the identity line in the first half of the experiment and then above in the second one (in brown), and vice versa (in yellow). On the other hand, apart from the fact that the the light green curves seem to be the smoothest, the differences between the learning curves of different clusters is not as clear. When it comes to the task with 2 seconds delay, the negative and positive scores of the first principal component are especially distinct. It is possible to notice that these two groups correspond to smooth (in dark purple and light cyan) and wiggly (in light purple and dark cyan) learning curves respectively, while it is more difficult to interpret the differences among the warping functions belonging to different clusters. A similar interpretation can be given to the scores and learning curves arising from the 16 seconds delay task, where the beige and light blue curves are somewhat more wiggly than the dark blue and chocolate ones.

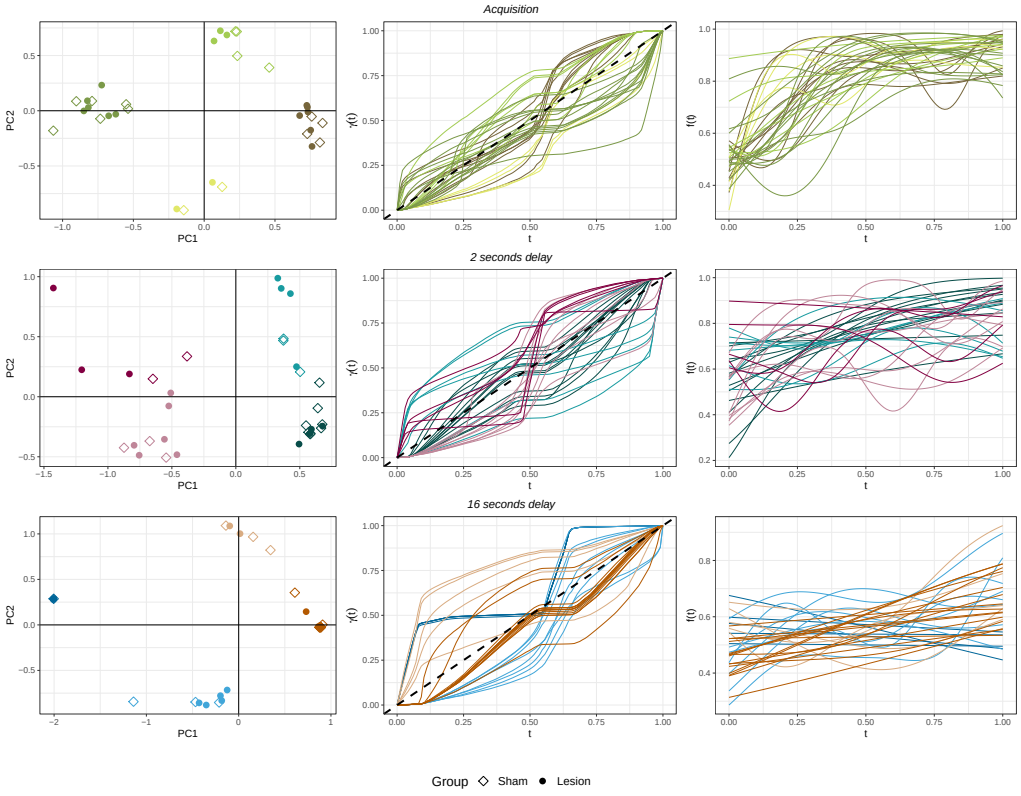

Figure 4.5: Scores from MFPCA (first column), warping functions (second column) and original learning curves (third column) coloured according to the different score clusters. Each row contains results related to one of the three scenarios, namely acquisition (first row), test with 2 seconds delay (second row) and test with 16 seconds delay (thirs row). The scores show the group the specific animal belongs to (Sham or Lesioned) via different point shapes.

	PVE_1	PVE_2
Acquisition	0.58	0.25
2 seconds delay	0.54	0.24
16 seconds delay	0.70	0.23

Table 4.2: Percentage of variance explained by the first and the second principal components (first and second columns respectively) for the different scenarios.

In summary, the analysis of the clusters of scores suggests that the two groups of animals share some similar behaviors, which are different across different scenarios. In particular, different clusters detect especially well different learning speed trends in the acquisition. On the other hand, for the tests with delays the most noticeable difference among clusters is the smoothness of the corresponding learning curves. This can can be interpreted as the difference between mice that can easily remember from the previous day of the experiment and those that need to re-adjust after some time off performing the task.

4.4 Discussion

This work is based on the framework discussed by Benoit et al. (2020). We compared the performance of two groups of mice, namely a control group and a group of animals with brain lesion, when it came to performing a memory-involving task repeatedly. In particular, we considered three scenarios: the acquisition, when the animals learn how to do the task, the test when short delays are introduced and the test when longer delays are introduced. We relied on the estimation methods described by Happ et al. (2019). Our findings in Section 4.3.2 indicate that, when the learning curves of the two groups are registered and analyzed with MFPCA separately, on average they differ less in the acquisition, reaching high probabilities of success in both cases, rather than when the delays are introduced. In particular, for the latter cases, even though the overall learning results are comparable at the end of the experiment for the 16 seconds delay, the mice in the sham group seem to have a smooth average learning curve, while it is piece-wise linear for lesioned mice. This might indicate that, even though the two groups of mice can learn equally well how to perform the task, once the short delays are introduced the lesioned mice need some re-adjustment periods in order to improve, while the mice in the sham group seem to remember better how to complete the task successfully. This is also the case for longer delays. Looking at the variations brought by the first principal component, it is interesting to notice how mice in the control group seem to reach equally high success probabilities with low overall variation in the acquisition and 2 seconds delay scenarios, no matter how much variability in amplitude and phase they exhibit in the first half of the experiment. This is not the case for lesioned mice, which show different behaviors in the acquisition and when short delays are introduced.

On the other hand, when registering and carrying out MFPCA on the joint collections of learning curves without accounting for the groups as in Section 4.3.3, we showed that the clustering of the scores resulted in the identification of behaviors that are shared across the two groups, rather than clusters of animals belonging to the same group. In particular, in the acquisition different learning speeds were identified, while in the scenarios where delays were introduced the smooth learning curves were separated from the wiggly ones.

Overall, our work suggests that the available data is rich and complex, and further analysis might be needed to model the learning behaviors of the two groups of mice. For instance, since Benoit et al. (2020) found out that the scenario in which there is the most significant difference between the two groups of mice is the one with 4 seconds delay, our plan is to include such scenario in our analysis. Moreover, it would be interesting to compare the results in Section 4.3.3 with those arising from separate FPCAs of amplitude and phase components, so as to check whether with those analysis clusters of scores corresponding only to observations from either sham or lesioned groups can be detected.

Other statistical tools could be employed as well. For instance, one could combine the method for cluster detection of multivariate objects proposed by Schmutz et al. (2020) with phase-amplitude separation. Another possible direction could be the development of tests aimed at comparing univariate amplitude and phase components of the two groups of mice, as well as the multivariate objects we considered for MFPCA.

Finally, when it comes to the pre-processing of data, we relied on smoothing techniques, somewhat similarly to the work of Wu and Srivastava (2014), who smoothed spike train data prior to performing registration. More recently, Wrobel et al. (2019) presented a registration approach that is likelihood-based, not requiring pre-smoothing of discrete data, and their data application consisted of sequences of binary data for every subject, similarly to what we handle in Section 4.3. It could be interesting to compare the results we obtained with those based on such phase-amplitude analysis, so as to verify how much information, if any, was lost in smoothing the discrete data.

4.5 Acknowledgements

The project was partly funded by the Danish Research Council (DFF grant 7014-00221).

4.6 Appendix

Further results of perturbations of the mean by amplitude and phase components on separate data

In this section we show the results related to the second principal component from the analysis carried out in Section 4.3.2.

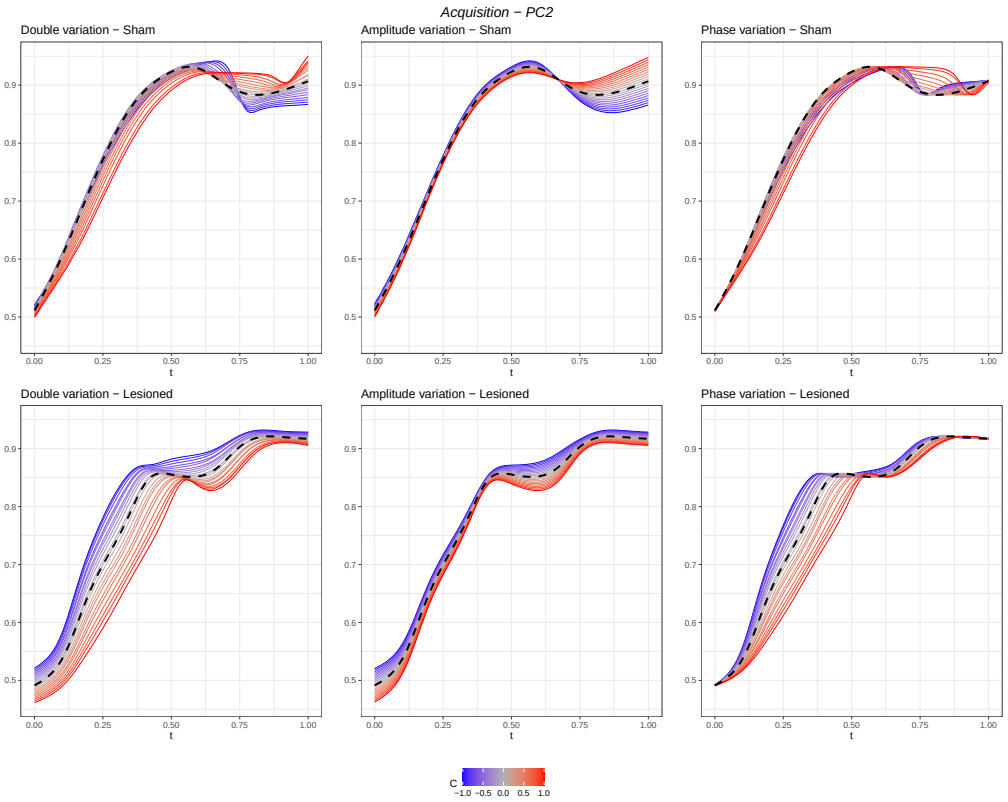

Figure 4.6: Double variation (4.2.10) (first column), amplitude variation (4.2.11) (second column) and phase variation (4.2.12) (third column) relative to the second PC of MFPCA on sham and lesioned mice acquisition data (first and second row respectively). The mean functions, corresponding to $C = 0$, are marked with a black dashed line.

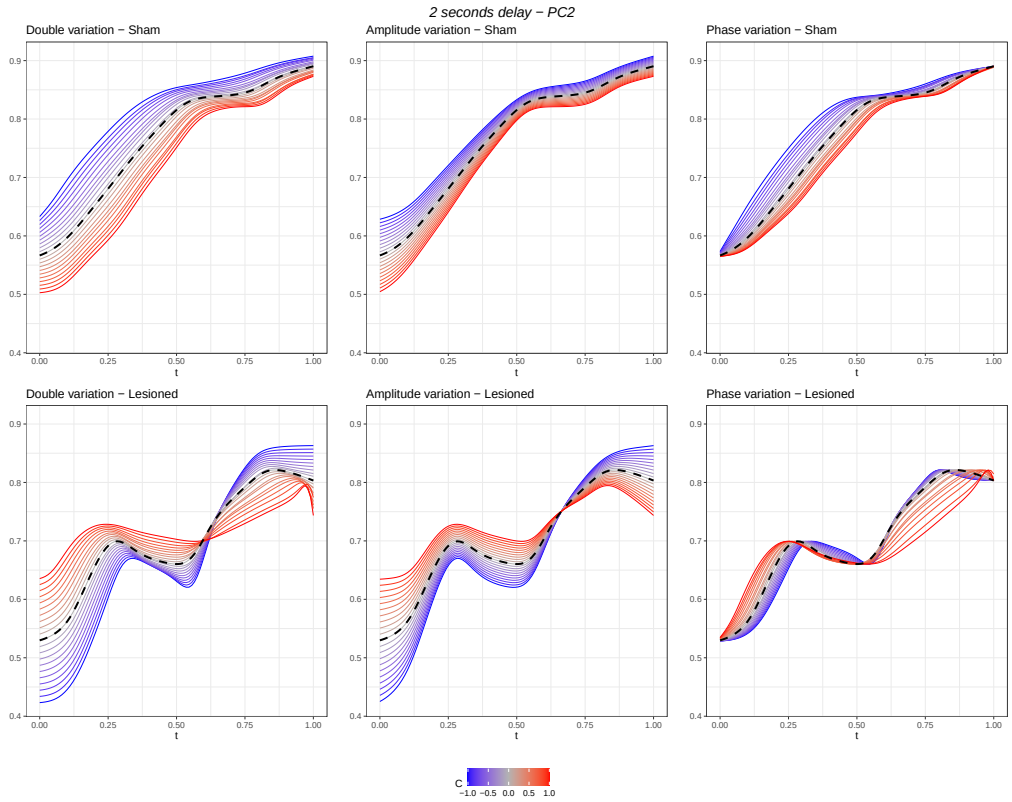

Figure 4.7: Double variation (4.2.10) (first column), amplitude variation (4.2.11) (second column) and phase variation (4.2.12) (third column) relative to the second PC of MFPCA on sham and lesioned mice 2 seconds delay data (first and second row respectively). The mean functions, corresponding to $C = 0$, are marked with a black dashed line.

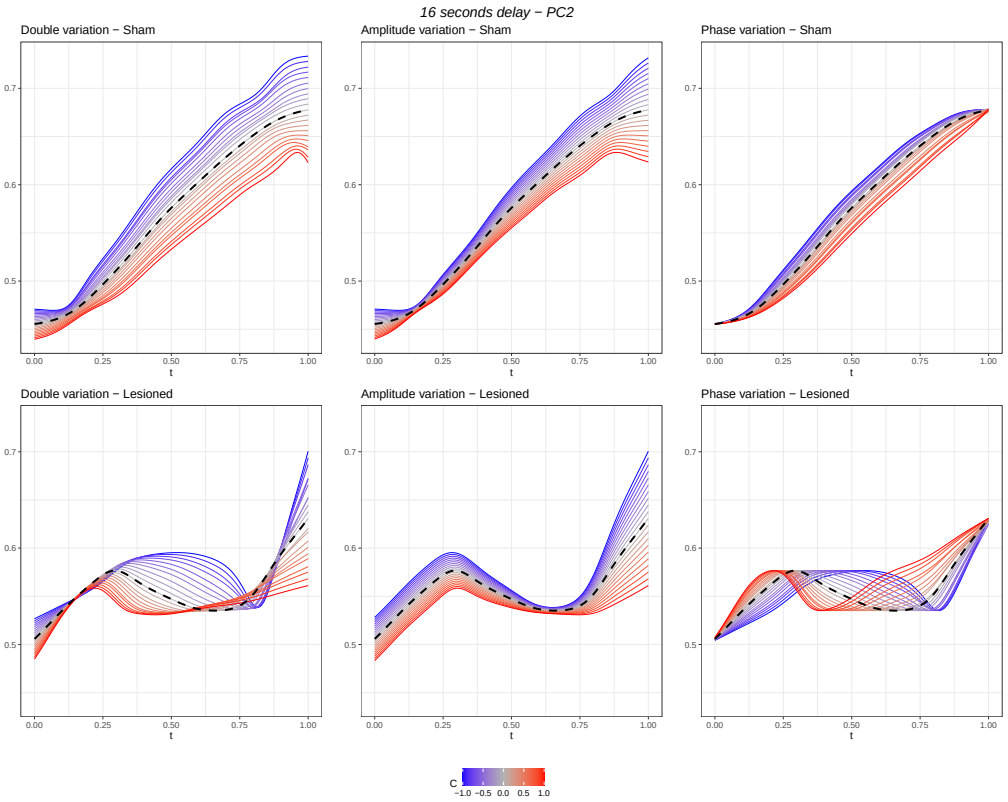

Figure 4.8: Double variation (4.2.10) (first column), amplitude variation (4.2.11) (second column) and phase variation (4.2.12) (third column) relative to the second PC of MFPCA on sham and lesioned mice 16 seconds delay data (first and second row respectively). The mean functions, corresponding to $C = 0$, are marked with a black dashed line.

Bibliography

Abramowicz, K., Häger, C., Pini, A., Schelin, L., Sjöstedt de Luna, S., and Vantini, S. (2018). Nonparametric inference for functional-on-scalar linear models applied to knee kinematic hop data after injury of the anterior cruciate ligament. *Scandinavian Journal of Statistics*, 45:1036–1061.

Abrevaya, J. and Dahl, C. M. (2008). The effects of birth inputs on birthweight. *Journal of Business & Economic Statistics*, 26(4):379–397.

Battagliola, M. L., Sørensen, H., Tolver, A., and Staicu, A.-M. (2021). A bias-adjusted estimator in quantile regression for clustered data. *Econometrics and Statistics*.

Bayer, S. (2018). Combining value-at-risk forecasts using penalized quantile regressions. *Econometrics and Statistics*, 8:56 – 77.

Benoit, L. J., Holt, E. S., Teboul, E., Taliaferro, J. P., Kellendonk, C., and Canetta, S. (2020). Medial prefrontal lesions impair performance in an operant delayed nonmatch to sample working memory task. *Behavioral neuroscience*.

Berrendero, J. R., Justel, A., and Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, 55(9):2619–2634.

Besstremyannaya, G. and Golovan, S. (2019). Reconsideration of a simple approach to quantile regression for panel data. *The Econometrics Journal*, 22(3):292–308.

Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):1103–1130.

Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838.

Bossoli, D. and Bottai, M. (2017). Marginal quantile regression for dependent data with a working odds-ratio matrix. *Biostatistics*, 19(4):529–545.

Brockhaus, S., Rügamer, D., and Greven, S. (2017). Boosting functional regression models with fdboost. *Journal of Statistical Software*, 94.

Bühlmann, P. and Hothorn, T. (2007). Boosting Algorithms: Regularization, Prediction and Model Fitting. *Statistical Science*, 22(4):477 – 505.

Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3):368–386.

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

Cardot, H., Ferraty, F., and Sarda, P. (1999a). Functional linear model. *Statistics & Probability Letters*, 45(1):11–22.

Cardot, H., Ferraty, F., and Sarda, P. (1999b). Functional linear model. *Statistics and Probability Letters*, 45:11–22.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, 13(3):571–591.

Carpenter, J. R., Goldstein, H., and Rasbash, J. (2003). A novel bootstrap procedure for assessing the relationship between class size and achievement. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):431–443.

Chen, K. and Müller, H.-G. (2012). Conditional quantile analysis when covariates are functions, with application to growth data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1):67–89.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, pages 1571–1596.

Davison, A. and Hinkley, D. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, New York.

Dette, H. and Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):609–627.

Dhaene, G. and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3):991–1030.

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2009). Multilevel functional principal component analysis. *The Annals of Applied Statistics*, 3(1):458 – 488.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Egozcue, J. J., Díaz-Barrero, J. L., and Pawlowsky-Glahn, V. (2006). Hilbert space of probability density functions based on aitchison geometry. *Acta Mathematica Sinica*, 22(4):1175–1182.

Fasiolo, M., Wood, S. N., Zaffran, M., Nedellec, R., and Goude, Y. (2020). Fast calibrated additive quantile regression. *Journal of the American Statistical Association*, 0(0):1–11.

Feng, X., He, X., and Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*, 98(4):995–999.

Fenske, N., Fahrmeir, L., Hothorn, T., Rzehak, P., and Höhle, M. (2013). Boosting structured additive quantile regression for longitudinal childhood obesity data. *The International Journal of Biostatistics*, 9(1):1–18.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). *Applied longitudinal analysis*, volume 998. John Wiley & Sons.

Galarza, C. E., Lachos, V. H., and Bandyopadhyay, D. (2017). Quantile regression in linear mixed models: a stochastic approximation em approach. *Statistics and its Interface*, 10(3):471.

Galvao, A. and Montes-Rojas, G. (2015). On bootstrap inference for quantile regression panel data: A Monte Carlo study. *Econometrics*, 3(3):654–666.

Galvao, A. F., Juhl, T., Montes-Rojas, G., and Olmo, J. (2017). Testing slope homogeneity in quantile regression panel data with an application to the cross-section of stock returns. *Journal of Financial Econometrics*, 16(2):211–243.

Galvao, A. F. and Kato, K. (2016). Smoothed quantile regression for panel data. *Journal of Econometrics*, 193(1):92–112.

Galvao, A. F. and Kato, K. (2017). Quantile regression methods for longitudinal data. In Koenker, R., Chernozhukov, V., He, X., and Peng, L., editors, *Handbook of Quantile Regression*, pages 363–380. Chapman and Hall/CRC.

Galvao, A. F. and Wang, L. (2015). Efficient minimum distance estimator for quantile regression fixed effects panel data. *Journal of Multivariate Analysis*, 133:1–26.

Geraci, M. (2014). Linear quantile mixed models: The lqmm package for Laplace quantile regression. *Journal of Statistical Software*, 57(13):1–29.

Geraci, M. (2019). Additive quantile regression for clustered data with an application to children's physical activity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(4):1071–1089.

Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, 8(1):140–154.

Geraci, M. and Bottai, M. (2014). Linear quantile mixed models. *Statistics and Computing*, 24(3):461–479.

Gervini, D. and Gasser, T. (2004). Self-modelling warping functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):959–971.

Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011a). Penalized functional regression. *Journal of computational and graphical statistics*, 20(4):830–851.

Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011b). Penalized functional regression. *Journal of Computational and Graphical Statistics*, 20:830–851.

Goldsmith, J., Crainiceanu, C., Caffo, B., and Reich, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society, Series C*, 61(3):453–469.

Goldsmith, J., Greven, S., and Crainiceanu, C. (2013). Corrected confidence bands for functional data using principal components. *Biometrics*, 69(1):41–51.

Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Di, C., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C., and Reiss, P. T. (2019). *refund: Regression with Functional Data*. R package version 0.1-21.

Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2010). Longitudinal functional principal component analysis. *Electronic Journal of Statistics*, 4:1022 – 1054.

Greven, S. and Scheipl, F. (2017). A general framework for functional regression modelling. *Statistical Modelling*, 17(1-2):1–35.

Gu, J. and Volgushev, S. (2019). Panel data quantile regression with grouped fixed effects. *Journal of Econometrics*, 213(1):68–91.

Hagemann, A. (2017). Cluster-robust bootstrap inference in quantile regression models. *Journal of the American Statistical Association*, 112(517):446–456.

Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113:649–659.

Happ, C., Scheipl, F., Gabriel, A., and Greven, S. (2019). A general framework for multivariate functional principal component analysis of amplitude and phase variation. *Stat*, 8:e220.

Happ-Kurz, C. (2020). *MFPCA: Multivariate Functional Principal Component Analysis for Data Observed on Different Dimensional Domains*. R package version 1.3-6.

Harding, M. and Lamarche, C. (2017). Penalized quantile regression with semiparametric correlated effects: An application with heterogeneous preferences. *Journal of Applied Econometrics*, 32(2):342–358.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.

Henry, K., Erice, A., Tierney, C., Balfour, H., Fischl, M., Kmack, A., Liou, S., Kenton, A., Hirsch, M., Phair, J., Martinez, A., and Kahn, J. (1998). A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced aids. aids clinical trial group 193a study team. *Journal of acquired immune deficiency syndromes and human retrovirology : official publication of the International Retrovirology Association*, 19(4):339—349.

Hron, K., Menafoglio, A., Templ, M., Hrůzová, K., and Filzmoser, P. (2016). Simplicial principal component analysis for density functions in bayes spaces. *Computational Statistics & Data Analysis*, 94:330–350.

Huang, J. Z., Shen, H., Buja, A., et al. (2008). Functional principal components analysis via penalized rank one approximation. *Electronic Journal of Statistics*, 2:678–695.

Huang, Y. and Chen, J. (2016). Bayesian quantile regression-based nonlinear mixed-effects joint models for time-to-event and longitudinal data with multiple features. *Statistics in Medicine*, 35(30):5666–5685.

Jacques, J. and Preda, C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis*, 71:92–106.

Karlsson, A. (2009). Bootstrap methods for bias correction and confidence interval estimation for nonlinear quantile regression of longitudinal data. *Journal of Statistical Computation and Simulation*, 79(10):1205–1218.

Kato, K. (2012). Estimation in functional linear quantile regression. *The Annals of Statistics*, 40(6):3108–3136.

Kato, K., F. Galvao, A., and Montes-Rojas, G. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170(1):76–91.

Kneip, A. and Ramsay, J. O. (2008). Combining registration and fitting for functional models. *Journal of the American Statistical Association*, 103(483):1155–1165.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89.

Koenker, R. (2005a). *Quantile Regression*. Cambridge University Press, New York.

Koenker, R. (2005b). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

Koenker, R. (2020). *quantreg: Quantile Regression*. R package version 5.61.

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

Koenker, R., Chernozhukov, V., He, X., and Peng, L. (2017). *Handbook of Quantile Regression*. CRC Press, Boca Raton.

Kundu, M. G., Harezlak, J., and Randolph, T. W. (2016). Longitudinal functional models with structured penalties. *Statistical Modelling*, 16(2):114–139.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974.

Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, 157(2):396–408.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2):391–413.

Lee, E. R., Noh, H., and Park, B. U. (2014). Model selection via bayesian information criterion for quantile regression models. *Journal of the American Statistical Association*, 109(505):216–229.

Lee, S. and Jung, S. (2016). Combined analysis of amplitude and phase variations in functional data. *arXiv: Methodology*.

Li, M., Wang, K., Maity, A., and Staicu, A.-M. (2016). Inference in functional linear quantile regression. *arXiv preprint arXiv:1602.08793*.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G., and Zhao, L. P. (1997). Quantile regression methods for longitudinal data with drop-outs: application to cd4 cell counts of patients infected with the human immunodeficiency virus. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(4):463–476.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Annals of Statistics*, 16(4):1696–1708.

Liu, Y. and Wu, Y. (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *Journal of Nonparametric Statistics*, 23(2):415–437.

Luo, Y., Lian, H., and Tian, M. (2012). Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, 82(11):1635–1649.

Maciak, M. (2021a). Quantile LASSO in arbitrage-free option markets. *Econometrics and Statistics*. In press.

Maciak, M. (2021b). Quantile LASSO with changepoints in panel data models applied to option pricing. *Econometrics and Statistics*. In press.

Marino, M. F. and Farcomeni, A. (2015). Linear quantile regression models for longitudinal experiments: an overview. *Metron*, 73(2):229–247.

Marino, M. F., Tzavidis, N., and Alfò, M. (2018). Mixed hidden markov quantile regression models for longitudinal data with possibly incomplete sequences. *Statistical Methods in Medical Research*, 27(7):2231–2246. PMID: 27899706.

Marron, J. S., Ramsay, J. O., Sangalli, L. M., and Srivastava, A. (2015). Functional Data Analysis of Amplitude and Phase Variation. *Statistical Science*, 30(4):468 – 484.

Marx, B. D. and Eilers, P. H. C. (1999). Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics*, 41:1–13.

Modugno, L. and Giannerini, S. (2015). The wild bootstrap for multilevel models. *Communications in Statistics – Theory and Methods*, 44(22):4812–4825.

Neuhaus, J. M., Kalbfleisch, J. D., and Hauck, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, 59(1):23–35.

Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Park, S. Y., Li, C., Benavides, S. M. M., van Heugten, E., and Staicu, A. M. (2019). Conditional Analysis for Mixed Covariates, with Application to Feed Intake of Lactating Sows. *Journal of Probability and Statistics*, 2019:1–14.

Park, S. Y. and Staicu, A.-M. (2015). Longitudinal functional data analysis. *Stat*, 4(1):212–226.

Pini, A., Sørensen, H., Tolver, A., and Vantini, S. (2021). Local inference for functional linear mixed models. Submitted.

R Core Team (2020a). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2020b). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer, New York, second edition.

Reich, B. J., Bondell, H. D., and Wang, H. J. (2009). Flexible bayesian quantile regression for independent and clustered data. *Biostatistics*, 11(2):337–352.

Sangalli, L. M., Secchi, P., Vantini, S., and Vitelli, V. (2010). k-mean alignment for curve clustering. *Computational Statistics & Data Analysis*, 54(5):1219–1233.

Schmid, M. and Hothorn, T. (2008). Boosting additive models using component-wise p-splines. *Computational Statistics & Data Analysis*, 53(2):298–311.

Schmutz, A., Jacques, J., Bouveyron, C., Cheze, L., and Martin, P. (2020). Clustering multivariate functional data in group-specific functional subspaces. *Computational Statistics*, pages 1–31.

Srivastava, A. and Klassen, E. (2016). *Functional and Shape Data Analysis*. Springer Series in Statistics. Springer New York.

Staicu, A.-M., Islam, M. N., Dumitru, R., and van Heugten, E. (2020). Longitudinal dynamic functional regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(1):25–46.

Tang, R. and Müller, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika*, 95(4):875–889.

Tucker, J. D. (2020). *fdasrvf: Elastic Functional Data Analysis*. R package version 1.9.4.

Wang, L., Van Keilegom, I., and Maidman, A. (2018a). Wild residual bootstrap inference for penalized quantile regression with homoscedastic errors. *Biometrika*, 105(4):859–872.

Wang, M., Chen, Z., and Wang, C. D. (2018b). Composite quantile regression for GARCH models using high-frequency data. *Econometrics and Statistics*, 7:115 – 133.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R, Second Edition*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press.

Wood, S., N., Pya, and S"afken, B. (2016). Smoothing parameter and model selection for general smooth models (with discussion). *Journal of the American Statistical Association*, 111:1548–1575.

Wood, S. and Scheipl, F. (2020). *gamm4: Generalized Additive Mixed Models using 'mgcv' and 'lme4'*. R package version 0.2-6.

Wrobel, J., Zipunnikov, V., Schrack, J., and Goldsmith, J. (2019). Registration for exponential family functional data. *Biometrics*, 75(1):48–57.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14(4):1261–1295.

Wu, W. and Srivastava, A. (2014). Analysis of spike train data: Alignment and comparisons using the extended fisher-rao metric. *Electron. J. Statist.*, 8(2):1776–1785.

Xiao, L., Li, C., Checkley, W., and Crainiceanu, C. (2018). Fast covariance estimation for sparse functional data. *Statistics and Computing*, 28:511–522.

Xiao, L., Zipunnikov, V., Ruppert, D., and Crainiceanu, C. (2016). Fast covariance estimation for high-dimensional functional data. *Statistics and computing*, 26(1-2):409–421.

Xiao, Z. (2017). QAR and quantile time series analysis. In Koenker, R., Chernozhukov, V., He, X., and Peng, L., editors, *Handbook of Quantile Regression*, pages 293–332. Chapman and Hall/CRC.

Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):3–25.

Yao, F., Müller, H.-G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A., and Vogel, J. S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics*, 59(3):676–685.

Yi, C. (2017). *hqreg: Regularization Paths for Lasso or Elastic-Net Penalized Huber Loss Regression and Quantile Regression*. R package version 1.4. `https://CRAN.R-project.org/package=hqreg`.

Yu, K. and Zhang, J. (2005). A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics - Theory and Methods*, 34(9-10):1867–1879.

Yuan, M. (2006). GACV for quantile smoothing splines. *Computational Statistics & Data Analysis*, 50(3):813–829.

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44(4):1049–1060.