Quantile regression in heteroscedastic varying coefficient models: testing and variable selection

Mohammed Abdulkerim Ibrahim

Promotor: Prof. Dr Anneleen Verhasselt
Co-promotor: Prof. Dr Irène Gijbels

Acknowledgments

First and foremost, I would like to thank Allah for surrounding me with wonderful people during my PhD. This thesis would not have been possible without the guidance, support and motivation from several individuals. I would like to use this opportunity to express my deepest gratitude to everyone whom in one way or another contributed to the successful completion of this thesis.

I would like to express my sincere gratitude to my promoters Prof. Dr Anneleen Verhasselt and Prof. Dr Irène Gijbels. It has been an honour and privilege working with you. Anneleen, thanks for giving me the opportunity to work with you. I cannot thank you enough for your patience, encouragement, motivation and advices over the past four years, whether on research or career perspectives. Your guidance and supportive feedback helped me during research and the writing of this dissertation. Irène, I would like to thank you for your kindness, time, constructive ideas and valuable comments. I would like to extend my gratitude to every member of the jury for their willingness to read my thesis. I would like to acknowledge with much appreciation your constructive comments and remarks which have improved the thesis.

I am indebted to the BOF doctoral fund of Hasselt University for funding my study. I would also like to acknowledge the VLIR-UOS financial support that gave me an opportunity to study my master's degree at Hasselt University, in 2012. I would like to thank Dr Yudhie Andriyana for sharing and explaining his R-code at the start of my PhD.

I would like to express my special appreciation and thanks to my colleagues at CenStat. Many thanks also to everyone (Trishanta, Yimer, Rudradev, Nolen, Thang,

Thao, Marijke, Ewoud, Alvaro, Oana, Annelies, Daniel, Evangelina, ...) who has been joining me for lunch and coffee. You made me feel at home. To my office mates: Mohamed, Kim, Tapiwa and Thang, I am very much grateful for the nice working environment and for our fruitful discussions and sharing information.

I would like to thank all my Ethiopian friends (Kassu, Alemu, Abebe, ...) for the time we had together. I did not feel alone because of you. My deepest thanks to Yimer, Tadesse and Belay for our remarkable friendship, discussions, advices and encouragements. Yimer, I am blessed to have you more as a brother than a friend during my stay in Belgium and thanks for all the support I received from you.

Finally, a very special thanks to my parents, my mother Aya Dahab Abdi Kelil and my father Awa Abdulkerim Ibrahim Ahmed, for their endless love and support. Similar profound gratitude goes to my brothers and sisters for their unconditional love, continued support and encouragement.

Mohammed Abdulkerim Ibrahim
Hasselt, June 2018

Summary

In this dissertation, quantile regression in varying coefficient models using a nonparametric technique called P-splines is investigated. In mean regression, we study the influence of the covariates on the conditional mean of the response. An alternative way to study the central location is median regression which is robust to heavy-tailed distributions. Quantile regression is a generalization of median regression to investigate the influence of the covariates on the quantiles/percentiles (entire distribution) of the response. It allows for a wide range of applications. For instance, investigating the 25th percentile of the response (e.g. weight of the child) might be of interest in studying severe malnutrition in children. In order to find the estimates, we need to minimize a quantile objective function. In contrast to that of mean regression, the objective function for quantile regression is not differentiable everywhere. Hence, the coefficient estimates have no explicit expression.

A varying coefficient model is an extension of a classical linear regression model, where each coefficient is varying with another variable. This model is important when we have a complex data setting like longitudinal data. In such data scheme, it is intuitive to allow the coefficients to vary with 'time'. We consider, in particular, a location-scale varying coefficient model. The key statistical tools are introduced in Chapter 1.

Population conditional quantiles cannot cross for different quantile levels (percentiles). However, individual conditional quantile estimators can cross each other. To avoid these crossings, we use an approach called 'AHe' (based on two assumptions). This approach enables us to estimate the scale (variability function), and by doing so estimate several quantiles with less computational time. In Chapter 2, we show the

consistency of the proposed estimator theoretically and illustrate it in a simulation study. Since estimation of the quantiles other than the median relies on the variability function, it is important to identify the correct structure of the variability function (e.g. homoscedastic or time varying). Under a homoscedastic variability structure, we can conclude that the influence of the covariates on all other quantiles of the response is similar to that of the median. We develop a testing procedure (Likelihood-Ratio-Type test) to investigate the structure of the variability function. The performance of the testing procedure is shown in the simulation study. The estimation and testing procedures are illustrated on data examples.

In Chapter 3, we focus on testing the shape of a coefficient function (e.g. constant, monotone or convex). Several testing procedures are proposed. A monotonicity test is important to check, for example, that the weight of a child is decreasing at some point in time. The performance of the testing procedures is illustrated in a simulation study. An application to a data example also illustrates the use of the procedures.

To further simplify the model, in Chapter 4, two types of variable selection techniques called 'grouped Adaptive Lasso' and 'NonNegative Garrote' are proposed. The performance of these techniques is illustrated in a simulation study. In Chapter 5, we also propose another two-stage variable selection technique called NNG_{SIS}, in an ultrahigh-dimensional setting (when the number of coefficients to estimate is possibly much bigger than the number of observations). Simulation studies illustrate the performance of the procedures. We also demonstrate the use of the methods on data examples.

Chapter 6 draws some conclusions with discussion of possible future perspectives. The R-code to implement the methodology developed in this dissertation is presented in Appendix A.

Samenvatting

In deze thesis bestuderen we kwantiel regressie in variërende coëfficiënten modellen waarbij we gebruik maken van P-splines, een niet-parametrische schattingsmethode. In klassieke regressiemodellen bestuderen we het effect van covariaten op de gemiddelde respons. In plaats van het gemiddelde te modelleren, kunnen we ook de mediaan bestuderen. Deze maat is robuust voor verdelingen met een zware staart. Kwantiel regressie is een veralgemening van mediaan regressie waarbij het effect van covariaten op de kwantielen/percentielen (volledige verdeling) van de responsvariabelen bestudeerd wordt. Deze regressietechniek heeft verschillende toepassingsmogelijkheden, bvb., bij de studie van ondervoeding bij kinderen is het vaak interessant om het 25ste percentiel van de respons (gewicht van het kind) te kennen. De schatters worden gevonden via het minimaliseren van een kwantiel doelfunctie. In tegenstelling tot regressiemodellen voor de gemiddelde respons, is deze doelfunctie niet overal afleidbaar waardoor er geen expliciete uitdrukking voor de schatters bestaat.

Een variërend coëfficiënten model is een uitbreiding van een lineair regressie model, waarbij iedere regressiecoëfficiënt een functie van een andere variabele kan zijn. Dit model is belangrijk bij complexe data structuren zoals longitudinale data, waarbij het natuurlijk is om aan te nemen dat coëfficiënten variëren over de tijd. We bestuderen in het bijzonder een locatie-schaal variërend coëfficiënten model. De belangrijke statistische aspecten worden in Hoofstuk 1 geïntroduceerd.

Populatie conditionele kwantielen kunnen elkaar niet kruisen over de verschillende percentielen. Individuele schatters voor deze conditionele kwantielen kunnen elkaar helaas wel kruisen. Om dergelijke kruisingen te voorkomen, gebruiken we de AHe methode (gebaseerd op twee aannames). De techniek maakt het mogelijk om de schaal

(variabiliteitsfunctie) te schatten waarbij verschillende kwantielen berekend kunnen worden in minder rekentijd. In Hoofdstuk 2 bewijzen we de consistentie van deze methode en illustreren we de techniek aan de hand van een simulatiestudie. Het is belangrijk om de correcte structuur van de variabiliteitsfunctie te kennen (bvb. homoscedastisch of variërend over de tijd) omdat ieder kwantiel (behalve de mediaan) beïnvloed wordt door deze variabiliteitsfunctie. In een homoscedastische setting geldt dat het effect van covariaten op alle mogelijke kwantielen van de respons gelijk is aan het effect op de mediaan van deze respons. We ontwikkelen een test om de structuur van de variabiliteitsfunctie te controleren. De kwaliteit van deze test wordt weergegeven in een simulatievoorbeeld. Zowel de schattingsmethoden als de testprocedures worden geïllustreerd op datavoorbeelden.

In Hoofdstuk 3 ligt de focus op de vorm van de coëfficiëntfuncties (bvb. constant, monotoon of convex). We stellen verschillende testprocedures voor. Een test voor monotoniciteit is belangrijk om, bij de studie omtrent ondervoeding, te onderzoeken of het gewicht van een kind daalt op een gegeven moment. De kwaliteit van de test wordt aangetoond in een simulatiestudie. De toepasbaarheid van de ontwikkelde techniek wordt weergegeven via een datavoorbeeld.

In Hoofdstuk 4 worden twee variabelenselectiemethoden, grouped adaptive Lasso en nonnegative garrote, voorgesteld om het variërend coëfficiënten model te vereenvoudigen. De kwaliteit van de procedures wordt geïllustreerd in een simulatiestudie. In Hoofdstuk 5 ontwikkelen we vervolgens een twee-stappen variabelenselectietechniek, NNG_{IS} genaamd voor een hoog-dimensionale setting (waarbij het aantal coëfficiënten mogelijk veel groter is dan het aantal observaties). Simulatiestudies tonen de kwaliteit van deze technieken. We illustreren het gebruik van de techniek in datavoorbeelden.

De conclusies en toekomst mogelijkheden omtrent dit onderzoek worden in Hoofdstuk 6 besproken. De R-code voor het implementeren van de methodologie, die in deze thesis ontwikkeld werd, is beschikbaar in Appendix A.

List of Publications

This dissertation is based on the following publications and reports:

- Gijbels, I., **Ibrahim, M. A.**, and Verhasselt, A., (2018). Testing the heteroscedastic error structure in quantile varying coefficient models. *Canadian Journal of Statistics*, 46(2):246–264.
- Gijbels, I., **Ibrahim, M. A.**, and Verhasselt, A., (2017). Shape testing in quantile varying coefficient models with heteroscedastic error. *Journal of Nonparametric Statistics*, 29(2):391–406.
- **Ibrahim, M. A.**, and Verhasselt, A., (2018a). A two-stage screening and variable selection method in ultrahigh-dimensional quantile varying coefficient models with heteroscedastic error. [***Submitted***].
- **Ibrahim, M. A.**, and Verhasselt, A., (2018b). Variable selection in quantile varying coefficient models with heteroscedastic error. [***Submitted***].
- Andriyana, Y., **Ibrahim, M. A.**, Gijbels, I., and Verhasselt, A., (2018). *QRegVCM: Quantile Regression in Varying-Coefficient Models.* R package version 1.2.

Contents

List of Tables

List of Figures

List of Abbreviations

$\xrightarrow{P}$	Convergence in probability
AHe	Adaptive He
EVT	Extreme Value Test
FN	Frisch-Newton interior point
gALassoP	grouped Adaptive Lasso using P-splines
Lasso	Least absolute shrinkage and selection operator
LP	Linear Programming
LRT	Likelihood-Ratio-Type test
NNG	NonNegative Garrote
P-splines	Penalized B-splines
RS	Rank Score
SCAD	Smoothly Clipped Absolute Deviation
SIC-criterion	Schwarz-type Information Criterion
SIS	Sure Independence Screening
SNR	Signal-to-Noise Ratio

Chapter 1

Introduction

Longitudinal data arise in various research areas including econometrics, epidemiology, clinical trials and environmental science. It refers to data observed over time (e.g. the CD4 data in Section 2.4.1). In the CD data, it is of interest to investigate the evolution of CD4 percentage over time. The 90th percentile of CD4 might evolve differently than the median or the 10th percentile. Further, a flexible smoothing technique may need to be considered in order to capture the evolution of the CD4 percentage over time.

The focus of this dissertation is on quantile regression in varying coefficient models using one particular non-parametric technique called penalized B-splines (P-splines). Therefore, in Section 1.1, the basics on quantile regression are introduced. The second basic tool is variable selection, which is briefly described in Section 1.2. Section 1.3 presents the basics on P-splines. Then, Section 1.4 discusses a varying coefficient model. Finally, Section 1.5 presents the contributions made in this thesis.

1.1 Quantile regression

Let $Y \in \mathbb{R}$ be our response variable and $(X^{(1)}, \ldots, X^{(p)})^{\mathrm{T}} \in \mathbb{R}^p$ ($\mathbf{A}^{\mathrm{T}}$ denotes the transpose of a vector or matrix $\mathbf{A}$) be p-dimensional covariate vector. We want to regress the response on the covariates, i.e. we want to predict/estimate the response based on the covariates.

As an alternative to least squares regression, quantile regression allows a wide range

of applications, where several conditional quantiles are of interest instead of just the conditional mean. Studying the entire distribution of the response rather than only the central tendency, as in mean regression, is important. Especially when the distribution has a heteroscedastic nature, only the central tendency cannot represent the entire distribution. If most of the observations are concentrated, for instance, on the 75th percentile of the distribution, then it is more appropriate to consider the 75% regression quantile than mean regression. If one is interested, for example, in studying severe malnutrition in children, then lower quantiles of 'weight' can be of interest. Further, quantile regression does not assume a specific form for the (conditional) distribution, and thus is able to accommodate non-normal errors. When the response variable given a set of covariates has a heavy-tailed distribution, quantile regression puts a reduced weight on the extreme observations. Furthermore, due to its robustness to outliers, there is a growing interest in the literature on quantile regression.

Quantile regression, as a generalization of median regression, was first introduced by Koenker and Bassett (1978) (refer to Koenker (2005) for a comprehensive introduction to quantile regression). Estimation of conditional quantiles relies on the non-differentiable and asymmetric check-loss function of Koenker and Bassett (1978), $\rho_\tau(z) = z[\tau - I(z < 0)]$ for $\tau \in (0, 1)$, with $I(A) = 1$ if A holds, and 0 otherwise, rather than the square loss function (in mean regression). Therefore, the computational and theoretical aspects of conditional quantile regression are different to that of mean regression.

1.1.1 Unconditional quantiles

As a starting point, let us compare the τth quantile of Y (denoted as $q_\tau(Y)$) with the mean of Y (denoted as $\mu = EY$). The mean of Y is the center of distribution which minimizes the expected L_2-loss deviation,

$$\mu = \arg\min_c E(Y - c)^2.$$

The τth quantile of Y rather minimizes the expected check-loss deviation,

$$q_\tau(Y) = \arg\min_c E[\rho_\tau(Y - c)].$$

Figure 1.1 depicts the loss functions that lead to the mean and quantiles of Y. As can be seen in this figure, the check-loss function $\rho_\tau(\cdot)$ (Koenker and Bassett, 1978) is

non-differentiable at zero, hence the minimizer has no explicit solution. This calls for the use of a Linear Programming (LP) optimization technique (Koenker and Bassett, 1978) like Frisch-Newton interior point (FN) algorithm (Portnoy and Koenker, 1997) or a convex programming optimization algorithms like a Matlab-based modeling system called CVX (Grant and Boyd, 2012). Linear Programming is considered in this dissertation. Median regression is a special case of quantile regression with $\tau = 0.5$,

$$q_{0.5}(Y) = \arg\min_{c} E[\rho_{0.5}(Y - c)] = \arg\min_{c} E[0.5|Y - c|]. \tag{1.1}$$

Equation (1.1) states that to obtain the median we need to minimize the expected absolute deviations. Hence, median regression depends mainly on the sign of the deviation rather than the magnitude. That is why median regression is robust to outliers. For median regression, as can be seen in Figure 1.1 and Equation (1.1), both negative and positive residuals are equally weighted. Where as for the other quantiles, the loss function is asymmetric.

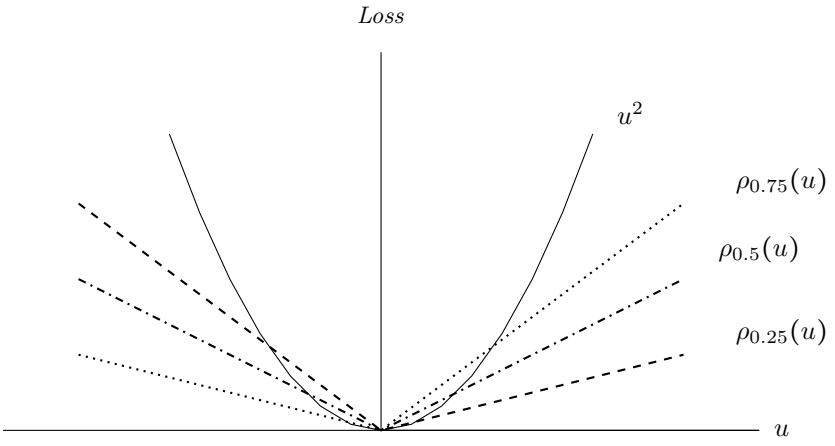


Figure 1.1. The check-loss and squared-loss functions.

1.1.2 Conditional quantiles

It is easy to extend quantile estimation to quantile regression. As in mean regression, we only need to incorporate the effect of covariates on the quantile of the response.

Suppose we have the following linear model:

$$Y = \beta_0 + \beta_1 X^{(1)} + \ldots + \beta_p X^{(p)} + \tilde{\epsilon} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \tilde{\epsilon}, \tag{1.2}$$

with $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^{\mathrm{T}}$ and $\mathbf{X} = (1, X^{(1)}, \ldots, X^{(p)})^{\mathrm{T}}$. Assuming that the conditional mean of the error term $\tilde{\epsilon}$ is zero, the conditional mean of Y is $E(Y|\mathbf{X}) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}$. The coefficient vector $\boldsymbol{\beta}$ is found solving $\min_{\boldsymbol{\beta}} E(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta})^2$. Proceeding similarly, assuming that the conditional τth quantile of the error term is zero, we obtain the coefficient vector of the τth conditional quantile of Y, $q_\tau(Y|\mathbf{X}) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_\tau$, solving

$$\arg\min_{\boldsymbol{\beta}} E[\rho_\tau(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta})], \tag{1.3}$$

with $\boldsymbol{\beta}_\tau = (\beta_{\tau 0}, \beta_{\tau 1}, \ldots, \beta_{\tau p})^{\mathrm{T}}$ the τ-dependent coefficient vector.

For an observed response Y_i and covariates $\mathbf{X}_i = (1, X_i^{(1)}, \ldots, X_i^{(p)})^{\mathrm{T}}$, for $i = 1, \ldots, n$, the empirical version of the objective function in Equation (1.3) is

$$\arg\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}). \tag{1.4}$$

The LP problem of Equation 1.4 can be written as,

$$\min_{\text{all } u_i, v_i} \left\{ \tau \sum_{i=1}^{n} u_i + (1-\tau) \sum_{i=1}^{n} v_i \right\}$$

subject to

$$\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} = Y_i/n \quad \text{for all } i = 1, \ldots, n,$$

where $u_i \geq 0$ and $v_i \geq 0$ are, respectively, the positive and the negative regression residuals.

The dual formulation of the above LP problem can then be solved by the FN algorithm.

Throughout the dissertation, we need the following two essential properties of the conditional quantiles.

Property:

1. *Equivariance.* Throughout the dissertation we transform the covariates such that they have equivalent scales. We expect such changes to have no fundamen-

tal changes on our coefficients. We call such invariant property 'equivariance'. Let us denote the coefficient vector of $q_\tau(Y|\mathbf{X})$ for the model in Equation (1.2) as $\boldsymbol{\beta}_{\tau;Y,\mathbf{X}}$, let $\mathbf{A}$ be any $(p+1)\times(p+1)$ nonsingular matrix, $\boldsymbol{\eta}\in\mathbb{R}^{p+1}$, $h(\cdot)$ is a nondecreasing function on $\mathbb{R}$ and $a>0$, then for any $\tau\in(0,1)$

- Scale equivariance: $\boldsymbol{\beta}_{\tau;aY,\mathbf{X}} = a\boldsymbol{\beta}_{\tau;Y,\mathbf{X}}$ and $\boldsymbol{\beta}_{\tau;-aY,\mathbf{X}} = -a\boldsymbol{\beta}_{1-\tau;Y,\mathbf{X}}$;
- Shift equivariance: $\boldsymbol{\beta}_{\tau;Y+\mathbf{X}^{\mathrm{T}}\boldsymbol{\eta},\mathbf{X}} = \boldsymbol{\beta}_{\tau;Y,\mathbf{X}} + \boldsymbol{\eta}$;
- Equivariance to reparametrization of design: $\boldsymbol{\beta}_{\tau;Y,\mathbf{X}^{\mathrm{T}}\mathbf{A}} = \left(\mathbf{A}^{-1}\right)^{\mathrm{T}}\boldsymbol{\beta}_{\tau;Y,\mathbf{X}}$;
- Equivariance to nondecreasing transformations: $q_\tau(h(Y)|\mathbf{X}) = h(q_\tau(Y|\mathbf{X}))$.

2. *Non-crossingness.* Conditional quantile curves are monotone increasing in the argument $\tau\in(0,1)$,

$$q_{\tau_1}(Y|\mathbf{X}) \le q_{\tau_2}(Y|\mathbf{X}), \qquad \text{for } 0<\tau_1\le\tau_2<1 \text{ and for all } \mathbf{X}.$$

Property 1 holds for $\hat{\boldsymbol{\beta}}_{\tau;\mathbf{Y},\boldsymbol{\mathcal{X}}}$, the minimizer of Equation (1.4), with observations $\mathbf{Y} = (Y_1,\ldots,Y_n)^{\mathrm{T}}$ and $\boldsymbol{\mathcal{X}} = (\mathbf{X}_1,\ldots,\mathbf{X}_n)^{\mathrm{T}}$. However, Property 2 holds only at the centroid of the design, $\bar{\mathbf{X}} = (1,\bar{X}^{(1)},\ldots,\bar{X}^{(p)})^{\mathrm{T}}$, with $\bar{X}^{(k)} = n^{-1}\sum_{i=1}^{n} X_i^{(k)}$ for $k=1,\ldots,p$, for the estimated conditional quantile functions $\hat{q}_\tau(Y|\mathbf{X}) = \mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{\tau;\mathbf{Y},\boldsymbol{\mathcal{X}}}$, as shown in Theorem 2.5 of Koenker (2005).

There are several ways to prevent the problem of crossing estimated conditional quantile curves, like the weighted simultaneous quantile functions (see e.g. Zou and Yuan, 2008; Zhao and Xiao, 2014), the quantile sheet (Schnabel and Eilers, 2013), the stepwise individual quantile regression (Wu and Liu, 2009a) and the restricted regression quantiles approach (He, 1997). In this thesis, we consider the restricted regression quantiles approach. The advantage of this approach is that, it enable us to estimate and identify the variability function (defined in Section 1.4). Using the estimate of the variability, several regression quantiles can be estimated in less computational time (especially when nonparametric techniques are considered to estimate the regression coefficients or when we have high-dimensional data, where the number of covariates is large).

1.2 Variable selection

When the number of covariates is large, the estimated model might not be easy to interpret or the prediction accuracy might be low, which calls for reducing the number of nuisance covariates from the model. This can be done by standard techniques such as subset selection and ridge regression. Subset selection leads to an interpretable result, but can be extremely variable (the selected model can vary due to small changes in the data). Whereas ridge regression shrinks the coefficients but does not set them to zero. In this thesis, several shrinkage approaches (the Least absolute shrinkage and selection operator (Lasso), the NonNegative Garrote (NNG), the Smoothly Clipped Absolute Deviation (SCAD)), that shrink the coefficients of irrelevant covariates to zeros, are considered. For high dimensional models, the Sure Independence Screening (SIS) is implemented to reduce the dimensionality from high to a moderate scale that is below the sample size.

Tibshirani (1996) propose the Lasso approach by adapting the ordinary least squares estimates, such that some coefficients are shrunken and others are set to zero. The Lasso retains the good feature (removing the irrelevant variables from the model) of both subset selection and ridge regression. The estimates for the model in Equation (1.2) are defined by,

$$\hat{\boldsymbol{\beta}}_{\tau}^{Lasso} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_{\tau}(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) + \lambda_L \sum_{k=0}^{p} |\beta_k|,$$

where $\lambda_L \geq 0$ is a tuning parameter that controls the amount of shrinkage.

The Lasso approach is motivated by Breiman's NNG (Breiman, 1995). The estimates are,

$$\hat{\mathbf{c}}_{\tau} = \arg\min_{c_0,\ldots,c_p} \sum_{i=1}^{n} \rho_{\tau}\left(Y_i - \sum_{k=0}^{p} c_k \hat{\beta}_{\tau k}^{\text{init}} X_i^{(k)}\right) + \lambda_{\text{NNG}} \sum_{k=0}^{p} c_k \text{ s.t. } c_k \geq 0 \ (k = 0, 1, \ldots, p),$$

where $\hat{\mathbf{c}}_{\tau} = (\hat{c}_{\tau 1}, \ldots, \hat{c}_{\tau p})^{\mathrm{T}}$; $\hat{\boldsymbol{\beta}}_{\tau}^{\text{init}}$ is the minimizer of Equation (1.4); and $\lambda_{NNG} \geq 0$ is a tuning parameter. The NNG estimator for $\beta_{\tau k}$ is given by,

$$\hat{\beta}_{\tau k}^{NNG} = \hat{c}_{\tau k} \hat{\beta}_{\tau k}^{\text{init}}.$$

As another attempt to retain the good features of both subset selection and ridge regression, Fan and Li (2001) propose the SCAD method. The coefficient estimates

are,

$$\hat{\boldsymbol{\beta}}_{\tau}^{SCAD} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_{\tau}(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) + \sum_{k=0}^{p} \varrho'_{\lambda_{\mathrm{SCAD}}}(|\beta_k|)|\beta_k|,$$

where for some constant $a > 2$,

$$\varrho'_{\lambda_{\mathrm{SCAD}}}(x) = \lambda_{\mathrm{SCAD}} \left(I(x \leq \lambda_{\mathrm{SCAD}}) + \frac{(a\lambda_{\mathrm{SCAD}} - x)_+}{(a-1)\lambda_{\mathrm{SCAD}}} I(x > \lambda_{\mathrm{SCAD}}) \right).$$

The above shrinkage approaches are applied to moderate-to-high dimensional models, i.e. when the number of parameters or variables p is comparable with the number of observations n. When p is much larger than n, Candes and Tao (2007) propose the Dantzig selector and show that it achieves the ideal risk up to a logarithmic factor $\log(p)$. Motivated by the fact that this factor can be large when the dimensionality is ultrahigh, Fan and Lv (2008) propose the SIS. They show that the SIS has the sure screening property for even exponentially growing dimensionality. SIS is based on correlation learning, to reduce the dimensionality from ultrahigh to a moderate scale (that is below the sample size n). Hence, the problems of the estimation accuracy and the computational cost, that is generated when applying the above shrinkage approaches on the ultrahigh-dimensional models, is resolved. When the dimension p is large, we assume that only a small number of predictors contribute to the response or in other words that $\boldsymbol{\beta}_{\tau}$ is sparse. It is defined by first obtaining $\boldsymbol{\upsilon} = (\upsilon_1, \ldots, \upsilon_p)^{\mathrm{T}}$, a vector of marginal correlations of predictors with the response variable. Then, the components of $\boldsymbol{\upsilon}$ are sorted in a decreasing order and a submodel is defined

$$M_a = \{1 \leq k \leq p : |\upsilon_k| \text{ is among the first } [an] \text{ largest of all}\},$$

where $[an]$ denotes the integer part of an.

1.3 P-splines

Suppose we have the following model:

$$Y = \beta(T) + \tilde{\epsilon}, \tag{1.5}$$

where Y is the response variable, $\beta(\cdot)$ is an unknown smooth function of a given covariate $T \in [a_0, a_1]$ and $\tilde{\epsilon}$ is the error term.

To estimate the function $\beta(\cdot)$, several smoothing methods are considered in the literature for quantile regression: local polynomial in Chaudhuri (1991), Honda (2004) and Cai and Xu (2008), smoothing splines in He et al. (1998), B-splines in He and Shi (1994) and Kim (2007) and P-splines in Bollaerts et al. (2006) and Andriyana et al. (2014, 2016).

B-splines (de Boor, 2001) are one particular type of splines, which are piecewise polynomial functions of a given degree with local support. They are determined by a degree ν and a knot sequence $l_1 \leq \ldots \leq a_0 = l_{\nu+1} \leq \ldots \leq a_1 = l_{u+\nu+1} \leq \ldots \leq l_{u+2\nu+1}$ (with $u+1$ knots in the T interval) and have the following properties (Eilers and Marx, 1996):

- it consists of $\nu+1$ polynomial pieces of degree ν;
- the polynomial pieces join at ν interior knots;
- at the joining points, the derivatives up to order $\nu-1$ are continuous;
- the B-spline is positive on the domain spanned by $\nu+2$ knots, it is zero everywhere else;
- except at the boundaries, it overlaps with 2ν polynomial pieces of its neighbors;
- at a given T, $\nu+1$ B-splines are nonzero.

The jth B-spline of degree $\nu \geq 1$, denoted by $B_j(T,\nu)$, can be determined from B-splines of a lower degree via the recursive formula:

$$B_j(T,\nu) = \frac{T-l_j}{l_{j+\nu}-l_j} B_j(T,\nu-1) + \frac{l_{j+\nu+1}-T}{l_{j+\nu+1}-l_{j+1}} B_{j+1}(T,\nu-1),$$

with $j = 1,\ldots,u+\nu$ and

$$B_j(T,0) = \begin{cases} 1 & \text{if } l_j \leq T < l_{j+1} \\ 0 & \text{otherwise.} \end{cases}$$

The B-splines are normalized such that $\sum_{l=1}^{u+\nu} B_l(T,\nu) = 1$ for all T. Figure 1.2 illustrates B-splines of degree one, when $u+1 = 3$ equidistant knots are used in the T interval. To construct the recursion, ν knots are added to both sides of the interval. As shown in the figure, three B-splines of degree one are constructed for $T \in [l_2, l_4]$. Each B-spline consists of two linear pieces; for instance, B_1 has one piece from l_1 to l_2, the other from l_2 to l_3 joined at the interior knot l_2.

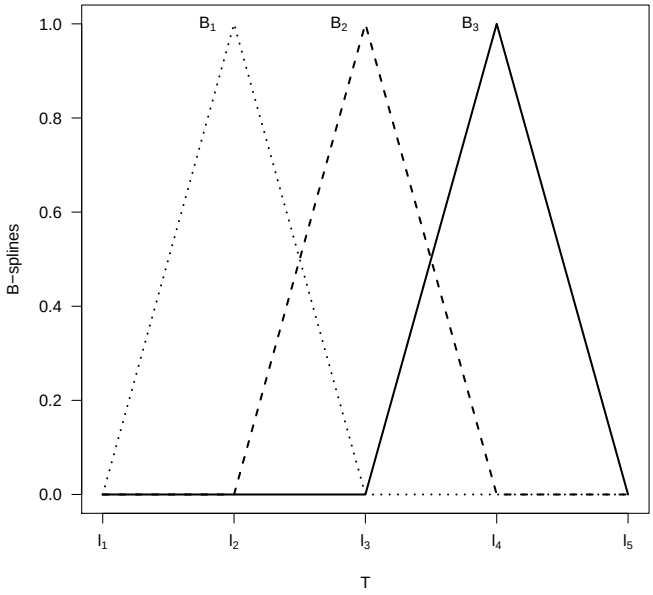


Figure 1.2. B-splines of degree 1.

We assume that the coefficient function $\beta(\cdot)$ in the model in Equation (1.5) is approximated by the following linear combination of B-splines:

$$\beta(T) \approx \sum_{j=1}^{m} \alpha_j B_j(T, \nu),$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^{\mathrm{T}}$ is the coefficient vector of the associated B-splines $B_j(\cdot, \nu)$, for $j = 1, \ldots, u + \nu = m$, of degree ν and $u + 1$ equidistant knots.

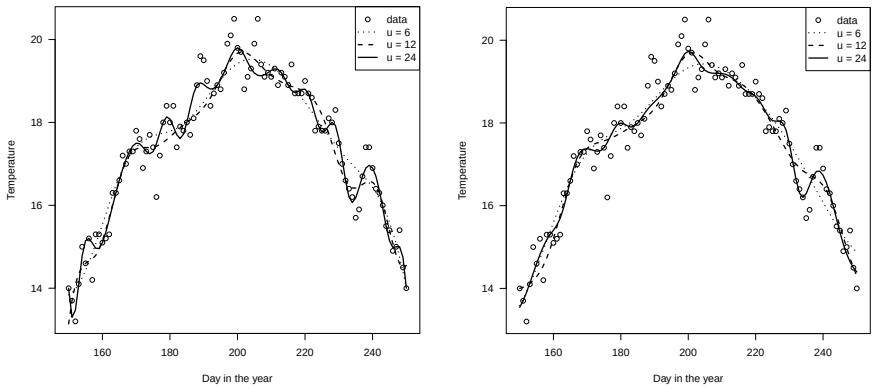
To overcome the overfitting problem of B-splines regression, Eilers and Marx (1996) proposed P-splines, where a difference penalty on the coefficient of adjacent B-splines is added to the loss function in the optimization problem. The adaptation of P-splines to quantile regression is given by Bollaerts et al. (2006). The objective function is given by

$$\min_{\boldsymbol{\alpha}} \left\{ \sum_{i=1}^{n} \rho_\tau \left(Y_i - \sum_{j=1}^{m} \alpha_j B_j(t_i, \nu) \right) + \lambda_P \sum_{j=d+1}^{m} |\Delta^d \alpha_j| \right\}, \tag{1.6}$$

where t_i is an observed value for T; $\lambda_P > 0$ is the smoothing parameter; d is the differencing order in the penalty term; and Δ^d is the dth order differencing operator:

$\Delta^d \alpha_j = \sum_{t=0}^{d} (-1)^t \binom{d}{t} \alpha_{j-t}$, for instance, $\Delta^d \alpha_j = \alpha_j - \alpha_{j-1}$ for $d = 1$. Based on the estimated coefficient vector $\hat{\boldsymbol{\alpha}}$ that minimizes Equation (1.6), we find that $\hat{\beta}(t_i) = \sum_{j=1}^{m} \hat{\alpha}_j B_j(t_i, \nu)$.

To illustrate the P-splines technique, we use the Canadian weather data in the fda R-package. The response variable Y is the average daily temperature in degree celcius for each day of the year, at a weather station in Canada, averaged over 1960 to 1994. The covariate T is the day of summer (the days 150 till 250 in the year). As can be seen in Figure 1.3 (a) the regular regression with B-splines tends to overfit when the number of knots grows. This overfitting problem is solved by using P-splines, and the choice of the number knots has now a minor effect on the regression curve, as shown in Figure 1.3 (b). Figure 1.4 shows the influence of the smoothing parameter λ_P on the fitted curve. As λ_P increases the fitted curve tends to be smoother. Hence, when applying the P-splines technique, a large number of knots with a good data driven smoothing parameter needs to be employed (see Chapter 2 for details).



(a) (b)

Figure 1.3. Canadian weather data. Conditional median estimators using $u = 6, 12,$ and 24 and $\nu = 3$. (a) B-splines and (b) P-splines with $d = 1$ and $\lambda_P = 1$.

1.4 Varying coefficient models

The model in Equation (1.2) is the classical multiple linear regression model, where the regression coefficient vector $\boldsymbol{\beta}$ is assumed to be a constant vector. This assumption might be too strong in some situations where the coefficient can vary over another

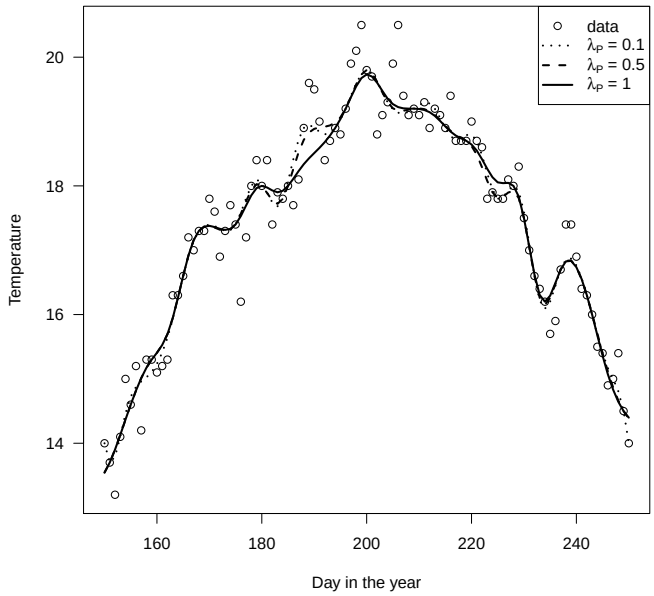


Figure 1.4. Canadian weather data. Conditional median estimators using P-splines with $\lambda_P = 0.1, 0.5,$ and 1, $u = 24$, $\nu = 3$ and $d = 1$.

variable in the data set.

Varying coefficient models are considered in this dissertation, as our data setting has a longitudinal nature and we need to allow the coefficients to vary with 'time'. As a completely nonparametric relationship between a response variable and several covariates is difficult to estimate and to alleviate the curse of dimensionality, varying coefficient models are introduced by Hastie and Tibshirani (1993). Due to the flexibility and interpretability of varying coefficient models, a lot of research has been done on parameter estimation and hypothesis testing, but mostly in a mean regression setting. Varying coefficient models are used to model longitudinal data (see e.g. Antoniadis et al., 2012b; Tang et al., 2013b; Andriyana et al., 2014).

In this dissertation, a longitudinal observational data scheme is considered, where each subject has measurements at different time points. The measurements from different subjects are assumed to be independent, but measurements at different time points within each subject can be correlated.

We have longitudinal observations $(Y(t_{ij}), \mathbf{X}^*(t_{ij})^{\mathrm{T}}, t_{ij})^{\mathrm{T}}$ of $(Y(T), \mathbf{X}^*(T)^{\mathrm{T}}, T)^{\mathrm{T}} \in$

$\mathbb{R}^{p+2}$ with $\mathbf{X}^*(T) = (X^{(1)}(T), \ldots, X^{(p)}(T))^{\mathrm{T}}$ for $i = 1, \ldots, n$, $j = 1, \ldots, N_i$, and $p \geq 1$, where t_{ij} is the jth measurement time for the ith subject, N_i is the number of repeated measurements for the ith subject, and $Y(t_{ij})$ and $\mathbf{X}^*(t_{ij}) = (X^{(1)}(t_{ij}), \ldots, X^{(p)}(t_{ij}))^{\mathrm{T}}$ are the observed outcome and covariates of the ith subject at time point t_{ij}. The time points t_{ij} take values in the space $\mathcal{T} \subset \mathbb{R}$, we denote $\mathbf{X}(T) = (X^{(0)}(T), X^{(1)}(T), \ldots, X^{(p)}(T))^{\mathrm{T}}$ with $X^{(0)}(T) \equiv 1$. We consider the following location-scale type varying coefficient model (which is a generalization of the linear model in Equation (1.2), where the coefficients $\boldsymbol{\beta}$ are allowed to vary with time variable T):

$$\begin{aligned} Y(T) &= \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T) + V(\mathbf{X}(T), T)\epsilon(T) \\ &= \beta_0(T) + \beta_1(T)X^{(1)}(T) + \ldots + \beta_p(T)X^{(p)}(T) + V(\mathbf{X}(T), T)\epsilon(T), \end{aligned} \tag{1.7}$$

where we have errors $\epsilon(t_{ij})$ of $\epsilon(T) \in \mathbb{R}$ for $i = 1, \ldots, n$, $j = 1, \ldots, N_i$ and $\boldsymbol{\beta}(T) = (\beta_0(T), \beta_1(T), \ldots, \beta_p(T))^{\mathrm{T}}$ is the vector of unknown regression coefficient functions at time T, with $\beta_0(T)$ the baseline effect. We call $V(\mathbf{X}(T), T)$ the variability function. To be short we write $Y(t_{ij}), \mathbf{X}(t_{ij}), \epsilon(t_{ij})$ as $Y_{ij}, \mathbf{X}_{ij}, \epsilon_{ij}$. The $\tilde{\epsilon}$ of the model in Equation (1.2) is written in Equation (1.7) as $V(\mathbf{X}(T), T)\epsilon(T)$ to allow for heteroscedasticity. Since the variability function is allowed to vary with $\mathbf{X}(T)$ and T, the coefficient functions for each covariate is τ specific. If the variability function is of a homoscedastic nature (i.e., $V(\mathbf{X}(T), T) = \gamma$, for a nonnegative constant γ), all quantile curves are as the median curve up to a shift (as only the intercept changes to $\beta_{\tau 0}(T) = \beta_0(T) + \gamma q_\tau(\epsilon(T)|\mathbf{X}(T), T)$). Some examples of heteroscedatic variability function are $V(\mathbf{X}(T), T) = \gamma(T)$ or $V(\mathbf{X}(T), T) = (\|\mathbf{X}(T)\|_1)^\theta$ (with $\theta \in \mathbb{R}$ and $\|\mathbf{X}(T)\|_1 = |X^{(0)}(T)| + |X^{(1)}(T)| + \ldots + |X^{(p)}(T)|$).

We estimate the τth quantile of Y, $\beta_k(T)$ (with $k = 0, \ldots, p$) and $V(\mathbf{X}(T), T)$ with the P-splines smoothing technique. The use of this smoothing technique is appealing here, since it allows for flexible modeling and has computational advantages in varying coefficient models.

In order to analyze longitudinal data, Koenker (2004) considers a model where the subject specific effects ($\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)^{\mathrm{T}}$) are introduced in the model as fixed effects. They use a shrinkage approach to control the variability introduced by the large number of estimated σ parameters. However, the subject specific effects are assumed to be independent of the quantile levels (i.e. the $\boldsymbol{\sigma}$ are common for all quantile levels) and the computational complexity increases as $n \to \infty$. Wang and Sun (2017) propose an iterative algorithm which solves an estimating equation that incorporates

a working correlation, to take into account the correlation within subject. Another suggestion is by considering the $\boldsymbol{\sigma}$ as random effects (Reich et al., 2010; Geraci and Bottai, 2014, to name a few).

In this thesis, we consider an estimation procedure which makes very minimal assumptions on the form of the error distribution and thus is able to accommodate nonnormal errors, and it is computationally convenient (Wang et al., 2009; Andriyana et al., 2014, to cite a few).

1.5 Outline of the dissertation

We investigate quantile regression in varying coefficient models with a heteroscedastic error structure. In Chapter 2, we propose an estimation technique for the variability function and show its consistency. When the variability function is misspecified, the estimated conditional quantiles other than the median are not correct, as is shown in Figure 2.4 (a). Hence, a Likelihood-Ratio-Type test is proposed to identify what type of function $V(\mathbf{X}(T), T)$ is. The proposed methods are implemented in the R-package QRegVCM (Andriyana, Ibrahim, Gijbels, and Verhasselt, 2018). The research of Chapter 2 is published in Gijbels, Ibrahim, and Verhasselt (2018).

In Chapter 3, we develop several new shape testing procedures for both the median and variability functions, that have not been considered in the literature before. A Likelihood-Ratio-Type test is proposed to test the shape (constancy, monotonicity and/or convexity) of the functional coefficients. Further, testing procedures based on L_1-norm, L_2-norm and L_∞-norm of the differences of the P-splines coefficients are introduced to test for constant functional coefficients. An extreme value test for testing monotonicity or convexity is also formulated. Chapter 3 summarizes the material presented in Gijbels, Ibrahim, and Verhasselt (2017).

There is a vast literature on variable selection in varying coefficient models when the focus is on quantile estimation (e.g., Noh, Chung, and Van Keilegom, 2012; Tang, Wang, and Zhu, 2013b). In Chapter 4, we propose an easy way to check the influence of the covariates on the distribution of the response, that combines smoothing and variable selection. Two types of variable selection techniques are considered: grouped Adaptive Lasso using P-splines (gALassoP) and NonNegative Garrote. The consistency of the grouped Adaptive Lasso is shown. This chapter discusses the methods presented in Ibrahim and Verhasselt (2018b).

Chapter 5 deals with selecting important covariates in the data set when the number of coefficients to be estimated is much bigger than the number of observations. A two-stage variable selection procedure for ultrahigh-dimensional data is introduced for both the location and the scale. In the first stage, we consider the Sure Independence Screening in order to reduce the dimension, and in the second stage we use the NonNegative Garrote on the reduced model. This chapter presents the results given in Ibrahim and Verhasselt (2018a).

Chapter 6 presents general conclusions and avenues for further research. Appendix A holds the R-codes to implement the results in this dissertation.

Part I

Testing in quantile varying coefficient models

Chapter 2

Testing the heteroscedastic error structure in quantile varying coefficient models

Quantile regression in varying coefficient models using P-splines estimation is studied in Andriyana (2015). In that work, simple and general heteroscedastic error structures for the variance of the errors (the variability function $V(\mathbf{X}(T), T)$) that are linear (in the parameters/covariates) are investigated using the model in Equation (1.7). In this chapter, we consider more elaborate structures for the variability function, like power and exponential functions. We also prove that our estimator is consistent when the variability function is of an exponential type. Furthermore, we provide a test for choosing between two variability functions based on a Likelihood-Ratio-Type test procedure.

In the mean regression case, Davidian and Giltinan (1995) consider the square of the mean function for the error variance; Fortin et al. (2007) propose an unknown power of the mean function. Inspired by these proposals, we consider variability functions of the following form:

$$V(\mathbf{X}(T), T) = \gamma(T) f_{\theta(T)}(\|\mathbf{X}(T)\|), \tag{2.1}$$

where $\gamma(T) \geqslant 0$, $f_{\theta(T)}(\|\mathbf{X}(T)\|) \geqslant 0$, for a well-defined vector norm, and $\theta(T) \in \mathbb{R}$. More specifically, the following four special cases of the variability function $V(\mathbf{X}(T), T)$ in Equation (2.1) are investigated:

V_1: $V(\mathbf{X}(T), T) = \gamma(\|\mathbf{X}(T)\|_1)^\theta$ (with $\gamma \geq 0$ and $\theta \in \mathbb{R}$);

V_2: $V(\mathbf{X}(T), T) = \gamma(T)(\|\mathbf{X}(T)\|_1)^\theta$;

V_3: $V(\mathbf{X}(T), T) = \gamma(T)\exp(\theta\|\mathbf{X}^*(T)\|_1) = \gamma(T)[\exp(\|\mathbf{X}^*(T)\|_1)]^\theta$;

V_4: $V(\mathbf{X}(T), T) = \gamma(T)\exp(\theta_1(T)|X^{(1)}(T)| + \ldots + \theta_p(T)|X^{(p)}(T)|)$.

The first variability function is a special case of the second one for $\gamma(T) = \gamma$. Modeling structure V_3 is a generalization of V_1 and V_2, in the sense that a first order approximation (for $\|\mathbf{X}^*(T)\|_1$ close to zero) of the exponential function $\exp(\|\mathbf{X}^*(T)\|_1)$ leads to $\|\mathbf{X}(T)\|_1$. Modeling structure V_4 is a further generalization of V_3. We say that structure V_1 is nested in V_2 (denoted hereafter as $V_1 \subseteq V_2$), V_2 is "nested" (in an approximative sense) in V_3, and V_3 is nested in V_4. Model V_2, with $\theta = 0$, is considered by Andriyana et al. (2016). A variability function similar to V_4 is considered by Van Keilegom and Wang (2010), with a partially linear structure. In V_4, we allow the variability function to depend on each covariate and each coefficient is allowed to vary with time T. In such a framework one can check the constancy of the varying coefficients or do variable selection. By Assumption **H1** stated in the next section, the above variability functions have an influence on all quantiles other than the median.

The rest of this chapter is organized as follows. Sections 2.1 and 2.2 deal with estimation and inference methods of the variability functions, respectively. Simulation studies are carried out in Section 2.3. The estimation method and the testing procedure are applied to data examples in Section 2.4. Section 2.5 concludes the results in this chapter. Finally, Section 2.6 presents the theoretical details. The R-code to implement the methods of this chapter on a data set is deferred to Appendix A.

2.1 Estimation procedure

The model in Equation (1.7) consists of three parts, the signal $\mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)$, the variability function $V(\mathbf{X}(T), T)$ and the error $\epsilon(T)$. Estimating all the components simultaneously is not feasible without imposing some assumption(s). Hence, we use an approach similar to that of Andriyana et al. (2016). Such an approach consists of adapting the approach of He (1997) to the context of varying coefficient models, and is therefore termed the 'Adaptive He (AHe) approach' in the literature. The advantage of this approach is two-fold: it avoids crossing of the estimated quantile curves (when several quantiles are estimated) and allows estimation (and identification) of the variability function. Note that the variability function is only identifiable when extra assumptions are made on the error structure.

We impose the following assumptions on the error structure:

H1. $q_{0.5}(\epsilon(T)|\mathbf{X}(T), T) = 0$.

H2. $q_{0.5}\left(\ln|\epsilon(T)| \,\middle|\, \mathbf{X}(T), T\right) = 0$.

Although the assumptions may appear strong, they are satisfied for a properly standardized error structure, as can be seen from Equation (2.14) in Section 2.3.

The estimation procedure then consists of the following three steps:

1. assuming **H1**, the median function is estimated,

2. assuming **H2**, the variability function is estimated, and

3. the τth quantile regression estimate is obtained, for various values of $\tau \in (0, 1)$, using the estimates from the previous two steps.

If the interest is on only the median function then Step 1 is enough. I.e, assuming **H1**, the conditional median function of the model in Equation (1.7) is given by:

$$\begin{aligned} q_{0.5}(Y(T)|\mathbf{X}(T), T) &= \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T) + V(\mathbf{X}(T), T)q_{0.5}(\epsilon(T)|\mathbf{X}(T), T) \\ &= \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T). \end{aligned} \tag{2.2}$$

However, to study other quantiles as well, the other steps need to be considered. This means that a correct specification of the variability function and Assumption **H2** are only important for the quantiles other than the median.

Remark 2.1 *Instead of the model in Equation* (1.7)*, one may consider the following quantile varying coefficient model (Tang et al., 2013b):*

$$q_\tau(Y(T)|\mathbf{X}(T), T) = \mathbf{X}^T(T)\boldsymbol{\beta}_\tau(T), \tag{2.3}$$

where the varying coefficients are allowed to vary with the τth quantile. However, for the model in Equation (2.3)*, we need to assume $q_\tau(\epsilon(T)|\mathbf{X}(T), T) = 0$ for the τ considered. Where in the AHe approach, only the two Assumptions* **H1** *and* **H2** *are needed.*

2.1.1 Step 1: estimation of the median

Under Assumption **H1**, the median function of the model in Equation (1.7) is given by:

$$q_{0.5}(Y(T)|\mathbf{X}(T),T) = \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T).$$

The coefficient functions $\beta_k(T)$ can be approximated via normalized B-splines of degree ν_k, for $k = 0, 1, \ldots, p$:

$$\begin{aligned}\beta_k(T) &\approx \alpha_{k1}B_{k1}(T,\nu_k) + \alpha_{k2}B_{k2}(T,\nu_k) + \ldots + \alpha_{km_k}B_{km_k}(T,\nu_k)\\ &= \sum_{l=1}^{m_k}\alpha_{kl}B_{kl}(T,\nu_k) = \boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{B}_k(T,\nu_k),\end{aligned} \tag{2.4}$$

where $\boldsymbol{\alpha}_k = (\alpha_{k1},\ldots,\alpha_{km_k})^{\mathrm{T}}$ denotes the coefficient vector of the associated B-splines $\mathbf{B}_k(T,\nu_k) = (B_{k1}(T,\nu_k),\ldots,B_{km_k}(T,\nu_k))^{\mathrm{T}}$, $m_k = u_k + \nu_k$, and $u_k + 1$ represents the number of equidistant knots in the B-splines approximation of the kth component $(X^{(k)}(T))$. A thorough discussion on the choice of the number of knots $u_k + 1$ and the degree of the splines ν_k can be found in Gijbels and Verhasselt (2010). The coefficient function $\beta_k(T)$ is estimated by,

$$\hat{\beta}_k(T) = \hat{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}_k(T,\nu_k), \tag{2.5}$$

where $\hat{\boldsymbol{\alpha}}_k = (\hat{\alpha}_{k1},\ldots,\hat{\alpha}_{km_k})^{\mathrm{T}}$ $(k = 0,...,p)$, and $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\alpha}}_0^{\mathrm{T}},\ldots,\hat{\boldsymbol{\alpha}}_p^{\mathrm{T}})^{\mathrm{T}}$ is obtained by minimizing the following P-splines objective function, using an L_1-type of penalty function (Andriyana et al., 2014):

$$\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(Y_{ij} - \mathbf{U}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}\right) + \lambda_P\sum_{k=0}^{p}\omega_{1k}\sum_{l=d_k+1}^{m_k}|\Delta^{d_k}\alpha_{kl}| \tag{2.6}$$

with respect to $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^{\mathrm{T}},\ldots,\boldsymbol{\alpha}_p^{\mathrm{T}})^{\mathrm{T}}$, where $\mathbf{U}_{ij} = \left(\left(\mathbf{U}_{ij}^{(0)}\right)^{\mathrm{T}},\ldots,\left(\mathbf{U}_{ij}^{(p)}\right)^{\mathrm{T}}\right)^{\mathrm{T}}_{m_{tot}\times 1}$ with $\mathbf{U}_{ij}^{(k)} = \mathbf{B}_k(t_{ij},\nu_k)X_{ij}^{(k)}$ and $m_{tot} = \sum_{k=0}^{p}m_k$; $\lambda_P > 0$ is the smoothing parameter, $\omega_{1k} = \mathrm{Range}(\hat{\beta}_k^B(T))^{-\eta_P}$; $\mathrm{Range}(\hat{\beta}_k^B(T))$ is the range of quantile regression estimators obtained using unpenalized B-splines (the minimizer of Equation (2.6) with $\lambda_P = 0$); $\eta_P \geq 0$; d_k is the differencing order in the penalty term; and Δ^{d_k} is the d_kth order differencing operator of the kth variable: $\Delta^{d_k}\alpha_{kl} = \sum_{t=0}^{d_k}(-1)^t\binom{d_k}{t}\alpha_{k(l-t)}$. The penalty weight ω_{1k} is used such that the functional coefficients with larger range values are allowed to vary more with T. We could, alternatively, use $\omega_{1k} = \mathrm{Sd}(\hat{\beta}_k^B(T))^{-\eta_P}$, with $\mathrm{Sd}(\hat{\beta}_k^B(T))$ the standard deviation of the regression estimators $\hat{\beta}_k^B(T)$ (Andriyana

et al., 2014).

2.1.2 Step 2: estimation of the variability function

From the model in Equation (1.7) and since $V(\mathbf{X}(T), T) \geq 0$, we have

$$\ln|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)| = \ln V(\mathbf{X}(T), T) + \ln|\epsilon(T)|. \tag{2.7}$$

Using Assumption **H2**, since for a monotone function $h(\cdot)$ the quantiles of the transformed random variable $h(Y)$ are simply the transformed quantiles of the original variable Y (see property 1 in Section 1.1.2), we estimate $V(\cdot,\cdot)$ based on the following equation:

$$q_{0.5}\left(\ln|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)| \,\middle|\, \mathbf{X}(T), T\right) = \ln V(\mathbf{X}(T), T).$$

In this step, we investigate how to estimate the four variability functions. First, the conditional median of $\ln|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)|$ is given by:

V_1: $q_{0.5}\left(\ln|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)| \,\middle|\, \mathbf{X}(T), T\right) = \delta + \theta \ln\{\|\mathbf{X}(T)\|_1\}$, with $\delta = \ln\{\gamma\}$

V_2: $q_{0.5}\left(\ln|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)| \,\middle|\, \mathbf{X}(T), T\right) = \ln\{\gamma(T)\} + \theta \ln\{\|\mathbf{X}(T)\|_1\}$,
where $\ln\{\gamma(T)\}$ is approximated on a basis (of size $m^{\mathrm{V}} = u^{\mathrm{V}} + \nu^{\mathrm{V}}$) of normalized B-splines of degree ν^{V} with $u^{\mathrm{V}} + 1$ equidistant knots.

V_3: $q_{0.5}\left(\ln|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)| \,\middle|\, \mathbf{X}(T), T\right) = \ln\{\gamma(T)\} + \theta\|\mathbf{X}^*(T)\|_1$,
where $\ln\{\gamma(T)\}$ is approximated as in V_2.

V_4:

$$q_{0.5}\left(\ln|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)| \,\middle|\, \mathbf{X}(T), T\right) = \sum_{k=0}^{p} \theta_k(T)|X^{(k)}(T)|, \tag{2.8}$$

with $\theta_0(T) = \ln\{\gamma(T)\}$ and $\theta_k(T)$ is approximated on a basis (of size $m_k^{\mathrm{V}} = u_k^{\mathrm{V}} + \nu_k^{\mathrm{V}}$) of normalized B-splines of degree ν_k^{V} with $u_k^{\mathrm{V}} + 1$ equidistant knots, for $k = 0, 1, \ldots, p$.

The conditional median of $\ln(|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)|)$ is therefore estimated by:

V_1: $\hat{q}_{0.5}\left(\ln|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)| \,\middle|\, \mathbf{X}(T), T\right) = \hat{\delta} + \hat{\theta} \ln\{\|\mathbf{X}(T)\|_1\}$,

V_2: $\hat{q}_{0.5}\left(\ln|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)| \,\middle|\, \mathbf{X}(T), T\right) = \sum_{l=1}^{m^{\mathrm{V}}} \hat{\alpha}_l^{\mathrm{V}} B_l(T, \nu^{\mathrm{V}}) + \hat{\theta} \ln\{\|\mathbf{X}(T)\|_1\}$,

V_3: $\hat{q}_{0.5}\left(\ln|Y(T)-\mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)|\,\middle|\,\mathbf{X}(T),T\right)=\sum_{l=1}^{m^{\mathrm{V}}}\hat{\alpha}_l^{\mathrm{V}}B_l(T,\nu^{\mathrm{V}})+\hat{\theta}\|\mathbf{X}^*(T)\|_1$,

V_4:

$$\hat{q}_{0.5}\left(\ln|Y(T)-\mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)|\,\middle|\,\mathbf{X}(T),T\right)=\sum_{k=0}^{p}\hat{\theta}_k(T)|X^{(k)}(T)|$$
$$=\sum_{k=0}^{p}\sum_{l=1}^{m_k^{\mathrm{V}}}\hat{\alpha}_{kl}^{\mathrm{V}}B_{kl}(T,\nu_k^{\mathrm{V}})|X^{(k)}(T)|. \quad (2.9)$$

The estimated parameters are obtained by solving for

V_1:

$$\min_{\delta,\theta}\left\{\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(\ln(|Y_{ij}-\mathbf{X}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\beta}}(t_{ij})|)-\delta-\theta\ln\{\|\mathbf{X}_{ij}\|_1\}\right)\right\},$$

where $\hat{\boldsymbol{\beta}}(T)=(\hat{\beta}_0(T),\hat{\beta}_1(T),\ldots,\hat{\beta}_p(T))^{\mathrm{T}}$ is the estimated vector of regression coefficients obtained from Step 1, in Section 2.1.1.

V_2:

$$\min_{\alpha_1^{\mathrm{V}},\alpha_2^{\mathrm{V}},\ldots,\alpha_{m^{\mathrm{V}}}^{\mathrm{V}},\theta}\left\{\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(\ln(|Y_{ij}-\mathbf{X}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\beta}}(t_{ij})|)-\sum_{l=1}^{m^{\mathrm{V}}}\alpha_l^{\mathrm{V}}B_l(t_{ij},\nu^{\mathrm{V}})\right.\right.$$
$$\left.\left.-\theta\ln\{\|\mathbf{X}_{ij}\|_1\}\right)+\lambda_P^{\mathrm{V}}\sum_{l=d^{\mathrm{V}}+1}^{m^{\mathrm{V}}}|\Delta^{d^{\mathrm{V}}}\alpha_l^{\mathrm{V}}|\right\},$$

where $\lambda_P^{\mathrm{V}}>0$ is the smoothing parameter, d^{V} is the differencing order in the penalty term.

V_3:

$$\min_{\alpha_1^{\mathrm{V}},\alpha_2^{\mathrm{V}},\ldots,\alpha_{m^{\mathrm{V}}}^{\mathrm{V}},\theta}\left\{\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(\ln(|Y_{ij}-\mathbf{X}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\beta}}(t_{ij})|)-\sum_{l=1}^{m^{\mathrm{V}}}\alpha_l^{\mathrm{V}}B_l(t_{ij},\nu^{\mathrm{V}})\right.\right.$$
$$\left.\left.-\theta\|\mathbf{X}_{ij}^*\|_1\right)+\lambda_P^{\mathrm{V}}\sum_{l=d^{\mathrm{V}}+1}^{m^{\mathrm{V}}}|\Delta^{d^{\mathrm{V}}}\alpha_l^{\mathrm{V}}|\right\},$$

V_4:

$$\min_{\boldsymbol{\alpha}_0^{\mathrm{V}},\ldots,\boldsymbol{\alpha}_p^{\mathrm{V}}} \left\{ \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\Big(\ln(|Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}} \hat{\boldsymbol{\beta}}(t_{ij})|) - (\mathbf{U}_{ij}^{\mathrm{Va}})^{\mathrm{T}} \boldsymbol{\alpha}^{\mathrm{V}} \Big) + \lambda_P^{\mathrm{V}} \sum_{k=0}^{p} \omega_{1k}^{\mathrm{V}} \|\mathbf{D}_{m_k^{\mathrm{V}}}^{d_k^{\mathrm{V}}} \boldsymbol{\alpha}_k^{\mathrm{V}}\|_1 \right\}, \tag{2.10}$$

where $\mathbf{U}_{ij}^{\mathrm{Va}} = \Big((\mathbf{U}^{\mathrm{Va}}{}_{ij}^{(0)})^{\mathrm{T}}, \ldots, (\mathbf{U}^{\mathrm{Va}}{}_{ij}^{(p)})^{\mathrm{T}} \Big)^{\mathrm{T}}_{m_{tot}^{\mathrm{V}} \times 1}$ with $m_{tot}^{\mathrm{V}} = \sum_{k=0}^{p} m_k^{\mathrm{V}}$; $\mathbf{U}^{\mathrm{Va}}{}_{ij}^{(k)} = \mathbf{B}_k^{\mathrm{V}}(t_{ij}, \nu_k^{\mathrm{V}})|X_{ij}^{(k)}|$; $\mathbf{B}_k^{\mathrm{V}}(T, \nu_k^{\mathrm{V}}) = (B_{k1}^{\mathrm{V}}(T, \nu_k^{\mathrm{V}}), \ldots, B_{km_k^{\mathrm{V}}}^{\mathrm{V}}(T, \nu_k^{\mathrm{V}}))^{\mathrm{T}}_{m_k^{\mathrm{V}} \times 1}$; $\boldsymbol{\alpha}^{\mathrm{V}} = \big((\boldsymbol{\alpha}_0^{\mathrm{V}})^{\mathrm{T}}, \ldots, (\boldsymbol{\alpha}_p^{\mathrm{V}})^{\mathrm{T}} \big)_{m_{tot}^{\mathrm{V}} \times 1}$; $\boldsymbol{\alpha}_k^{\mathrm{V}} = (\alpha_{k1}^{\mathrm{V}}, \ldots, \alpha_{km_k^{\mathrm{V}}}^{\mathrm{V}})^{\mathrm{T}}_{m_k^{\mathrm{V}} \times 1}$; $\omega_{1k}^{\mathrm{V}} = \mathrm{Range}(\hat{\theta}_k^B(T))^{-\eta_P^{\mathrm{V}}}$ with $\eta_P^{\mathrm{V}} \geq 0$ and $\hat{\theta}_k^B(T)$ is the estimated coefficient using unpenalized B-splines; d_k^{V} is the differencing order in the penalty term; and $\mathbf{D}_{m_k}^{d_k^{\mathrm{V}}}$ is the matrix-representation of the differencing operator $\Delta^{d_k^{\mathrm{V}}}$,

$$\text{for } d_k^{\mathrm{V}} = 1,\ \mathbf{D}_{m_k^{\mathrm{V}}}^{1} = \begin{pmatrix} 1 & -1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(m_k^{\mathrm{V}}-1) \times m_k^{\mathrm{V}}}.$$

Finally, the variability function is estimated by:

V_1: $\hat{V}(\mathbf{X}(T), T) = \hat{\gamma} \|\mathbf{X}(T)\|_1^{\hat{\theta}} = \exp\{\hat{\delta}\} \|\mathbf{X}(T)\|_1^{\hat{\theta}}$,

V_2: $\hat{V}(\mathbf{X}(T), T) = \hat{\gamma}(T) \|\mathbf{X}(T)\|_1^{\hat{\theta}} = \exp\{\sum_{l=1}^{m^{\mathrm{V}}} \hat{\alpha}_l^{\mathrm{V}} B_l(T, \nu^{\mathrm{V}})\} \|\mathbf{X}(T)\|_1^{\hat{\theta}}$,

V_3: $\hat{V}(\mathbf{X}(T), T) = \hat{\gamma}(T) \exp\{\hat{\theta} \|\mathbf{X}^*(T)\|_1\} = \exp\Big\{ \sum_{l=1}^{m^{\mathrm{V}}} \hat{\alpha}_l^{\mathrm{V}} B_l(T, \nu^{\mathrm{V}}) + \hat{\theta} \|\mathbf{X}^*(T)\|_1 \Big\}$,

V_4: $\hat{V}(\mathbf{X}(T), T) = \hat{\gamma}(T) \exp\{\sum_{k=1}^{p} \hat{\theta}_k(T) |X^{(k)}(T)|\} = \exp\big\{ \sum_{k=0}^{p} \hat{\theta}_k(T) |X^{(k)}(T)| \big\}$.

2.1.3 Step 3: estimation of the τth quantile

The model in Equation (1.7) yields

$$q_\tau(Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T) | \mathbf{X}(T), T) = V(\mathbf{X}(T), T) q_\tau(\epsilon(T) | \mathbf{X}(T), T).$$

Using the estimators $\hat{\boldsymbol{\beta}}(T)$ and $\hat{V}(\mathbf{X}(T), T)$ from Steps 1 and 2, respectively, we create a pseudo-response $Y(T) - \mathbf{X}^{\mathrm{T}}(T)\hat{\boldsymbol{\beta}}(T)$ and a pseudo-covariate $\hat{V}(\mathbf{X}(T), T)$.

We approximate the unknown conditional τth quantile of the error term $\epsilon(T)$, $q_\tau(\epsilon(T)|\mathbf{X}(T),T)$, by B-spline basis functions of degree ν^q with u^q+1 equidistant knots, leading to $q_\tau(\epsilon(T)|\mathbf{X}(T),T) \approx \sum_{l=1}^{m^q} \alpha_l^q B_l^q(T,\nu^q)$. We hereby also assume that $q_\tau(\epsilon(T)|\mathbf{X}(T),T)$ satisfies the Hölder condition: $q_\tau(\epsilon(T)|\mathbf{X}(T),T) \in \mathcal{H}_{r^q}$ for some $r^q > 1/2$ (see Definition 2.2 in Section 2.6.1). The coefficients $(\alpha_1^q, \alpha_2^q, ..., \alpha_{m^q}^q)$ are estimated by minimizing the following P-splines objective function:

$$\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_\tau \left(Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}} \hat{\boldsymbol{\beta}}(t_{ij}) - \hat{V}(\mathbf{X}_{ij}, t_{ij}) \sum_{l=1}^{m^q} \alpha_l^q B_l^q(t_{ij}, \nu^q) \right) + \sum_{l=d^q+1}^{m^q} \lambda^q |\Delta^{d^q} \alpha_l^q|$$

with respect to $(\alpha_1^q, \alpha_2^q, ..., \alpha_{m^q}^q)$, where $\lambda^q > 0$ is the smoothing parameter, and d^q is the differencing order.

The τth quantile of the error is then estimated by: $\hat{q}_\tau(\epsilon(T)|\mathbf{X}(T),T) = \sum_{l=1}^{m^q} \hat{\alpha}_l^q B_l^q(T,\nu^q)$. Finally, the estimated conditional quantile function of the response $Y(T)$ is given by:

$$\hat{q}_\tau(Y(T)|\mathbf{X}(T),T) = \mathbf{X}^{\mathrm{T}}(T)\hat{\boldsymbol{\beta}}(T) + \hat{V}(\mathbf{X}(T),T)\hat{q}_\tau(\epsilon(T)|\mathbf{X}(T),T).$$

Recall that based on Assumption **H1**, we obtain the coefficient estimates $\hat{\boldsymbol{\beta}}(T)$. Then, based on Assumption **H2**, we get the estimated variability function $\hat{V}(\mathbf{X}(T),T)$. Finally, using the estimates obtained from the previous two steps ($\hat{\boldsymbol{\beta}}(T)$ and $\hat{V}(\mathbf{X}(T),T)$), we can easily get $\hat{q}_\tau(Y(T)|\mathbf{X}(T),T)$ by estimating $\hat{q}_\tau(\epsilon(T)|\mathbf{X}(T),T)$.

2.1.4 Implementation of the estimation method

Due to the non-differentiability of the optimization problem in Equation (2.6), the estimators have no explicit solutions. However, one can use available LP optimization techniques (Koenker and Bassett, 1978) like FN algorithm (Portnoy and Koenker, 1997) or a convex programming optimization algorithms like a Matlab-based modeling system called CVX (Grant and Boyd, 2012). We apply the FN algorithm, used in Andriyana et al. (2014), which is faster than CVX and efficient even for very large problems, particularly when dealing with sparse matrices (see also Koenker and Ng,

2003, 2005). The optimization problem can be written as:

$$\min_{\text{all } u_{ij},v_{ij},s_{0l},\ldots,s_{pl},t_{0l},\ldots,t_{pl}} \Big\{0.5\sum_{i=1}^{n}\sum_{j=1}^{N_i} u_{ij} + (1-0.5)\sum_{i=1}^{n}\sum_{j=1}^{N_i} v_{ij} + \lambda_P \sum_{k=0}^{p} \omega_{1k} \sum_{l=d_k+1}^{m_k} (s_{kl}+t_{kl})\Big\}$$

subject to

$$\begin{aligned} \mathbf{U}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}/N_i + u_{ij} - v_{ij} &= Y_{ij}/N_i \quad \text{for all } i = 1,\ldots,n \text{ and } j = 1,\ldots,N_i, \\ \Delta^{d_k}\alpha_{kl} - s_{kl} + t_{kl} &= 0 \quad \text{for all } l = d_k+1,\ldots,m_k \text{ and } k = 0,\ldots,p, \end{aligned}$$

where $u_{ij} \geq 0$ and $v_{ij} \geq 0$ are the positive and negative regression residuals, respectively; $s_{kl} = |\Delta^{d_k}\alpha_{kl}|I(\Delta^{d_k}\alpha_{kl} \geq 0)$; and $t_{kl} = |\Delta^{d_k}\alpha_{kl}|I(\Delta^{d_k}\alpha_{kl} < 0)$.

The FN algorithm solves the dual formulation of the above LP problem.

The optimization problem in (Step 2) can also be written in a similar way by changing appropriately $\Big[\mathbf{U}_{ij}, u_{ij}, v_{ij}, s_{kl}, t_{kl}, \lambda_k, d_k, m_k, \alpha_{kl}, Y_{ij}\Big]$.

The R-code to implement the estimation method is given in Appendix A and is available in QRegVCM R-package (Andriyana et al., 2018).

2.1.5 Choice of the smoothing parameter

In the first step of our estimation procedure, the optimal value for λ_P is found by minimizing the following Schwarz-type Information Criterion (SIC-criterion), introduced by Schwarz (1978), suggested by Koenker et al. (1994) and also applied by Andriyana et al. (2014):

$$\mathrm{SIC}(\lambda_P) = \ln\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}(Y_{ij} - \hat{q}_{0.5}(Y_{ij}|\mathbf{X}_{ij}, t_{ij}))\Big) + \frac{\ln(N)}{2N}p_{\lambda_P} \tag{2.11}$$

with respect to λ_P, where $N = \sum_{i=1}^{n} N_i$ and p_{λ_P} is the effective degrees of freedom and taken to equal the size of the elbow set $\mathcal{E}_{\lambda_P}$ (Koenker, 2011),

$$\mathcal{E}_{\lambda_P} = \{(i,j) : Y_{ij} - \hat{q}_{0.5}(Y_{ij}|\mathbf{X}_{ij}, t_{ij}) = 0\}. \tag{2.12}$$

For the last two steps of our procedure, a similar approach is considered.

2.1.6 Consistency of the estimated variability function for V_4

In this section, we investigate the consistency of the proposed estimation procedure for V_4. We show that the P-splines quantile estimator for V_4 converges in L_2-norm to the unknown quantile function $q_{0.5}\left(\ln|Y(T)-\mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)|\,\middle|\,\mathbf{X}(T),T\right)=\sum_{k=0}^{p}\theta_k(T)|X^{(k)}(T)|$, by proving that each $\hat{\theta}_k(T)$ converges in L_2-norm to the function $\theta_k(T)$.

The following theorem, taken from Theorem 2.1 in Andriyana et al. (2014), states that each individual estimate $\hat{\beta}_k$ (in the signal part of the model), defined in Equation (2.5), converges in L_2-norm to the regression coefficient function β_k for an increasing number of knots. The definitions and assumptions are stated in Section 2.6.1.

Theorem 2.1 *Suppose Assumptions* **2.1** – **2.7** *hold. Furthermore, assume that* $u_{\max}^{3/4}\lambda_{P\max}n^{-1/2}\to 0$ *as* n *tends to* ∞, *where* $u_{\max}=\max(u_0,\ldots,u_p)$ *and* $\lambda_{P\max}=\max(\lambda_{P0},\ldots,\lambda_{Pp})$ *with* $\lambda_{Pk}=\lambda_P\omega_{1k}$. *Then,* $\|\hat{\beta}_k-\beta_k\|_2=O_p(n^{-r_{\min}/(2r_{\min}+1)})$, *where for any function* β_k: $\|\beta_k\|_2^2=\int_{\mathcal{T}}\beta_k(t)^2dt$ *and* $r_{\min}=\min\{r_0,\ldots,r_p\}$, *with* r_k *as defined in Definition 2.2.*

In the second step of our procedure, we assume that $\ln|\epsilon(T)|$ has median zero, given $(\mathbf{X}(T),T)$. Note that relying on Theorem 2.1, we have:

$$\begin{aligned}\ln|Y(T)-&\mathbf{X}^{\mathrm{T}}(T)\{\boldsymbol{\beta}(T)+[\hat{\boldsymbol{\beta}}(T)-\boldsymbol{\beta}(T)]\}|\\&=\ln|Y(T)-\mathbf{X}^{\mathrm{T}}(T)\{\boldsymbol{\beta}(T)+O_p(n^{-r_{\min}/(2r_{\min}+1)})\mathbf{1}_{p+1}\}|\\&\xrightarrow{P}\ln|Y(T)-\mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)|,\quad\text{as } n\to\infty,\end{aligned}\tag{2.13}$$

where $\mathbf{1}_{p+1}$ is the vector of ones of dimension $p+1$.

Our main result on the consistency of the θ_k, in the variability part of the model, follows. The proof is provided in Section 2.6.

Theorem 2.2 *Suppose Assumptions* **2.6** – **2.13** *hold. Furthermore, assume that* $(u_{\max}^V)^{3/4}\lambda_{P\max}^V n^{-1/2}\to 0$ *as* n *tends to* ∞, *where* $u_{\max}^V=\max(u_0^V,\ldots,u_p^V)$ *and* $\lambda_{P\max}^V=\max(\lambda_{P0}^V,\ldots,\lambda_{Pp}^V)$ *with* $\lambda_{Pk}^V=\lambda_P^V\omega_{1k}^V$. *Then,*

$$\|\hat{\theta}_k^{(j)}-\theta_k^{(j)}\|_2=O_p(n^{-(r_{\min}^V-j)/(2r_{\min}^V+1)}),\qquad j=0,1,\ldots,\nu_k^V-1,$$

where $\theta_k^{(j)}$ *is the* j*th order derivative of* θ_k *and* $r_{\min}^V=\min\{r_0^V,\ldots,r_p^V\}$, *where* $r_k^v>1/2$ *as in Assumption* **2.8**.

The convergence rates in Theorem 2.2 attain the optimal global rates established by Stone (1982) for non-parametric regression with independent and identically distributed data. Indeed if θ_k has bounded second order derivative and splines of at most degree 3 are used, the convergence rate is $n^{-2/5}$.

The proof of Theorem 2.2 is based on the fact that the number of knots does not grow too fast with the sample size (see for example Assumption **2.10**). This is similar to the approach of Claeskens et al. (2009) which leads to a scenario close to regression splines.

2.2 Model testing

Section 2.1.2 explains how to estimate a specific variability function. This section discusses how to identify the specific shape of the variability function.

2.2.1 Variability function

The following hypothesis tests are considered:

$$H_0 : V(\mathbf{X}(T), T) \text{ in } V_j \text{ versus}$$
$$H_1 : V(\mathbf{X}(T), T) \text{ in } V_k$$

where $V_j \subset V_k$.

We use the Likelihood-Ratio-Type test (LRT) considered by Kim (2007). The test statistic is defined as:

$$G = 2\left\{\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(|Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\beta}}(t_{ij})| - \hat{m}_0\right) - \sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(|Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\beta}}(t_{ij})| - \hat{m}_1\right)\right\},$$

where $\hat{m}_0$ and $\hat{m}_1$ are the estimated variability functions under H_0 and under H_1, respectively. It is expected, when H_0 is false, that $\rho_{0.5}\left(|Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\beta}}(t_{ij})| - \hat{m}_0\right)$ is larger than $\rho_{0.5}\left(|Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\beta}}(t_{ij})| - \hat{m}_1\right)$. Therefore, with test statistic G, we assess the generality of the variability function. We consider the four structures of the variability function, V_1 to V_4, as well as:

V_0: $V(\mathbf{X}(T), T) = \gamma$ (the homoscedastic model);

V_5: $V(\mathbf{X}(T), T) = \gamma(T)$ (the simple heteroscedastic model).

Note that these six variability structures are "nested" in the following sense:

$$V_0 \subseteq (V_1, V_5) \subseteq V_2 \underset{\sim}{\subset} V_3 \subseteq V_4,$$

where $\underset{\sim}{\subset}$ means that V_3 is a generalization of V_2 in the sense that a first order approximation of the exponential function $\exp(\|\mathbf{X}^*(T)\|_1)$ leads to $\|\mathbf{X}(T)\|_1$. So, V_4 is considered the "full model" here.

The null hypothesis is rejected when G is too large. The p-value is obtained using a resampling subject bootstrap. Since the data are longitudinal in nature and it is important to retain the dependence structure within a subject, the observations in each subject are treated as blocks. The procedure is given below.

1. Resample n subjects with replacement from $i = 1, \ldots, n$, to obtain the bootstrap sample $\{(Y^b(t_{ij}^b), \mathbf{X}^b(t_{ij}^b), t_{ij}^b) : i = 1, \ldots, n, j = 1, \ldots, N_i^b\}$ from $\{(Y^p(t_{ij}), \mathbf{X}(t_{ij}), t_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N_i\}$, with

$$Y^p(t_{ij}) = \sum_{k=0}^{p} \hat{\beta}_k(t_{ij}) X^{(k)}(t_{ij}) + \hat{m}_0 \epsilon^p(t_{ij}),$$
$$\epsilon^p(t_{ij}) = \frac{Y(t_{ij}) - \sum_{k=0}^{p} \hat{\beta}_k(t_{ij}) X^{(k)}(t_{ij})}{\hat{m}_4},$$

   where $\hat{m}_0$ and $\hat{m}_4$ are the estimated variability functions under H_0 and under the full model (V_4), respectively.

2. Repeat the above sampling procedure B times, where B is the number of bootstrap replications.

3. Obtain the test statistic G^b for each bootstrap sample to obtain its empirical distribution.

4. Get the p-value using the empirical probability of $G^b \geq G$.

The above testing procedure is implemented in the QRegVCM R-package (Andriyana et al., 2018). The R-code is also available in Appendix A.

2.2.2 Extensions

V_4 is the most general model for the variability function considered here. It has a varying coefficient for each covariate $X^{(1)}, \ldots, X^{(p)}$. Hence, it can be of interest to investigate whether the coefficients are constant (or even not significant). This again

can be tested in a similar way to that presented in Section 2.2.1, using the LRT, by putting the reduced model (where the varying coefficients are assumed constant or not significant) in H_0 and the full model (V_4) in H_1 (see Chapter 3).

2.3 Simulation study

In this section, we conduct simulation studies to compare the various variability structures discussed in the previous section. Four simulation settings, defined in Table 2.1, are used, assuming the variability function structures V_1, V_2, V_3 and V_4. These are referred to as Settings 1, 2, 3 and 4 here, with $\theta = 2$. We also consider a simulation setting, called Setting 5, which combines Settings 1 and 2 to check the robustness of our testing procedure.

Table 2.1. Description of the coefficients.

Coefficients	Setting 1	Setting 2	Setting 3	Setting 4
$\beta_0(T)$	$2\sqrt{T}$	$2\sqrt{T}$	$2\sqrt{T}$	$2\sqrt{T}$
$\beta_1(T)$	$\frac{2}{5}(\pi T + 10)$	$\frac{5}{4}(\pi T + 10)$	$\pi T + 10$	$\frac{5}{4}(\pi T + 10)$
$\beta_2(T)$	$\sin(\pi T/30) + 3$	$\sin(\pi T/30) + 3$	$\sin(\pi T/30) + 3$	$\sin(\pi T/30) + 3$
$\beta_3(T)$	$\frac{(20-T)^2}{1000} - 4$	$(20 - T)^2 - 4$	$(20 - T)^2 - 4$	$\frac{(20-T)^2}{1000} - 4$
$\gamma(T)$	1	$T/4$	$T/8$	$\beta_0(T)/8$
$\theta_1(T)$				$\sqrt{\beta_1(T)}/2$
$\theta_2(T)$				$\beta_2(T)/50$
$\theta_3(T)$				$\beta_3(T)/50$

In order to ensure that the estimation tasks are of comparable difficulty in the four simulation settings, the coefficient functions are such that the Signal-to-Noise Ratio (SNR) is approximately seven (as in Andriyana and Gijbels, 2017) in each setting:

$$\text{SNR} = \frac{\text{Sample variance of } \sum_{k=0}^{3} \beta_k(T) X^{(k)}(T)}{\text{Sample variance of } V(\mathbf{X}(T), T)\epsilon(T)} \approx 7.$$

The error term is generated from a transformed multivariate normal distribution to ensure that **H1** and **H2** hold. First, for $i = 1, \ldots, n$, we generate $(\zeta_{i1}, \ldots, \zeta_{iN_i})$ from $N(\mathbf{0}, \mathbf{C})$, a multivariate normal with zero mean vector and covariance matrix $\mathbf{C}$, whose

element on the jth row and zth column is given by $cov(\zeta_{ij}, \zeta_{iz}) = 30\exp(-|j-z|))$. Then, the error term $\epsilon(T)$ is obtained from the following transformation,

$$\epsilon_{ij} = \frac{\zeta_{ij} - \hat{q}_{0.5}(\zeta_{ij})}{\hat{q}_{0.5}(|\zeta_{ij} - \hat{q}_{0.5}(\zeta_{ij})|)}, \tag{2.14}$$

where $\hat{q}_{0.5}(\zeta_{ij})$ is the sample median of ζ_{ij}, for $i = 1, \ldots, n$ and $j = 1, \ldots, N_i$.

The covariate $X^{(1)}(T)$ is generated from a standard exponential distribution, $X^{(2)}(T)$ is generated from a standard normal distribution, and $X^{(3)}(T)$ is generated from a uniform distribution $U[-1, 1]$. The covariates are standardized in the following way:

$$X^{(k)}(T) = \frac{X^{(k)}(T) - \min(X^{(k)}(T))}{3\{\max(X^{(k)}(T)) - \min(X^{(k)}(T))\}}.$$

This standardization is done to ensure that the exponential function $\exp(\|\mathbf{X}^*(T)\|_1)$ is well approximated by its first order approximation $\|\mathbf{X}(T)\|_1$. We simulate 100 data sets with $n = 100$ from each of the above settings. The time variable ranges from 0 to 49. For each case i, the probability of having a measurement in each time point is 0.6, creating an unbalanced number of measurements (the number of measurement for each subject i is different). The actual time points are calculated by adding a generated value from a $U[0, 0.5]$ to the non-skipped time points.

To analyze the data sets, for each simulation setting, B-splines of degree three with 11 equidistant knots in the time interval and differencing order 1 (similarly as in Andriyana et al., 2014) are used.

As discussed in Section 2.2.1, the six variability models are compared using the LRT hypothesis testing based on bootstrap resampling, with $B = 200$ bootstrap samples for all 100 simulated data sets. Proportions of significant tests are presented using 5% level of significance.

2.3.1 Setting 1

It is expected, under this setting, that the estimated variability structure assuming V_1 performs better than or equivalently to V_0 and V_5, and has an equivalent performance to V_2, V_3 and V_4. Figure 2.1 (a) shows the root approximate integrated square error

(RAISE), for the six aforementioned models, defined by

$$RAISE(\hat{V}^{(s)}(\cdot,\cdot)) = \left(\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}[\hat{V}^{(s)}(\mathbf{X}_{ij},t_{ij}) - V^{(s)}(\mathbf{X}_{ij},t_{ij})]^2\right)^{\frac{1}{2}},$$

for simulation s.

In all figures we refer to a model with structure V_j as "Model j".

It can be seen from Figure 2.1 (a) that V_0 and V_5 have poorer performance than the other models. This is also observed from Figure 2.1 (b), the root approximate integrated square error of the quantiles from all simulated data sets, defined by

$$RAISE(\hat{q}_\tau^{(s)}(\cdot)) = \left(\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}[\hat{q}_\tau^{(s)}(Y_{ij}|\mathbf{X}_{ij},t_{ij}) - q_\tau^{(s)}(Y_{ij}|\mathbf{X}_{ij},t_{ij})]^2\right)^{\frac{1}{2}}$$

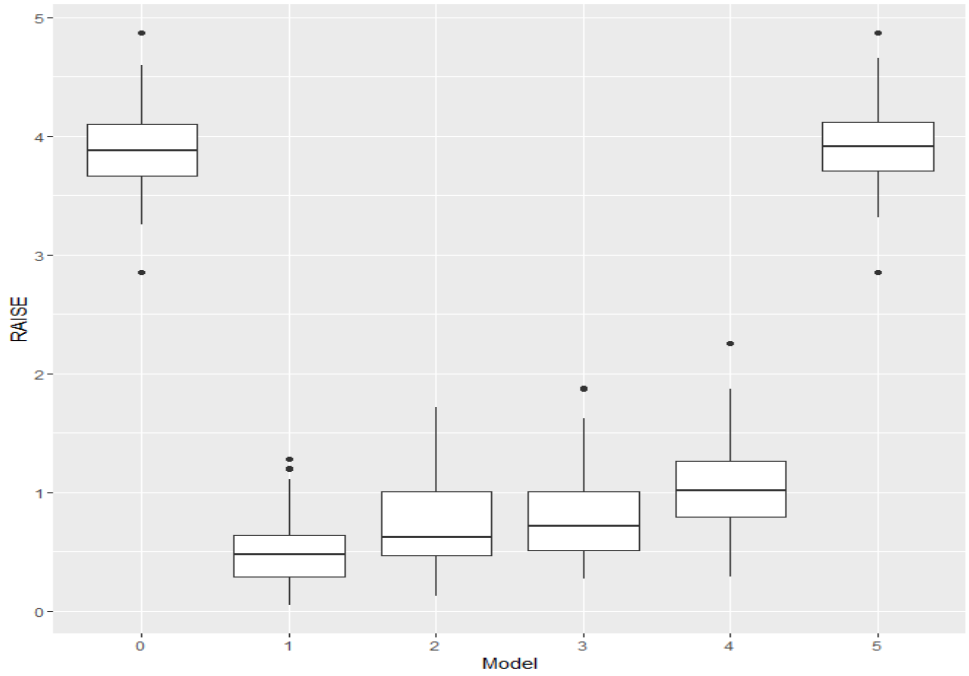
for simulation s.

The true V_1 (Figure 2.2 (a)) as well as the estimated variability function (Figure 2.2 (b)) are presented for the sample associated with a 50th percentile performance in terms of $RAISE(\hat{V}^{(s)}(\cdot,\cdot))$. It is observed that the estimated variability function of V_1 mimics the true variability function. The true and estimated quantiles evaluated at the maximum values of the covariates, using V_1, are presented in Figures 2.2 (c) and (d), respectively. As can be observed from these figures, V_1 mimics the true quantiles quite well.

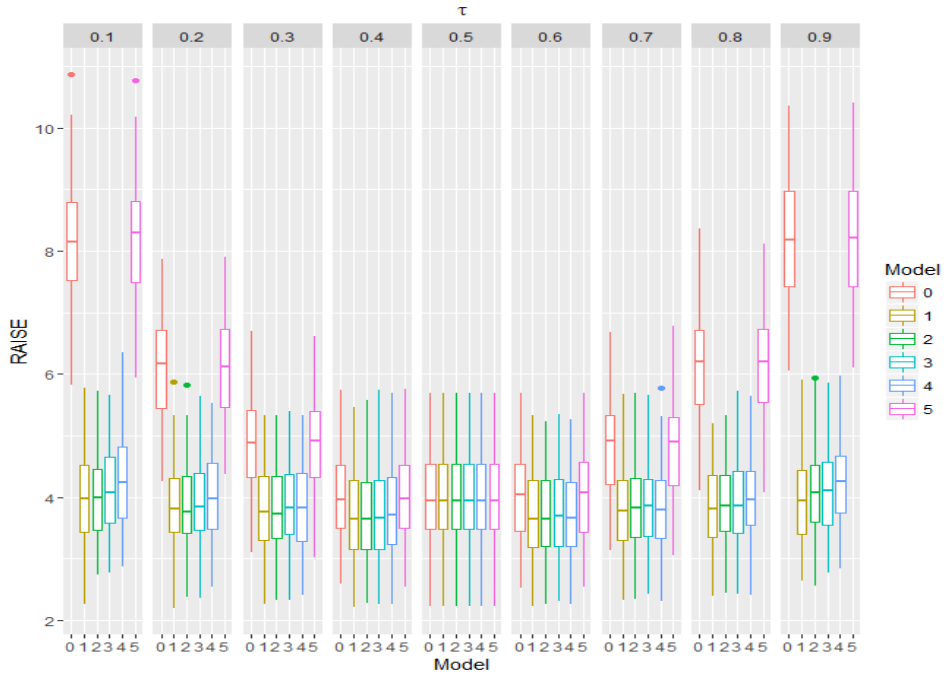
Several comparisons of the models using the LRT test are presented in Table 2.2. The hypothesis for the first testing procedure is:

$$H_0 : V(\mathbf{X}(T),T) \text{ in } V_0 \text{ versus}$$
$$H_1 : V(\mathbf{X}(T),T) \text{ in } V_1.$$

Table 2.2 shows that V_0 is significantly worse than V_1, indicating that θ is different from zero or the variability function depends on the covariates via $\|\mathbf{X}(T)\|_1$. Furthermore, V_1 is not significantly worse than V_2, V_3 or V_4. Table 2.2 also reveals that V_0 is not significantly worse than V_5 and that V_5 is worse than V_2, V_3 or V_4. This indicates that V_5 is also worse than V_1. This result coincides with the simulation result given in Figure 2.1 (a). Similar results are obtained when the SNR is reduced to 4 or 0.5.



(a)



(b)

Figure 2.1. Setting 1. Boxplots of (a) $RAISE(\hat{V}^{(s)})$; (b) $RAISE(\hat{q}_\tau^{(s)})$.

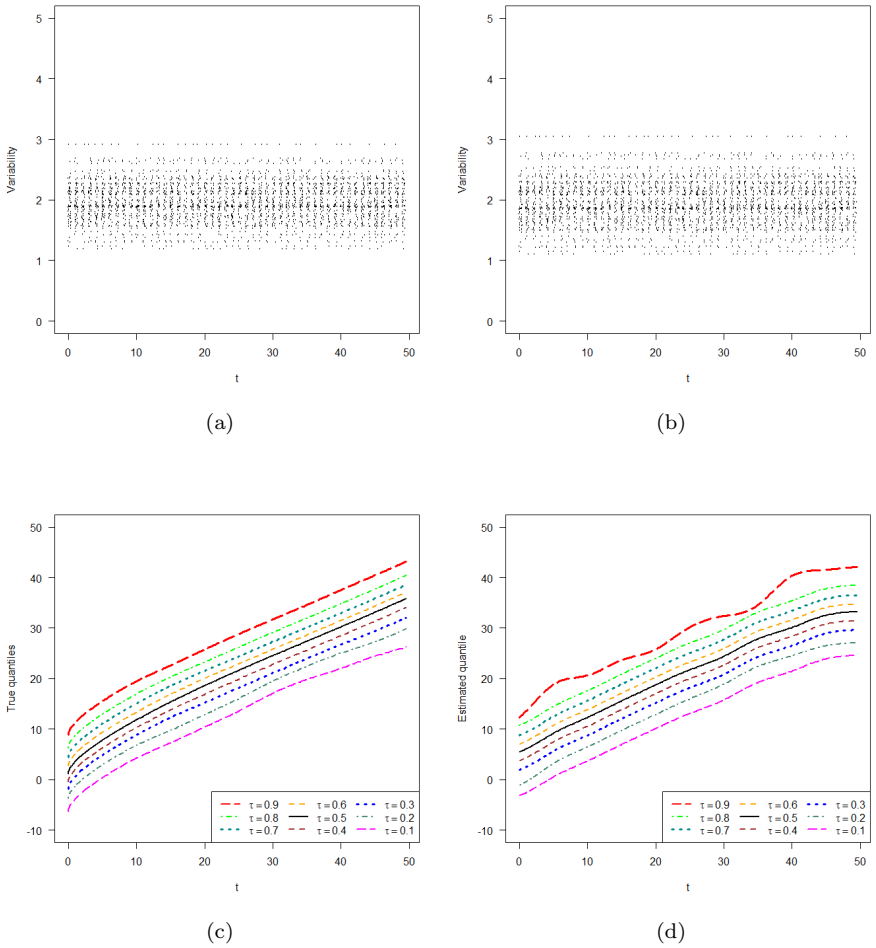


Figure 2.2. Setting 1. (a) True variability function; (b) Estimated variability function based on V_1; (c) True quantile curves; (d) Estimated quantile curves based on V_1.

2.3.2 Setting 2

We investigate the performance of the estimation and testing procedures by simulating from Setting 2 (defined in Table 2.1). Figure 2.3 (a) shows that V_2, V_3 and V_4 have better performance with respect to $\hat{V}(\cdot,\cdot)$. The boxplots for $RAISE(\hat{q}_\tau(\cdot))$ lead to similar conclusions, but are not presented here for brevity.

V_2 is compared to the other models using the LRT test. It is noted from Table 2.3 that

Table 2.2. Setting 1. Proportion of significant tests.

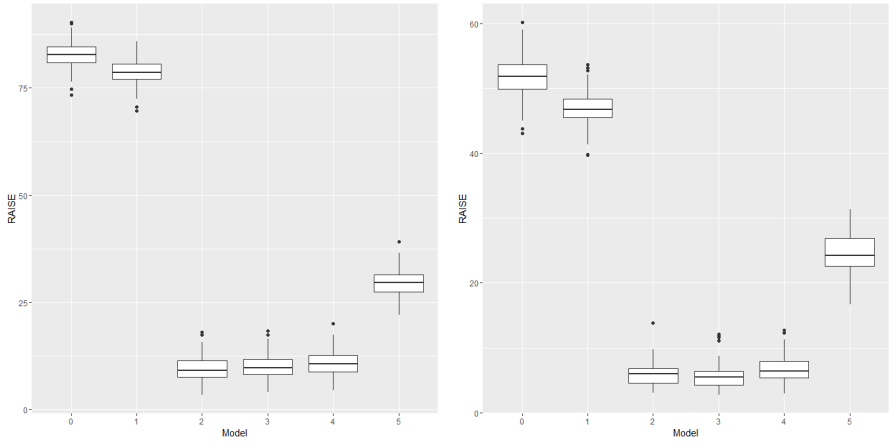
Comparison	Proportion		
	SNR ≈ 7	SNR ≈ 4	SNR ≈ 0.5
V_0 vs V_1	1.00	1.00	1.00
V_1 vs V_2	0.02	0.00	0.00
V_1 vs V_3	0.01	0.00	0.01
V_1 vs V_4	0.02	0.02	0.02
V_0 vs V_5	0.02	0.00	0.01
V_5 vs V_2	1.00	1.00	1.00
V_5 vs V_3	1.00	1.00	1.00
V_5 vs V_4	1.00	1.00	1.00

the power, namely the probability P(H_0 is rejected $|\neg H_0$) is close to one. Furthermore, the probability P(H_0 is rejected $|H_0$) is less than the nominal significance level (5%). We also conclude that V_0, V_1 and V_5 are significantly worse than V_2, indicating that θ is different from zero and the variability function depends on time T. However, V_2 is not significantly worse than V_3 and V_4. This coincides with Figure 2.3 (a). Similar results are obtained when the SNR is reduced to 4 or even to 0.5 (the last two columns of Table 2.3) and when it is 16. We see that for large values of SNR (SNR ≈ 7 or SNR ≈ 16), the power is less than one when comparing V_5 with V_2.

2.3.3 Setting 3

We simulate data from Setting 3 (defined in Table 2.1). As shown in Figure 2.3 (b), V_2, V_3 and V_4 are found to have better performance in estimating the variability function. A similar conclusion can be drawn from the boxplots (not presented here) of $RAISE(\hat{q}_\tau(\cdot))$.

As in the previous section, the LRT test is used on this setting (presented in Table 2.4). V_0, V_1 and V_5 are found to be significantly worse than V_3. However, V_2 is not significantly worse than V_3, and this model on its turn is not significantly worse than V_4. This coincides with the simulation result given in Figure 2.3 (b). Here also, the



(a) (b)

Figure 2.3. Setting 2 (a) and 3 (b). Boxplots of $RAISE(\hat{V}^{(s)})$.

Table 2.3. Setting 2. Proportion of significant tests.

Comparison	Proportion			
	SNR ≈ 16	SNR ≈ 7	SNR ≈ 4	SNR ≈ 0.5
V_0 vs V_2	1.00	1.00	1.00	1.00
V_1 vs V_2	1.00	1.00	1.00	1.00
V_5 vs V_2	0.99	0.98	1.00	1.00
V_2 vs V_3	0.01	0.01	0.02	0.01
V_2 vs V_4	0.02	0.02	0.02	0.02

power of the test is close to one and the probability P(H_0 is rejected $|H_0$) is less than the nominal significance level. When the SNR reduced to 4, similar results are obtained.

2.3.4 Setting 4

The data sets are simulated from Setting 4. V_4 has the best performance in estimating the quantiles as well as the variability function (see Figures 2.4 (a) and (b)). From Figure 2.5, it is clear that the true variability and quantile functions are estimated

Table 2.4. Setting 3. Proportion of significant tests.

Comparison	Proportion	
	SNR ≈ 7	SNR ≈ 4
V_0 vs V_3	1.00	1.00
V_1 vs V_3	1.00	1.00
V_2 vs V_3	0.01	0.02
V_5 vs V_3	1.00	1.00
V_3 vs V_4	0.01	0.01

Table 2.5. Setting 4. Proportion of significant tests.

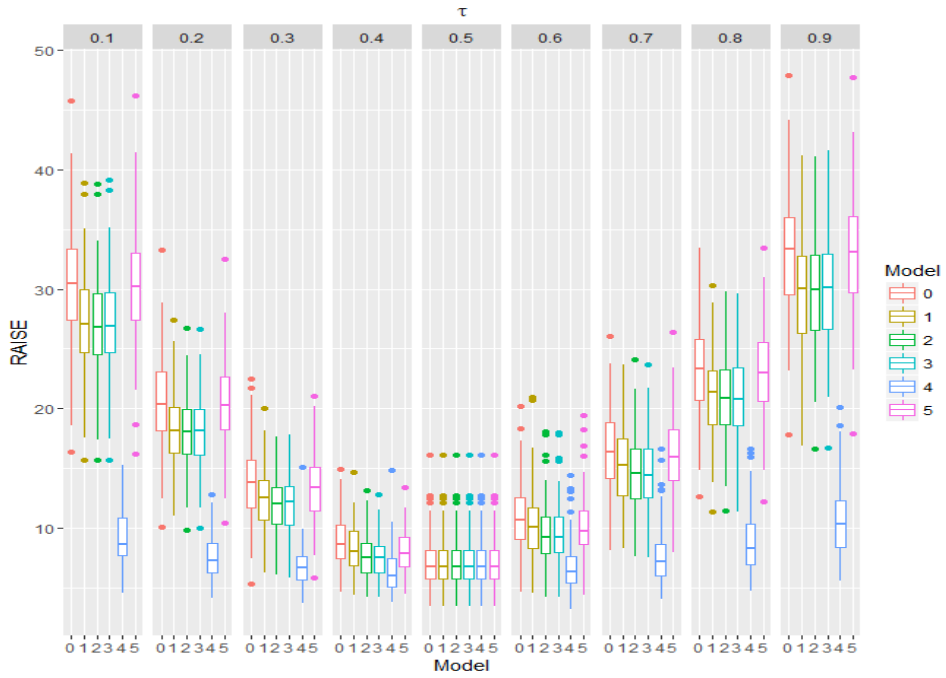
Comparison	Proportion		
	SNR ≈ 7	SNR ≈ 4	SNR ≈ 2
V_0 vs V_4	1.00	1.00	1.00
V_1 vs V_4	1.00	1.00	1.00
V_2 vs V_4	1.00	1.00	1.00
V_3 vs V_4	1.00	1.00	1.00
V_5 vs V_4	1.00	1.00	1.00

quite well using V_4.

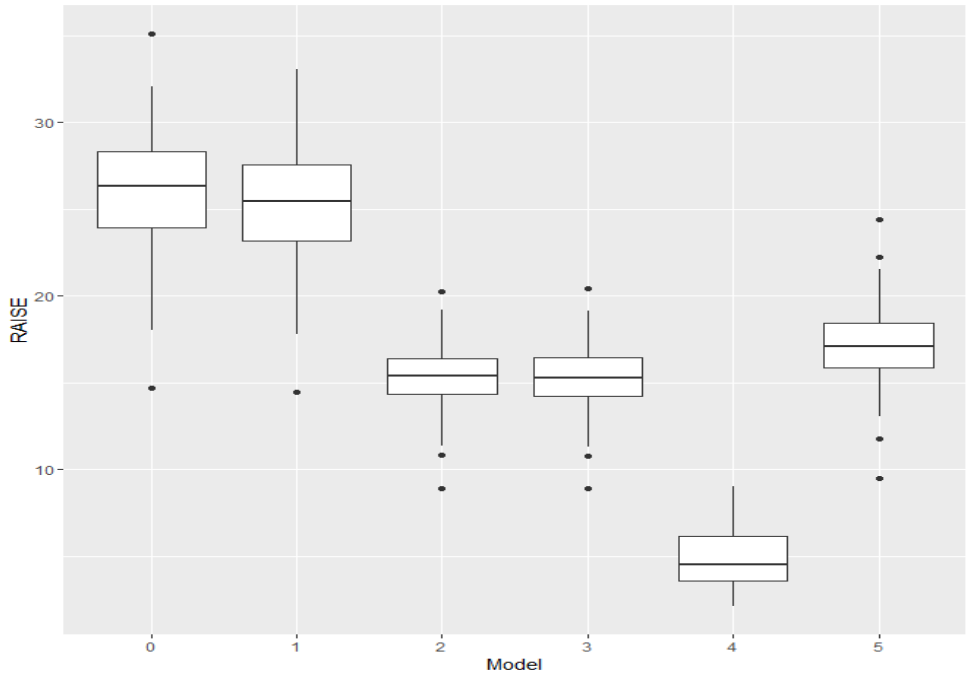
Using the LRT test, V_4 is compared to the other models. Table 2.5 shows that all the other models are worse than V_4 with power equal to one. This was expected since V_4 is the most flexible model. This coincides with the simulation result given in Figure 2.4. Table 2.5 also shows that when we reduce the SNR to 4 or 2, our test still maintains its power.

2.3.5 Setting 5

We simulate data from Setting 2 with SNR $\approx$ 4, where the variability function $V(\mathbf{X}(T), T)$ has a structure which is a linear combination of the structures V_1 and



(a)



(b)

Figure 2.4. Setting 4. Boxplots of (a) $RAISE(\hat{q}_\tau^{(s)})$; (b) $RAISE(\hat{V}^{(s)})$.

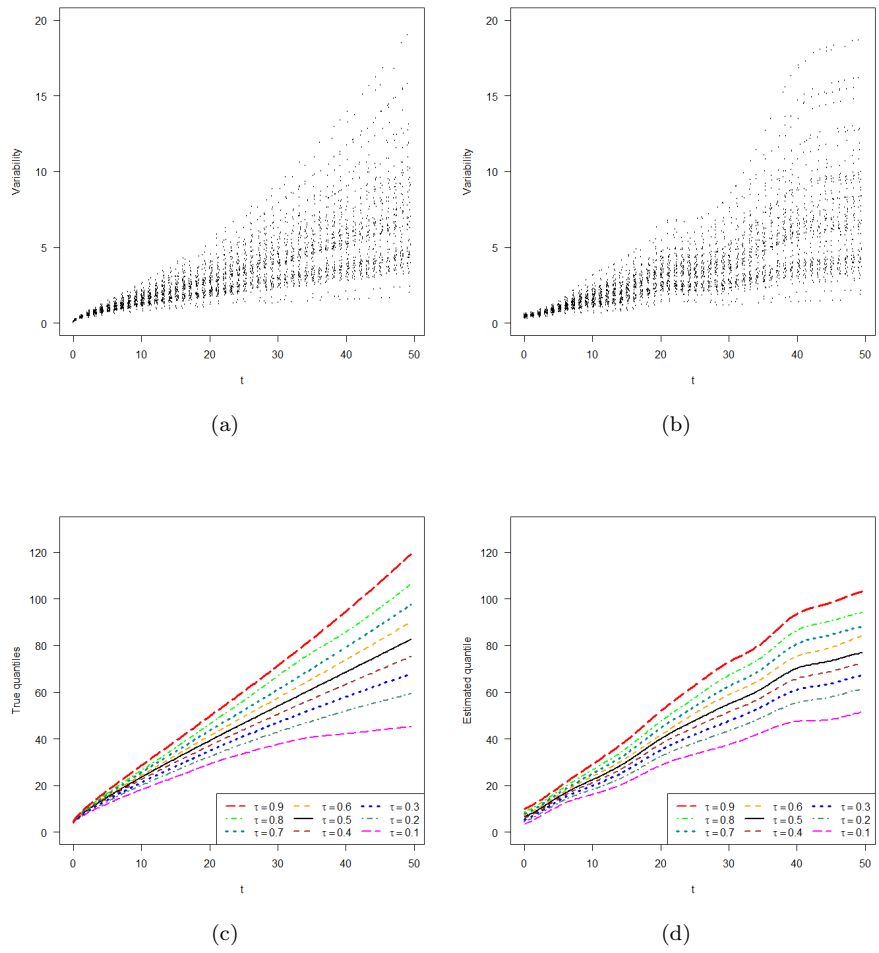
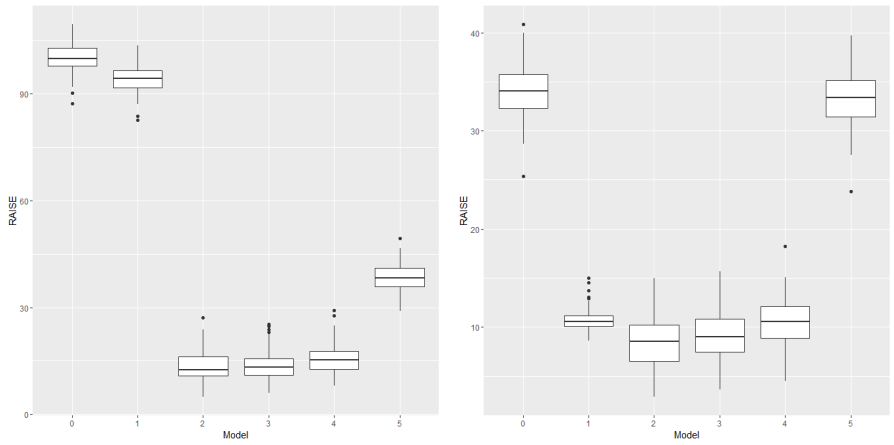


Figure 2.5. Setting 4. (a) True variability function; (b) Estimated variability function based on V_4; (c) True quantile curves; (d) Estimated quantile curves based on V_4.

V_2, i.e. of the form $aV_1 + (1-a)V_2$, with $\gamma = 8.394$ in structure V_1. Under this setting, the variability function is close to either V_1 or V_2 depending on the value of a – for a close to 1, $V(\mathbf{X}(T), T) \approx V_1$. We consider $a = 0.1$ and $a = 0.9$. For $a = 0.1$, Figure 2.6 (a) shows that V_2, V_3 and V_4 have better performance. Where as, for $a = 0.9$, V_1 has also better performance (Figure 2.6 (b)). This result is confirmed by the LRT test given in Table 2.6.



(a) (b)

Figure 2.6. Setting 5. $a = 0.1$ (a) and $a = 0.9$ (b). Boxplots of $RAISE(\hat{V}^{(s)})$.

Table 2.6. Setting 5. Proportion of significant tests.

Comparison	Proportion	
	$a = 0.1$	$a = 0.9$
V_0 vs V_2	1.00	1.00
V_1 vs V_2	1.00	0.06
V_5 vs V_2	0.96	0.96
V_2 vs V_3	0.02	0.09
V_2 vs V_4	0.02	0.05

2.4 Data examples

In this section, the estimation and testing procedures developed in Sections 2.1 and 2.2 are applied to two data sets. To analyze both data sets, B-splines of degree three with 11 equidistant knots in the time interval and differencing of order one are used, similar as in Section 2.3. The covariates are standardized as in Section 2.3. All tests are done at the 5% level of significance.

2.4.1 CD4 data

The data set, which is a subset of the Multicenter AIDS cohort study, contains repeated measurements of physical examinations and CD4 percentages of 283 homosexual men who became HIV-positive between 1984 and 1991. Each individual has a different number of measurements ranging from 1 to 14, with median equals to 6, due to missing of their appointments and the random HIV infection moments. There are 59 distinct time points in total. More details on the data can be found in Kaslow et al. (1987).

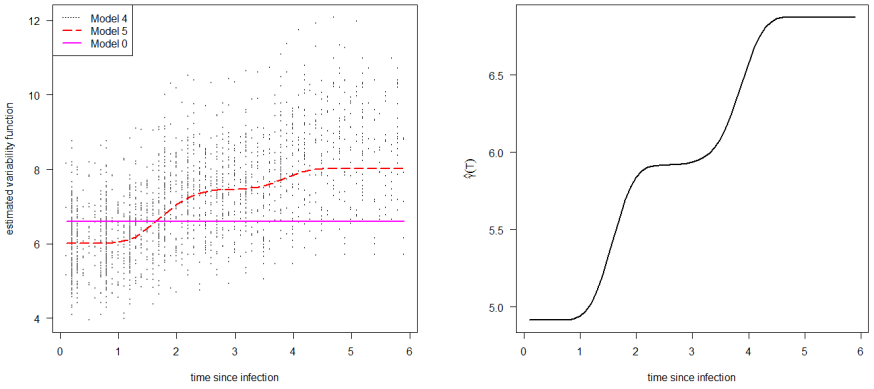
The covariates are $X_i^{(1)}$: the smoking status (1 or 0 according to whether the individual ever or never smoked cigarettes), $X_i^{(2)}$: age at HIV infection, and $X_i^{(3)}$: the pre-infection CD4 percentage of the individual. The response $Y(t_{ij})$ is the CD4 percentage at time t_{ij}, where t_{ij} denotes the time in years of the jth measurement on the ith individual after HIV infection. HIV attacks the CD4 cells. The CD4 percentage reflects an individual's immune system status, the lower the CD4 percentage the weaker the immune system and the more vulnerable an individual is to infections. Hence, it is of interest for investigating the evolvement of an HIV/AIDS infection over time.

The simple heteroscedastic model has been previously used (by Andriyana et al. (2016)) with these data and the validity of the two assumptions of the AHe approach checked (and fulfilled).

As discussed in Section 2.2, several comparisons are performed using the LRT test with B=200 bootstrap samples. Table 2.7 shows that V_0 and V_1 are significantly worse than V_4. But, V_5, V_2 and V_3 are not significantly worse than V_4. Hence, for this data set the simple heteroscedastic model (V_5) is recommended. The estimated variability functions for V_0, V_4 and V_5 are presented in Figure 2.7 (a). By using V_5, we see that the variability function varies with time. Furthermore, the estimated variability function of V_5 is located around the middle of the estimated variability function of V_4. The estimated function $\hat{\gamma}(T)$ in the variability function based on V_4 is given in Figure 2.7 (b). The estimators of the other coefficient functions in the variability function are approximately constant and equal to −0.2270, −0.0078 and 0.0147 for smoking status, age and pre-infection CD4, respectively (with only small deviations from a constant of magnitude 10^{-12}, 10^{-14} and 10^{-14}, respectively).

Table 2.7. CD4 and PM10 data. P-values.

Comparison	CD4	PM10
V_0 vs V_4	0.0450	0.7800
V_1 vs V_4	0.0400	0.7150
V_5 vs V_4	0.1100	0.9150
V_2 vs V_4	0.0950	0.9100
V_3 vs V_4	0.1700	0.9150



(a) (b)

Figure 2.7. CD4 data. (a) Estimated variability functions; (b) $\hat{\gamma}(T)$ based on V_4.

2.4.2 PM10 data

This data set is collected by the Norwegian Public Road Administration. It contains repeated measurements of air pollution, traffic volume, and meteorological variables for 273 days, at Alnabru in Oslo, Norway, between October 2001 and August 2003. During each day, measurements are performed at different time points (hours), with a range of 1 to 6 measurements per day, and a median of 2 measurements per day. The response $Y(t_{ij})$ is the logarithm of the concentration of PM10 at hour t_{ij}. PM10 is a mixture of solid and liquid droplets with diameter less than 10 micrometers, known as "particulate matter", which is one of the main constituent air pollutants with negative effect on human health. The covariates are $X^{(1)}(t_{ij})$ the logarithm of number of cars

per hour at time t_{ij}, $X^{(2)}(t_{ij})$ the wind speed (in meters/second), and $X^{(3)}(t_{ij})$ the temperature (2 meters above the ground in degree Celsius), where t_{ij} denotes the time in hours of the jth measurement of the ith day. See Statlib (http://lib.Stat.cmu.edu) and truncSP R-package (Karlsson and Lindmark, 2014) for more information on the data. It is of interest to investigate the evolution of the concentration of PM10 during a day.

This data set is analyzed in Andriyana (2015) using the simple and a general heteroscedastic model. It is seen from his analysis that the quantile curves obtained from the simple and the general heteroscedastic models are comparable. Further, the coefficients of the variability function are plotted. It can be seen from the plots in Andriyana (2015) that the coefficients of the covariates for the variability function are close to zero. The homoscedastic model is not considered previously for these data.

Several comparisons of the models are performed (presented in Table 2.7) using the LRT test with B = 200 bootstrap samples. The table shows that all the reduced models are not significantly worse than V_4, indicating that θ is zero and the variability function depends neither on time T nor on the covariates. Hence, for this data set V_0 (the homoscedastic model) can be used, which is better than all the other models in terms of model parsimony.

2.5 Conclusion

Various structures of the variability function are investigated. An estimation procedure for these functions is introduced without any reliance on global distributional assumptions and we show its consistency for an increasing number of knots. The simulation results show that the nesting structure of the different variability models is respected by the proposed estimation method. For instance, as is observed from simulation Setting 2, V_2 outperforms V_0, V_1 and V_5 and has an equivalent performance when compared to the more complicated variability structures V_3 and V_4. The Likelihood-Ratio-Type test proposed in this chapter, to check the structure of the variability function, confirms this finding. We illustrate the methodology on two data examples.

Hypothesis testing has also been performed using B = 99 bootstrap samples for the simulation settings and the data sets (not presented). Similar results as B = 200 bootstrap samples are obtained for each setting. The variability structure V_4 is the

most flexible modeling structure depending on the covariates. Hence, based on this model, one can do variable selection or test the constancy of the varying coefficient functions using the Likelihood-Ratio-Type test (see Chapter 3).

Although the Likelihood-Ratio-Type test has a good performance in the simulation study, it is of interest to prove the consistency of the test. This consistency relies on the consistency of the estimators of the parameters as well as on the consistency of the bootstrap procedure. If no modeling bias is assumed (i.e. $\theta_k(T) = \sum_{l=1}^{m_k^{\mathrm{V}}} \alpha_{kl}^{\mathrm{V}} B_{kl}(T, \nu_k^{\mathrm{V}})$ in V_4), then the consistency of the bootstrap procedure can be shown similarly as Knight (1999) (see also Andriyana et al., 2014). However, as we do not assume that the $\theta_k(T)$ are splines (i.e. there is modeling bias), the extension to this setting is not straight forward.

Due to the non-differentiability of the quantile loss function, there is no explicit formulation for the coefficients. Hence, the objective function is translated to a primal-dual linear programming problem, such that it is easy to implement it by using all available algorithms in the literature for the linear programming problems. We use a sparse implementation of the Frisch-Newton interior point algorithm due to its ability to exploit the sparsity that we are facing by adding a block diagonal matrix to penalize the adjacent coefficients of the B-splines. The proposed methods can be easily implemented using an R package called "QRegVCM" (Andriyana et al., 2018).

2.6 Assumptions, auxiliary results and proofs

2.6.1 Definitions and assumptions

Definition 2.1 *For two sequences of nonnegative real numbers a_n and b_n, $a_n \sim b_n$ denotes that there are constants $0 < a < A < \infty$ such that $a \leq a_n/b_n \leq A$ for all n.*

Definition 2.2 *The collection of all functions on $\mathcal{T}$ for which the lth order derivative satisfies the Hölder condition of order $\mathfrak{H}$ with $r = l + \mathfrak{H}$ is denoted by $\mathcal{H}_r$. That is, for each $h \in \mathcal{H}_r$, $|h^{(l)}(s) - h^{(l)}(t)| \leq c|s-t|^{\mathfrak{H}}$ for any $s, t \in \mathcal{T}$, and a positive constant c.*

We introduce additional notation.

$$R_{ij}^{\mathrm{V}} = \sum_{k=0}^{p}(\tilde{\theta}_k(t_{ij}) - \theta_k(t_{ij}))|X_{ij}^{(k)}|; \qquad m_{\max}^{\mathrm{V}} = \max(m_0^{\mathrm{V}}, \ldots, m_p^{\mathrm{V}});$$

$$\mathbf{Z}^{\mathrm{Va}} = \left(\mathbf{U}_{11}^{\mathrm{Va}}, \ldots, \mathbf{U}_{1N_1}^{\mathrm{Va}}, \ldots, \mathbf{U}_{n1}^{\mathrm{Va}}, \ldots, \mathbf{U}_{nN_n}^{\mathrm{Va}}\right)_{N\times 1};$$

$$(\mathbf{H}^{\mathrm{Va}})^2 = (\mathbf{Z}^{\mathrm{Va}})^{\mathrm{T}}\mathbf{Z}^{\mathrm{Va}}; \qquad \mathbf{z}_{ij} = (\mathbf{H}^{\mathrm{Va}})^{+}\mathbf{U}^{\mathrm{Va}}(t_{ij});$$

$$\hat{\boldsymbol{\alpha}}^{\mathrm{V}} = ((\hat{\boldsymbol{\alpha}}_0^{\mathrm{V}})^{\mathrm{T}}, (\hat{\boldsymbol{\alpha}}_1^{\mathrm{V}})^{\mathrm{T}}, \ldots, (\hat{\boldsymbol{\alpha}}_p^{\mathrm{V}})^{\mathrm{T}})^{\mathrm{T}}_{m_{tot}^{\mathrm{V}}\times 1}; \quad \hat{\boldsymbol{\theta}}^{*} = \mathbf{H}^{\mathrm{Va}}\hat{\boldsymbol{\alpha}}^{\mathrm{V}};$$

$$\tilde{\boldsymbol{\alpha}}^{\mathrm{V}} = ((\tilde{\boldsymbol{\alpha}}_0^{\mathrm{V}})^{\mathrm{T}}, (\tilde{\boldsymbol{\alpha}}_1^{\mathrm{V}})^{\mathrm{T}}, \ldots, (\tilde{\boldsymbol{\alpha}}_p^{\mathrm{V}})^{\mathrm{T}})^{\mathrm{T}}_{m_{tot}^{\mathrm{V}}\times 1}; \quad \tilde{\boldsymbol{\theta}}^{*} = \mathbf{H}^{\mathrm{Va}}\tilde{\boldsymbol{\alpha}}^{\mathrm{V}};$$

$$\epsilon_{ij}^{\mathrm{V}} = \epsilon^{\mathrm{V}}(t_{ij});$$

where $(\mathbf{H}^{\mathrm{Va}})^{+}$ denotes the Moore inverse of $\mathbf{H}^{\mathrm{Va}}$, and for a vector $\mathbf{u}$, $\|\mathbf{u}\|_2$ denotes its Euclidean norm.

Assumptions:

2.1. For $k = 0, 1, \ldots, p$, $\beta_k \in \mathcal{H}_{r_k}$, for some $r_k > 1/2$. Let c_k denote the Hölder constant for the kth varying coefficient function β_k.

2.2. $\limsup_{n\to\infty}\left(\frac{u_{\max}}{u_{\min}}\right) < \infty$, with $u_{\min} = \min(u_0, \ldots, u_p)$.

2.3. The error terms ϵ_{ij} and ϵ_{lz} are independent for all $i \neq l$ and for all j and z, with $i, l \in \{1, \ldots, n\}$, $j \in \{1, \ldots, N_i\}$ and $z \in \{1, \ldots, N_l\}$. The distributions of $(\epsilon_{i1}, \ldots, \epsilon_{iN_i})$ are the same for all i. The density function of ϵ_{ij} given $(\mathbf{X}_{ij}, t_{ij})$, f_ϵ, is continuous at zero and bounded away from zero and infinity.

2.4. $\lim_{n\to\infty} u_{\max} n^{\delta^M - 1} = 0$ for some $\delta^M \in (0, 1)$.

2.5. $u_{\max} \sim n^{1/(2r_{\min}+1)}$, where $r_{\min} = \min(r_0, \ldots, r_p)$.

2.6. The conditional distribution of T, given $\mathbf{X}$=$\mathbf{x}$ has a bounded density $f_{T|\mathbf{X}} : 0 < b \leq f_{T|\mathbf{X}}(t|\mathbf{x}) \leq B < \infty$, uniformly in $\mathbf{x}$ and t for some positive constants b and B.

2.7. $E(X^{(k)}(T)|T) = 0$ and $P(|X^{(k)}(T)| < M) = 1$ for some $M < \infty$, $k = 1, \ldots, p$, and there exist two positive definite matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ such that $\boldsymbol{\Sigma}_1 \leq Var(\mathbf{X}(T)|T) \leq \boldsymbol{\Sigma}_2$ uniformly in T, where $Var(\mathbf{X}(T)|T)$ denotes the conditional covariance matrix of $\mathbf{X}(T)$ given T.

2.8. For $k = 0, 1, \ldots, p$, $\theta_k \in \mathcal{H}_{r_k^{\mathrm{V}}}$, for some $r_k^{\mathrm{V}} > 1/2$. Let c_k^{V} denote the Hölder constant for the kth varying coefficient function θ_k.

2.9. Let $\epsilon^{\mathrm{V}}(T) = \ln(|\epsilon(T)|)$. The error terms $\epsilon_{ij}^{\mathrm{V}}$ and $\epsilon_{lz}^{\mathrm{V}}$ are independent for all $i \neq l$ and for all j and z, with $i, l \in \{1, \ldots, n\}$, $j \in \{1, \ldots, N_i\}$ and $z \in \{1, \ldots, N_l\}$. The distribution of $\epsilon_{ij}^{\mathrm{V}}$, given $(\mathbf{X}_{ij}, t_{ij})$, has a density function $f_{\epsilon^{\mathrm{V}}}$ which is continuous at zero with $f_{\epsilon^{\mathrm{V}}}(0) > 0$.

2.10. $\lim_{n\to\infty}(u_{\max}^{\mathrm{V}})^2 n^{\delta^{\mathrm{V}}-1} = 0$ for some $\delta^{\mathrm{V}} \in]0, 1[$, where $u_{\max}^{\mathrm{V}} = \max(u_0^{\mathrm{V}}, \ldots, u_p^{\mathrm{V}})$.

2.11. $u_{\max}^{\mathrm{V}} \sim n^{1/(2r_{\min}^{\mathrm{V}}+1)}$, with $r_{\min}^{\mathrm{V}} = \min(r_0^{\mathrm{V}}, \ldots, r_p^{\mathrm{V}})$.

2.12. $\limsup_{n\to\infty}\left(\frac{u_{\max}^{\mathrm{V}}}{u_{\min}^{\mathrm{V}}}\right) < \infty$, with $u_{\min}^{\mathrm{V}} = \min(u_0^{\mathrm{V}}, \ldots, u_p^{\mathrm{V}})$.

2.13. $\theta_k(t)$ is $\nu_k^{\mathrm{V}} - 1$ times continuously differentiable on $t \in \mathcal{T}$, with $\nu_k^{\mathrm{V}} \geq 1$.

2.6.2 Preliminary lemmas

The proof of Theorem 2.2 is based on the following lemmas.

Lemma 2.1 is a straightforward application of Corollary 6.21 in Schumaker (1981).

Lemma 2.1 *Suppose Assumption* **2.8** *holds. For some constant $W_{1,k}$ that depends only on ν_k^V and c_k^V, there exists a spline coefficient vector $\tilde{\boldsymbol{\alpha}}_k^V$ such that,*

$$sup_{t\in\mathcal{T}}|(\tilde{\boldsymbol{\alpha}}_k^V)^T\mathbf{B}_k^V(t, \nu_k^V) - \theta_k(t)| \leq W_{1,k}(u_k^V)^{-r_k^V}.$$

The following two lemmas are derived from He and Shi (1994) (Lemmas 3.4 and 3.5, respectively).

Lemma 2.2 *Under Assumption* **2.10**, *there exists a positive constant $W_{2,k}$ depending only on b, B, ν_k^V and c_k^V such that, except on an event whose probability tends to zero with n, for all $\hat{\theta}_k$ and $\theta_k \in \mathcal{H}_{r_k^V}$,*

$$\|\hat{\theta}_k - \theta_k\|_2^2 \leq W_{2,k}\left\{(u_{\max}^V)^{-2r_{\min}^V} + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}(\hat{\theta}_k(t_{ij}) - \theta_k(t_{ij}))^2\right\}.$$

Lemma 2.3 *There exists a positive constant $W_{3,k}$ such that, for all $\hat{\theta}_k$ and $\theta_k \in \mathcal{H}_{r_k^V}$,*

$$\|\hat{\theta}_k^{(j)} - \theta_k^{(j)}\|_2^2 \leq W_{3,k}\left\{(u_{\max}^V)^{2(j-r_{\min}^V)} + (u_{\max}^V)^{2j}\|\hat{\theta}_k - \theta_k\|_2^2\right\}, \qquad j = 0, 1, \ldots, \nu_k^V - 1.$$

Furthermore, we need the following lemma.

Lemma 2.4 *Suppose Assumptions* **2.6** – **2.9** *and* **2.11** *hold, and* $\boldsymbol{\xi} \in \mathbb{R}^{m^V_{tot}}$.

(a) *For any sequence* $\{L_n\}$ *satisfying* $1 \le L_n \le (u^V_{\max})^{\delta_0/10}$, *for some* $0 < \delta_0 < (r^V_{\min} - 1/2)/(2r^V_{\min} + 1)$,

$$\sup_{\|\boldsymbol{\xi}\|_2 \le 1} (u^V_{\max})^{-1} \Bigg| \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \Bigg[\rho_{0.5}\Big\{\epsilon^V_{ij} - L_n (u^V_{\max})^{1/2} \mathbf{z}^T_{ij}\boldsymbol{\xi} - R^V_{ij}\Big\} - \rho_{0.5}\Big\{\epsilon^V_{ij} - R^V_{ij}\Big\} + L_n (u^V_{\max})^{1/2} \mathbf{z}^T_{ij}\boldsymbol{\xi}(0.5 - I(\epsilon^V_{ij} < 0)) - E_{\epsilon^V}\Big[\rho_{0.5}\Big\{\epsilon^V_{ij} - L_n (u^V_{\max})^{1/2} \mathbf{z}^T_{ij}\boldsymbol{\xi} - R^V_{ij}\Big\} - \rho_{0.5}\Big\{\epsilon^V_{ij} - R^V_{ij}\Big\}\Big] \Bigg] \Bigg| = o_p(1),$$

where E_{ϵ^V} *is the conditional expectation given* $(\mathbf{X}_{ij}, t_{ij})$ *for* $i = 1, \ldots, n; j = 1, \ldots, N_i$.

(b) *For any* $\omega > 0$, *there exists* L *(sufficiently large) such that as* $n \to \infty$,

$$P\Bigg\{ (u^V_{\max})^{-1} \Bigg(\inf_{\|\boldsymbol{\xi}\|_2 = 1} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} E_{\epsilon^V}\Big[\rho_{0.5}\Big\{\epsilon^V_{ij} - L(u^V_{\max})^{1/2} \mathbf{z}^T_{ij}\boldsymbol{\xi} - R^V_{ij}\Big\} - \rho_{0.5}\Big\{\epsilon^V_{ij} - R^V_{ij}\Big\}\Big] - L(u^V_{\max})^{1/2} \Big\| \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{z}_{ij}(0.5 - I(\epsilon^V_{ij} < 0)) \Big\|_2 \Bigg) > 1 \Bigg\} > 1 - \omega.$$

The proof of Lemma 2.4 follows similar lines to those of Lemmas 3.2 and 3.3 in He and Shi (1994). The difference arises from the calculations of Equations (3.7) and (3.10) in He and Shi (1994); here the correlation of observations within subject i has to be taken into account. However, the goal is to find an upper bound for the aforementioned two expressions and since the correlation is at most one, we still attain the same upper bounds for the expressions. Therefore, we omit the proof.

Lemma 2.5 *Suppose Assumptions* **2.6** – **2.9** *and* **2.11** *hold. Furthermore, assume that* $(u^V_{\max})^{3/4} \lambda^V_{P\max} n^{-1/2} \to 0$ *as* n *tends to* ∞. *Let* $\tilde{\theta}_k(T) = (\tilde{\boldsymbol{\alpha}}^V_k)^T \mathbf{B}^V_k(T, \nu^V_k)$, *where the* $\tilde{\boldsymbol{\alpha}}^V_k$ *are the coefficients of the best possible spline approximation of* $\theta_k(T)$ *as defined in Lemma 2.1. Then,*

$$\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} (\hat{\theta}_k(t_{ij}) - \tilde{\theta}_k(t_{ij}))^2 = O_p(u^V_{\max}).$$

Proof 2.1 (Proof of Lemma 2.5) *The proof of Lemma 2.5 is along the same lines as the proof of Theorem 2.1 of He and Shi (1994). Using Lemma 2.4, for any* $\omega > 0$, *there exists* L *such that as* $n \to \infty$,

$$P\Bigg\{ (u^V_{\max})^{-1} \Bigg(\inf_{\|\boldsymbol{\xi}\|_2 = L(u^V_{\max})^{1/2}} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\Big\{\epsilon^V_{ij} - \mathbf{z}^T_{ij}\boldsymbol{\xi} - R^V_{ij}\Big\} - \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\Big\{\epsilon^V_{ij} - R^V_{ij}\Big\} \Bigg) > 1 \Bigg\} > 1 - \omega. \tag{2.15}$$

With probability one, we have

$$\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left\{\epsilon_{ij}^{V}-\mathbf{z}_{ij}^{T}(\hat{\boldsymbol{\theta}}^{*}-\tilde{\boldsymbol{\theta}}^{*})-R_{ij}^{V}\right\}=\inf_{\|\boldsymbol{\xi}\|_2\in\mathbb{R}^{m_{tot}^{V}}}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left\{\epsilon_{ij}^{V}-\mathbf{z}_{ij}^{T}\boldsymbol{\xi}-R_{ij}^{V}\right\}. \tag{2.16}$$

From the proof of Lemma 2.3 in Andriyana et al. (2014) and using $\|\hat{\boldsymbol{\theta}}^{*}-\tilde{\boldsymbol{\theta}}^{*}\|_2 = L(u_{\max}^{V})^{1/2}$*, we have*

$$\sum_{k=0}^{p}\lambda_{Pk}^{V}\|\mathbf{D}_{m_k^{V}}^{d_k^{V}}\hat{\boldsymbol{\alpha}}_k^{V}\|_1-\sum_{k=0}^{p}\lambda_{Pk}^{V}\|\mathbf{D}_{m_k^{V}}^{d_k^{V}}\tilde{\boldsymbol{\alpha}}_k^{V}\|_1=o_p(u_{\max}^{V}). \tag{2.17}$$

By Equations (2.16), (2.17) *and since*

$$\sum_{k=0}^{p}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left(\hat{\theta}_k(t_{ij})|X_{ij}^{(k)}|-\tilde{\theta}_k(t_{ij})|X_{ij}^{(k)}|\right)^2=\|\hat{\boldsymbol{\theta}}^{*}-\tilde{\boldsymbol{\theta}}^{*}\|_2^2,$$

Equation (2.15) *implies that*

$$\begin{aligned}P\left(\sum_{k=0}^{p}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left(\hat{\theta}_k(t_{ij})|X_{ij}^{(k)}|-\tilde{\theta}_k(t_{ij})|X_{ij}^{(k)}|\right)^2\le L^2u_{\max}^{V}\right)\\ =P(\|\hat{\boldsymbol{\theta}}^{*}-\tilde{\boldsymbol{\theta}}^{*}\|_2\le L(u_{\max}^{V})^{1/2})>1-\omega,\end{aligned}$$

which proves the lemma.

2.6.3 Proof of Theorem 2.2

Proof 2.2 (Proof of Theorem 2.2) *First, by the fact that* $2ab\le a^2+b^2$ *for any* $(a,b)\in\mathbb{R}^2$*, we have*

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{\theta}_k(t_{ij})-\theta_k(t_{ij})\right)^2\\ &\qquad\le\frac{2}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{\theta}_k(t_{ij})-\tilde{\theta}_k(t_{ij})\right)^2+2\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\tilde{\theta}_k(t_{ij})-\theta_k(t_{ij})\right)^2.\end{aligned}$$

By Lemma 2.5, we have that

$$\frac{2}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{\theta}_k(t_{ij})-\tilde{\theta}_k(t_{ij})\right)^2=\frac{2}{n}O_p(u_{\max}^V)=O_p\left(u_{\max}^V n^{-1}\right). \tag{2.18}$$

By Lemma 2.1 we have that

$$2\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\tilde{\theta}_k(t_{ij})-\theta_k(t_{ij})\right)^2\le 2\left(\max_{k=0,\ldots,p}(W_{1,k})\right)^2(u_{\max}^V)^{-2r_{\min}^V}=O_p\left((u_{\max}^V)^{-2r_{\min}^V}\right). \tag{2.19}$$

Combining Equations (2.18) *and* (2.19) *leads to*

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{\theta}_k(t_{ij})-\theta_k(t_{ij})\right)^2&=O_p\left(u_{\max}^V n^{-1}+(u_{\max}^V)^{-2r_{\min}^V}\right)\\&=O_p\left(n^{-2r_{\min}^V/(2r_{\min}^V+1)}\right).\end{aligned} \tag{2.20}$$

Using Equation (2.20)*, Lemmas 2.2 and 2.3, the proof of our main result is now complete.*

Chapter 3

Shape testing in quantile varying coefficient models with heteroscedastic error

Investigating several regression quantiles is important in order to get a nuanced picture for the relationship between a response and covariates. However, this is time consuming if the variance of the errors (the variability function) is of a homoscedastic structure (since it is sufficient, in this case, to consider one regression quantile). In the previous chapter, we study several structures of the variability function, like power and exponential functions; proposing also a Likelihood-Ratio-Type test to choose between two variability structures. In this chapter, we want to test the constancy of the coefficients in the variability function as well as in the signal part of the model (the median function). We investigate also other type of tests in addition to the LRT proposed in the previous chapter. For mean regression, Huang et al. (2002), Li et al. (2011) and Ahkim and Verhasselt (2018) develop testing procedure on the constancy of the varying coefficients. Using B-splines, Kim (2007), Wang et al. (2009), Tang et al. (2012) and Feng and Zhu (2016) propose the corresponding test for quantile regression.

Furthermore, we discuss shape testing for the coefficient of a specific covariate, such as a monotonicity test to check whether a covariate has a non-decreasing/non-increasing effect over the evolution of another covariate. An estimation of monotone B-spline smoothing in a simple non-parametric quantile regression was pioneered by He and Shi (1998). He and Ng (1999) extend these results to a general framework for constrained L_1-norm regression. The extended framework includes monotonicity, convexity/con-

cavity, periodicity and pointwise constraints. As a consequence, they develop an algorithm called COBS (constrained B-splines smoothing) to S-plus users. Since the above results were for the case of univariate smoothing, where the unknown function is a function of a single variable, Kim (2006) extends them to the varying coefficient model. Bollaerts et al. (2006) propose, for the univariate case, an estimation technique for non-parametric monotone quantile regression using P-splines.

In mean regression, estimation methods under monotonicity constraints and testing procedures for monotonicity have been widely discussed in univariate regression (Bowman et al., 1998; Ghosal et al., 2000; Wang and Meyer, 2011, and refereces therein). Zhang et al. (2013) and Ahkim et al. (2017) extend testing the features of the functional coefficients to varying coefficient models. Zhang et al. (2013) propose a robust scenario of the SiZer (significant zero crossing of derivative) inference approach (Zhang and Mei, 2012) based on the local least absolute deviation fitting procedure, using the local polynomial estimation technique. Ahkim et al. (2017) propose a generalization of the testing procedure developed by Wang and Meyer (2011) and another testing procedure called Extreme Value Test, discussed in Section 3.1.2, using the B-splines technique.

A monotonicity testing procedure in quantile regression is important in a real life application, e.g. to test whether the weight of a child is significantly decreasing at some points in time with the age of the child.

In this chapter we consider the varying coefficient model in Equation (1.7) with the variability function having the following form (the "full" model in the previous chapter):

$$V(\mathbf{X}(T), T) = \gamma(T) \exp\{\theta_1(T)X^{(1)}(T) + \ldots + \theta_p(T)X^{(p)}(T)\}, \tag{3.1}$$

where $\gamma(T) \geqslant 0$ and $\theta_k(T) \in \mathbb{R}$ for $k = 1, 2, \ldots, p$.

Remark 3.1 *A similar variability function to Equation* (3.1) *is considered by Van Keilegom and Wang (2010), with a partially linear structure. The structure in Equation* (3.1) *ensures that the variability function is positive and the functional coefficients are flexible enough.*

The goal of this study is to propose a testing procedure that (i) checks whether the coefficients $\beta_k(T)$ and $\theta_k(T)$ can be considered constant over T, for $k = 0, \ldots, p$; (ii) checks the shape (monotonicity or convexity/concavity) of $\beta_k(T)$ over T.

The rest of the chapter is organized as follows. Section 3.1 deals with testing procedures for constancy, monotonicity and convexity/concavity. Simulation studies are

carried out in Section 3.2. The testing procedures in Section 3.1 are applied on a data example in Section 3.3. Finally, Section 3.4 concludes the results in this chapter. The R-code to implement the methods of this chapter on the data set is deferred to Appendix A and is available in QRegVCM R-package (Andriyana et al., 2018).

3.1 Testing

For the signal, the hypotheses are defined as:

$$H_0 : q_{0.5}(Y(T)|\mathbf{X}(T), T) \in \mathcal{M}_0 \text{ Vs. } H_1 : q_{0.5}(Y(T)|\mathbf{X}(T), T) \in \mathcal{M}_1,$$

where $\mathcal{M}_0 \subset \mathcal{M}_1$.

This can be extended to other hypotheses, such as constancy, monotonicity or convexity of the coefficients in the signal, and also in the variability function or a combination of these.

3.1.1 Constancy tests

In this section, four different types of testing procedures for the signal as well as the variability function are discussed. They are the Likelihood-Ratio-Type (LRT) (considered in the previous chapter), the L_1, the L_2 and the $L_{\max}$ (discussed below) tests.

The LRT test

The test is based on the objective function instead of the likelihood function. The test statistic, for the signal, is defined as:

$$G_m = 2\left\{\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\{Y(t_{ij}) - \hat{m}_0(t_{ij})\} - \sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\{Y(t_{ij}) - \hat{m}_1(t_{ij})\}\right\},$$

where $\hat{m}_0(\cdot)$ and $\hat{m}_1(\cdot)$ are the estimated signal under H_0 and under H_1, respectively.

The null hypothesis is rejected when G_m is too large, as in the previous chapter. The p-value is obtained by a re-sampling subject bootstrap:

1. re-sample n subjects with replacement from $i = 1, \ldots, n$, to obtain the bth bootstrap sample

$$\{(Y^b(t_{ij}^b), \mathbf{X}^b(t_{ij}^b), t_{ij}^b) : i = 1, \ldots, n, j = 1, \ldots, N_i^b\}$$

from $\{(Y^P(t_{ij}), \mathbf{X}(t_{ij}), t_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N_i\}$, with

$$Y^P(t_{ij}) = \hat{m}_0(t_{ij}) + \epsilon^P(t_{ij}), \text{ and } \epsilon^P(t_{ij}) = Y(t_{ij}) - \hat{m}_2(t_{ij}),$$

where $\hat{m}_0(\cdot)$ and $\hat{m}_2(\cdot)$ are the estimated signal under H_0 and under the most complex signal (where all coefficients are varying with T), respectively.

2. repeat the above sampling procedure B times (hence $b = 1, \ldots, B$).
3. calculate the test statistic G_m^b for each bth bootstrap sample to obtain its empirical distribution.
4. get the p-value using the empirical probability of $G_m^b \geq G_m$.

Similarly, we can conduct the same procedure for the variability function (see Chapter 2). The test statistic is also defined similarly but changing $(Y(t_{ij}),\ \hat{m}_0(t_{ij})$ and $\hat{m}_1(t_{ij}))$ to $(\ln|Y(t_{ij}) - \mathbf{X}^{\mathrm{T}}(t_{ij})\hat{\boldsymbol{\beta}}(t_{ij})|,\ \hat{v}_0(t_{ij})$ and $\hat{v}_1(t_{ij}))$, where $\hat{v}_0(\cdot)$ and $\hat{v}_1(\cdot)$ are the estimated variability functions under H_0 and under H_1, respectively. The p-value is obtained by a re-sampling subject bootstrap, similar to the test for the signal, but now we obtain the bootstrap sample

$$\{(Y^b(t_{ij}^b), \mathbf{X}^b(t_{ij}^b), t_{ij}^b) : i = 1, \ldots, n, j = 1, \ldots, N_i^b\}$$

from $\{(Y^P(t_{ij}), \mathbf{X}(t_{ij}), t_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N_i\}$, with

$$Y^P(t_{ij}) = \sum_{k=0}^{p} \hat{\beta}_k(t_{ij}) X^{(k)}(t_{ij}) + \hat{v}_0(t_{ij})\epsilon^P(t_{ij}),$$
$$\epsilon^P(t_{ij}) = \frac{Y(t_{ij}) - \sum_{k=0}^{p} \hat{\beta}_k(t_{ij}) X^{(k)}(t_{ij})}{\hat{v}_2(t_{ij})},$$

where $\hat{v}_2(\cdot)$ is an estimate for the most complex variability function.

The L_1, the L_2 and the $L_{\max}$ tests

We introduce in this paragraph three types of test statistics for testing that a coefficient function is constant or varying. When the coefficients $\boldsymbol{\alpha}_k$ are a constant

vector, the corresponding spline $\beta_k(\cdot)$ is constant. From de Boor (2001), we have that the first order derivative of $\beta_k(T)$ in Equation (2.4), with distance $1/u_k$ between the equidistant knots (for $T \in [0,1]$), is given by

$$\beta_k'(T) = u_k \sum_{j=2}^{m_k} \Delta^1 \alpha_{kj} \mathbf{B}_k(T; \nu_k - 1) = u_k \mathbf{B}_k(T; \nu_k - 1)^{\mathrm{T}} \mathbf{D}^1_{m_k} \boldsymbol{\alpha}_k. \tag{3.2}$$

Degree one B-splines have support at three knots, and equal zero at the two end knot points and one at the middle knot point (see Figure 1.2). Therefore, for quadratic B-splines and by Equation (3.2), $\beta_k'(t) \neq 0$ for all $t \in \mathcal{T}$ if and only if $\mathbf{D}^1_{m_k} \boldsymbol{\alpha}_k \neq \mathbf{0}$.

The test statistics are based on a vector norm of the differences of the B-splines coefficients for the $\beta_k(\cdot)$. They are defined as follows, for testing constancy of the $\beta_k(\cdot)$:

$$L_{1m} = \sum_{k=0}^{p} \|\mathbf{D}^1_{m_k} \hat{\boldsymbol{\alpha}}_k\|_1, \qquad L_{2m} = \sum_{k=0}^{p} \|\mathbf{D}^1_{m_k} \hat{\boldsymbol{\alpha}}_k\|_2, \text{ and} \qquad L_{\max m} = \sum_{k=0}^{p} \|\mathbf{D}^1_{m_k} \hat{\boldsymbol{\alpha}}_k\|_\infty,$$

where $\|\boldsymbol{\alpha}_k\|_1 = \sum_{j=1}^{m_k} |\alpha_{kj}|$, $\|\boldsymbol{\alpha}_k\|_2 = \sqrt{\sum_{j=1}^{m_k} |\alpha_{kj}|^2}$, and $\|\boldsymbol{\alpha}_k\|_\infty = \max_{j=1,\ldots,m_k} |\alpha_{kj}|$, are respectively the L_1-norm, the L_2-norm and the L_∞-norm of the vector $\boldsymbol{\alpha}_k$.

Each of the above test statistics looks at the consecutive differences of the B-splines coefficients to check whether the coefficient for the corresponding covariate varies over time or not. The testing procedure for each of the above test statistics is based on bootstrap re-sampling, to calculate the p-values.

Similar to that of the signal, the three test statistics are also considered for testing the constancy of the coefficients $\theta_k(\cdot)$ in the variability function, defined as follows:

$$L_{1v} = \sum_{k=0}^{p} \|\mathbf{D}^1_{m_k^{\mathrm{V}}} \hat{\boldsymbol{\alpha}}_k^{\mathrm{V}}\|_1, \qquad L_{2v} = \sum_{k=0}^{p} \|\mathbf{D}^1_{m_k^{\mathrm{V}}} \hat{\boldsymbol{\alpha}}_k^{\mathrm{V}}\|_2, \text{ and} \qquad L_{\max v} = \sum_{k=0}^{p} \|\mathbf{D}^1_{m_k^{\mathrm{V}}} \hat{\boldsymbol{\alpha}}_k^{\mathrm{V}}\|_\infty.$$

When we use B-splines of degree three or more, we use the $L_{\max}$ test given by (for the signal, similar for the variability function):

$$L_{\max m} = \sum_{k=0}^{p} \max_{t \in \mathcal{T}} |\mathbf{B}_k(t; \nu_k - 1)^{\mathrm{T}} \mathbf{D}^1_{m_k} \boldsymbol{\alpha}_k|.$$

Remark 3.2 *We can similarly test for linearity of the functional coefficients (H_0 : $\beta_k''(t) = 0$ for all $t \in \mathcal{T}$). For cubic B-splines and by Equation (3.7), $\beta_k''(t) \neq 0$ for all $t \in \mathcal{T}$ if and only if $\mathbf{D}^2_{m_k} \boldsymbol{\alpha}_k \neq \mathbf{0}$. Hence, the test statistics are constructed based on*

the second order differences of the B-splines coefficients.

3.1.2 Monotonicity tests

By Equation (3.2), as u_k and $\mathbf{B}_k(T; \nu_k - 1)$ are all positive by definition, restricting $\Delta^1 \alpha_{kj}$ to be positive (resp. negative) is a sufficient condition for $\beta'_k(T)$ to be positive (resp. negative). Hence, to restrict $\beta_k(T)$ to be monotone increasing or decreasing, we use the objective function considered by Bollaerts et al. (2006), which adds asymmetric weights ϖ_{kj} to the objective function in Equation (2.6)

$$S_m(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5} \left(Y(t_{ij}) - \sum_{k=0}^{p} X^{(k)}(t_{ij}) \boldsymbol{\alpha}_k^{\mathrm{T}} \mathbf{B}_k(t_{ij}; \nu_k) \right) + \lambda_P \sum_{k=0}^{p} \omega_{1k} \| \mathbf{D}_{m_k}^{d_k} \boldsymbol{\alpha}_k \|_1 + \sum_{k=0}^{p} \kappa_k \sum_{j=2}^{m_k} \varpi_{kj} |\Delta^1 \alpha_{kj}| \tag{3.3}$$

with κ_k a user-defined constraint parameter chosen as large as possible to avoid the violation of the constraints and

$$\varpi_{kj} = \begin{cases} 0 & \text{if } \Delta^1 \alpha_{kj} \geq 0 \text{ for increasing (resp. } \Delta^1 \alpha_{kj} \leq 0 \text{ for decreasing),} \\ 1 & \text{otherwise.} \end{cases} \tag{3.4}$$

The implementation of the estimation method for the objective function in Equation (3.3), for a model with only one covariate, is discussed in Section 3 of Bollaerts et al. (2006). It can be extended, for a model with several covariates, by adding the additional constraints in Equation (3.4) for each coefficient of the covariates.

To test the monotonicity of β_k, we construct two types of tests. The first one is the LRT used in Section 3.1.1. The second test is the algorithm used by Ahkim et al. (2017) for mean regression, which we call Extreme Value Test (EVT).

The hypothesis for testing β_k is a non-decreasing function is:

$$H_0 : \beta'_k(t) \geq 0 \text{ for all } t \in \mathcal{T} \qquad \text{Versus} \qquad H_1 : \neg H_0. \tag{3.5}$$

To test β_k is a non-increasing function, the hypothesis is

$$H_0 : \beta'_k(t) \leq 0 \text{ for all } t \in \mathcal{T} \qquad \text{Versus} \qquad H_1 : \neg H_0. \tag{3.6}$$

Degree one B-splines have support at three knots, and equal zero at the two end knot points and one at the middle knot point (Figure 1.2). Therefore, for quadratic B-splines and by Equation (3.2), β_k is non-decreasing if and only if $\mathbf{D}^1_{m_k}\boldsymbol{\alpha}_k \geq \mathbf{0}$. Then, the algorithm for the testing procedure reads as follows:

1. $d_{\min} = \min_{2\leq j\leq m_k}(\Delta^1\hat{\alpha}_{kj})$

2. if $d_{\min} \geq 0$, the p-value equals one

3. if $d_{\min} < 0$, determine the distribution of $d_{\min}$ under H_0

4. get the p-value using the empirical probability of $d^b_{\min} \leq d_{\min}$, where $d^b_{\min}$ is the test statistic based on the bth bootstrap sample obtained under the null hypothesis.

The bootstrap samples ($b = 1, \ldots, B$)

$$\{(Y^b(t^b_{ij}), \mathbf{X}^b(t^b_{ij}), t^b_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N^b_i\}$$

are obtained from $\{(Y^P(t_{ij}), \mathbf{X}(t_{ij}), t_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N_i\}$, with

$$Y^P(t_{ij}) = \sum_{k=0}^{p} \hat{\beta}^m_k(t_{ij})X^{(k)}(t_{ij}) + \epsilon^P(t_{ij}),$$
$$\epsilon^P(t_{ij}) = Y(t_{ij}) - \sum_{k=0}^{p} \hat{\beta}_k(t_{ij})X^{(k)}(t_{ij}),$$

where $\hat{\beta}^m_k(t_{ij})$ and $\hat{\beta}_k(t_{ij})$ are the monotone constrained (solution of Equation (3.3)) and un-constrained estimated coefficients, respectively.

When we use B-splines of degree three or more, our test is based on

$$d_{\min} = \min_{t\epsilon\mathcal{T}}(\mathbf{B}_k(t; \nu_k - 1)^{\mathrm{T}}\mathbf{D}^1_{m_k}\boldsymbol{\alpha}_k).$$

The test for hypothesis in Equation (3.6) can be carried out similarly with the obvious adjustment of the inequality and now using $d_{\max} = \max_{2\leq j\leq m_k}(\Delta^1\hat{\alpha}_{kj})$.

3.1.3 Convexity/Concavity tests

The second derivative of $\beta_k(T)$ in Equation (2.4), with distance $1/u_k$ between the equidistant knots (for $T \in [0,1]$) is given by

$$\beta_k''(T) = u_k^2 \sum_{j=3}^{m_k} \Delta^2 \alpha_{kj} \mathbf{B}_k(T; \nu_k - 2). \tag{3.7}$$

Hence, to restrict $\beta_k(T)$ to be convex or concave, we use the following objective function, which adds asymmetric weights ϑ_{kj} to the objective function in (2.6)

$$S_c(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left(Y(t_{ij}) - \sum_{k=0}^{p} X^{(k)}(t_{ij}) \boldsymbol{\alpha}_k^{\mathrm{T}} \mathbf{B}_k(t_{ij}; \nu_k)\right) + \lambda_P \sum_{k=0}^{p} \omega_{1k} \|\mathbf{D}_{m_k}^{d_k} \boldsymbol{\alpha}_k\|_1 + \sum_{k=0}^{p} \kappa_k^c \sum_{j=3}^{m_k} \vartheta_{kj} |\Delta^2 \alpha_{kj}| \tag{3.8}$$

with κ_k^c a user-defined constraint parameter and

$$\vartheta_{kj} = \begin{cases} 0 & \text{if } \Delta^2 \alpha_{kj} \geq 0 \text{ for convex (resp. } \Delta^2 \alpha_{kj} \leq 0 \text{ for concave),} \\ 1 & \text{otherwise.} \end{cases}$$

The implementation of the estimation method for the objective function in Equation (3.8) can also be written similarly as that of monotone quantile regression but now using the second order differences.

Here also we can use the two tests discussed in Section 3.1.2, but now we work with the second order derivative (for EVT test). The hypothesis for testing β_k is a convex function reads as:

$$H_0 : \beta_k''(t) \geq 0 \text{ for all } t \in \mathcal{T} \qquad \text{Versus} \qquad H_1 : \neg H_0.$$

The EVT test, for cubic splines, is given as $d_{\min} = \min_{3 \leq j \leq m_k} (\Delta^2 \hat{\alpha}_{kj})$. When splines of degree four or more are used the test is based on

$$d_{\min} = \min_{t \in \mathcal{T}} (\mathbf{B}_k(t; \nu_k - 2)^{\mathrm{T}} \mathbf{D}_{m_k}^2 \boldsymbol{\alpha}_k).$$

To test whether $\beta_k(\cdot)$ is a concave function, the hypothesis is

$$H_0 : \beta_k''(t) \leq 0 \text{ for all } t \in \mathcal{T} \qquad \text{Versus} \qquad H_1 : \neg H_0.$$

3.1.4 Shape testing in both signal and variability function

In this section, we introduce the LRT to test, simultaneously, convexity of a coefficient in the signal and constancy of a coefficient in the variability function (can be generalized for other shapes). The hypothesis reads as follows:

$$H_0 : \beta_k''(t) \geq 0 \ \& \ \theta_k'(t) = 0 \text{ for all } t \in \mathcal{T} \qquad \text{Versus} \qquad H_1 : \neg H_0.$$

The test statistic is defined as:

$$\begin{aligned} G = 2\Bigg\{ \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \Big[& \rho_{0.5}\{Y(t_{ij}) - \hat{m}_0(t_{ij})\} - \rho_{0.5}\{Y(t_{ij}) - \hat{m}_1(t_{ij})\} \\ & + \rho_{0.5}\{\ln|Y(t_{ij}) - \mathbf{X}^{\mathrm{T}}(t_{ij})\hat{\boldsymbol{\beta}}(t_{ij})| - \hat{v}_0(t_{ij})\} \\ & - \rho_{0.5}\{\ln|Y(t_{ij}) - \mathbf{X}^{\mathrm{T}}(t_{ij})\hat{\boldsymbol{\beta}}(t_{ij})| - \hat{v}_1(t_{ij})\} \Big] \Bigg\}, \end{aligned}$$

where $\hat{m}_0(\cdot)$ and $\hat{m}_1(\cdot)$ are, respectively, the estimated signal under H_0 and under H_1, and $\hat{v}_0(\cdot)$ and $\hat{v}_1(\cdot)$ are, respectively, the estimated variability function under H_0 and under H_1.

The p-value is obtained by the re-sampling subject bootstrap, similar procedure as discussed in Section 3.1.1. The bootstrap samples ($b = 1, \ldots, B$)

$$\{(Y^b(t_{ij}^b), \mathbf{X}^b(t_{ij}^b), t_{ij}^b) : i = 1, \ldots, n, j = 1, \ldots, N_i^b\}$$

are obtained from $\{(Y^P(t_{ij}), \mathbf{X}(t_{ij}), t_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N_i\}$, with

$$\begin{aligned} Y^P(t_{ij}) &= \sum_{k=0}^{p} \hat{\beta}_k^c(t_{ij}) X^{(k)}(t_{ij}) + \hat{v}_0(t_{ij}) \epsilon^P(t_{ij}), \\ \epsilon^P(t_{ij}) &= \frac{Y(t_{ij}) - \sum_{k=0}^{p} \hat{\beta}_k(t_{ij}) X^{(k)}(t_{ij})}{\hat{v}_2(t_{ij})}, \end{aligned}$$

where $\hat{\beta}_k^c(\cdot)$ and $\hat{\beta}_k(\cdot)$ are, respectively, the convex constrained (solution of Equation (3.8)) and un-constrained estimated coefficients for the signal, and $\hat{v}_2(\cdot)$ is an estimate

for the most complex variability function.

3.2 Simulation study

In this section, we conduct simulation studies to investigate the performances of the testing procedures discussed in Section 3.1. Three simulation settings are considered (i) testing constancy of the coefficients in the signal as well as in the variability function; (ii) monotonicity test for the coefficients in the signal; (iii) testing, simultaneously, for convexity of the functional coefficient of a covariate in the signal and constancy of the functional coefficient of a covariate in the variability function. To analyze the data sets, B-splines of degree two with six equidistant knots on the time interval (similar as in Ahkim et al., 2017) and differencing order 1 are used.

3.2.1 Simulation study: constancy

We investigate the performances of the testing procedures in Section 3.1.1. We further compare the testing procedures with the Rank Score (RS) test of Wang et al. (2009) (using the compound symmetry error structure, and cubic splines for B-spline estimators). A simulation setting (defined in Table 3.1) of the following type is used:

$$\begin{aligned} Y(T) =& \beta_0(T) + \beta_1(T)X^{(1)}(T) + \beta_2(T)X^{(2)}(T) + \\ & \gamma(T)\exp\{\theta_1(T)X^{(1)}(T) + \theta_2(T)X^{(2)}(T)\}\epsilon(T). \end{aligned}$$

Table 3.1. Description of the coefficients.

$\beta_0(T)$	$\beta_1(T)$	$\beta_2(T)$	$\gamma(T)$	$\theta_1(T)$	$\theta_2(T)$
$2\sqrt{T}$	$\frac{\pi T+10}{15}$	$9\sin(\pi T/25)+3$	$\frac{T+1}{25}$	$\frac{-T}{5}+5$	$\frac{(20-T)^2}{100}-4$

To have an equivalent complexity in the data for the simulated samples, the coefficients are formulated such that the SNR is approximately seven.

As in Chapter 2, the error term is generated from a transformed multivariate normal distribution given in Equation (2.14) (such that we attain the two assumptions **H1** and **H2**, defined in Chapter 2) with ζ_{ij} generated from a Gaussian process with zero

mean and a covariance

$$cov(\zeta_{ij}, \zeta_{lz}) = \begin{cases} 0.6^2/5 & \text{if } i = l, j \neq z \\ 0.6^2 & \text{if } i = l, j = z \\ 0 & \text{if } i \neq l. \end{cases}$$

The following time dependent covariates are considered

$$\begin{pmatrix} X^{(1)}(T) \\ X^{(2)}(T) \end{pmatrix} \sim N(\mathbf{0}, \mathbf{\Sigma}(T)), \qquad \mathbf{\Sigma}(T) = \begin{pmatrix} 1 & 1/(4+T) \\ 1/(4+T) & 1 \end{pmatrix}.$$

Then, the covariates $X^{(k)}(t_{ij})$ are standardized as follows,

$$\frac{X^{(k)}(t_{ij}) - \min_{1\leq i\leq n; 1\leq j\leq N_i}\left(X^{(k)}(t_{ij})\right)}{\max_{1\leq i\leq n; 1\leq j\leq N_i}\left(X^{(k)}(t_{ij})\right) - \min_{1\leq i\leq n; 1\leq j\leq N_i}\left(X^{(k)}(t_{ij})\right)},$$

to have them on a comparable scale, which is important to determine the individual smoothing parameters λ_{Pk}'s with $\lambda_{Pk} = \lambda_P \omega_{1k}$, chosen via the SIC in Chapter 2.

We simulated 100 data sets of size $n = 100$ from the above model. The time variable ranges from 0 to 49. For each case i, the probability of having a measurement in each time point is 0.6, creating an unbalanced number of measurements. Then, the actual time points are calculated by adding a generated value from a $U[0, 0.5]$ to the non-skipped time points. For the test, we use $B = 200$ bootstrap samples and proportion of significant tests are presented with 5% significance level.

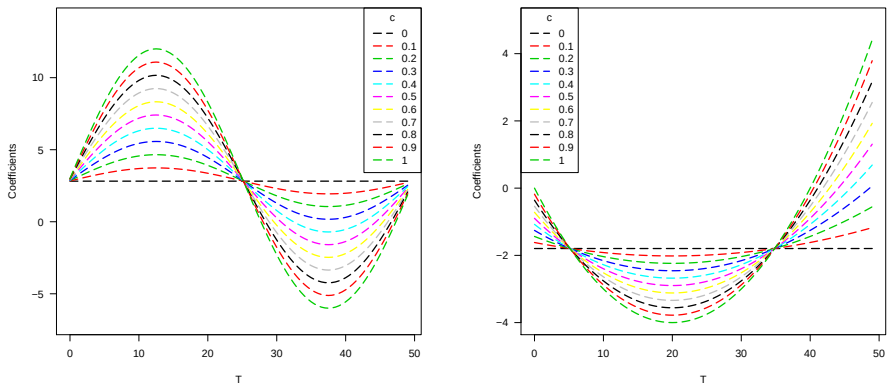
We test whether the coefficients for $X^{(2)}(T)$ are varying over time or not (it can similarly be done for the other covariates) for the signal ($\beta_2(T)$) as well as for the variability function ($\theta_2(T)$). To study the power of the tests in more detail we use a sequence of alternative models indexed by c for the aforementioned coefficients:

$$\beta_2(T, c) = c_1 + c\{\beta_2(T) - c_1\}, \qquad c_1 = \int_0^{49} \beta_2(t)dt/49$$

$$\theta_2(T, c) = c_2 + c\{\theta_2(T) - c_2\}, \qquad c_2 = \int_0^{49} \theta_2(t)dt/49$$

for $c \in \{0, 0.1, \ldots, 1\}$, presented in Figure 3.1.

For $\beta_2(T)$, we see in Figure 3.2 (a) that all the tests preserve the nominal significance



(a) $\beta_2(T, c)$ (b) $\theta_2(T, c)$

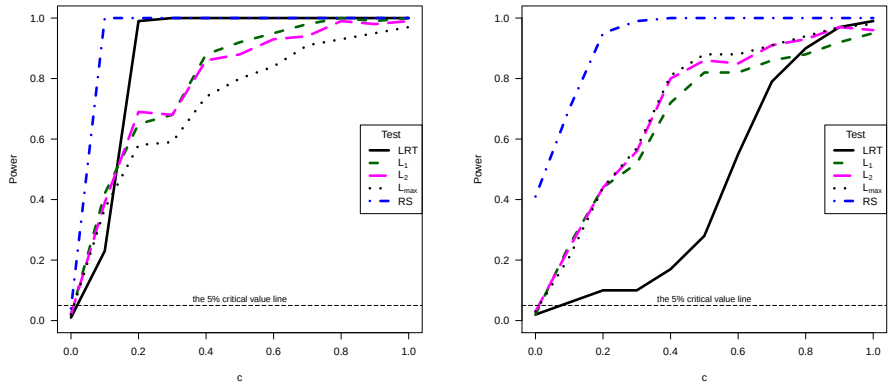
Figure 3.1. The true coefficients of $X^{(2)}(T)$ for the signal (left panel) and for the variability function (right panel) for various values of c.

level for $c = 0$ and the power increases to 1 for large values of c. Note that the power curve of the RS test outperforms these of the other tests. Among these, the L_1, L_2 and $L_{\max}$ tests perform better than the LRT test for small values of c, whereas the latter test performs best for moderate to large values of c. Since the power curves of the L_1, L_2, $L_{\max}$ and LRT tests are crossing, there is no uniformly best test among them. For the variability function (see Figure 3.2 (b)) we see that all the tests other than the RS test preserve the nominal level. Among the tests that best keep the nominal level, we observe that the power properties of the L_1, L_2 and $L_{\max}$ tests are substantially better than these of the LRT test, for a large range of c-values. The power further increases to 1 for c large for all tests.

3.2.2 Simulation study: monotonicity

In this section we conduct a simulation study to investigate the performances of the testing procedures discussed in Section 3.1.2, for the signal. We consider a homoscedastic varying coefficient model with $V(\cdot, \cdot) = 0.4$. We use the coefficient functions $\beta_0(T) = 0.25 + 2T$ and $\beta_2(T) = -0.5 + 10(T - 0.5)^2$. We consider four simulation settings to test the hypothesis in Equation (3.5) for $\beta_1(T)$, which are based on the following two functions,

- $f_{1,a}(T) = -2 + 2(1 + T - a\exp(-50(T - 0.5)^2))$ and



(a) $\beta_2(T)$ (b) $\theta_2(T)$

Figure 3.2. The power curves for $\beta_2(T)$ and $\theta_2(T)$.

- $f_2(T) = 1.1$.

The first three simulation settings are based on $f_{1,a}(T)$ (presented in Figure 3.3), which is taken from Bowman et al. (1998). This function is strictly increasing for $a = 0.15$, a dip appears when $a = 0.3$ and the dip appears more profoundly with $a = 0.45$. The fourth simulation setting is based on $f_2(T)$. The coefficients are formulated such that the SNR is approximately seven.

The error term, the covariates and the time variable are generated as in the previous section. Then the time variable t_{ij} is standardized via

$$\frac{t_{ij} - \min_{1\leq i\leq n;1\leq j\leq N_i}(t_{ij})}{\max_{1\leq i\leq n;1\leq j\leq N_i}(t_{ij}) - \min_{1\leq i\leq n;1\leq j\leq N_i}(t_{ij})}, \tag{3.9}$$

to ensure that the time variable is between zero and one (and hence T takes values in $[0, 1]$).

To avoid the violation of the monotonicity constraint we use $\kappa_1 = 1000$.

We present the proportion of significant tests for each testing procedure under the four simulation settings in Table 3.2. Under the monotone functions $f_2(T)$ and $f_{1,0.15}(T)$, we see that the P(rejecting $H_0|H_0$) is around the nominal significance level, for both tests. Whereas, under the non-monotone functions, $f_{1,0.3}(T)$ and $f_{1,0.45}(T)$, the power of the test is around one, except for LRT having less power under $f_{1,0.3}(T)$. Here we see that $d_{\min}$ test perform better than the LRT test.

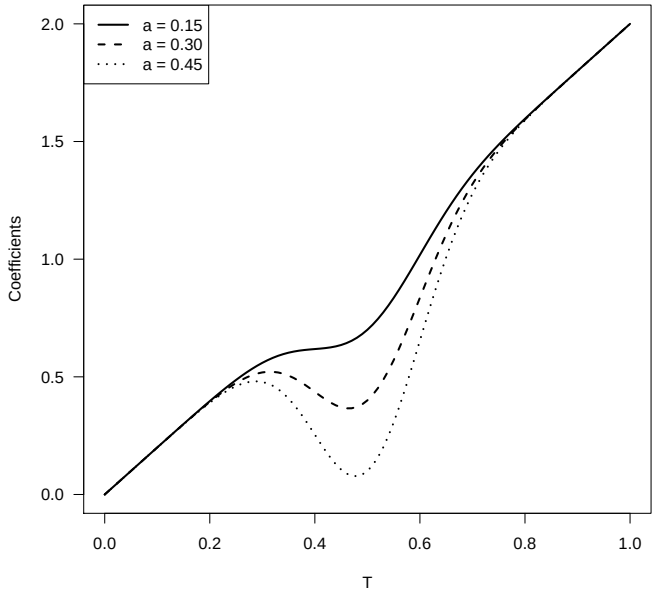


Figure 3.3. $f_{1,a}(T)$.

Table 3.2. Proportion of significant tests.

Setting	LRT	$d_{\min}$	Ideal
f_2	0.06	0.04	0.05
$f_{1,0.15}$	0.07	0.06	0.05
$f_{1,0.3}$	0.51	0.87	1.00
$f_{1,0.45}$	0.95	0.99	1.00

3.2.3 Simulation study: convex in signal and constant in variability

Using the LRT testing procedure, as discussed in Section 3.1.4, we test whether a covariate has a coefficient function that is convex for the signal and is constant for the variability function. We consider the simulation setting of Section 3.2.2 and take $\beta_1(T)$ equal to $f_{1,0.45}(T)$. The proportion of significant tests, respectively, are 0.01 and 0.97 for $X^{(0)}(T)$ ($H_0 : \beta_0''(t) \geq 0$ & $\theta_0'(t) = 0$ for all $t \in \mathcal{T}$) and $X^{(1)}(T)$ ($H_0 : \beta_1''(t) \geq 0$ & $\theta_1'(t) = 0$ for all $t \in \mathcal{T}$). These results show that the test has good

performance, since the nominal significance level is respected and the power is close to one.

3.3 Data example: Wages data

In this section, we illustrate the use of the testing procedures in Section 3.1 on the 'Wages' data. To analyze the data set, similarly as in Section 3.2, B-splines of degree two with six equidistant knots on the time interval and differencing order one are used.

The data example is taken from the National Longitudinal Survey of Youth, collected in USA. It was originally reported in Murnane et al. (1999). It consists of 888 individuals of age 14 to 17 years old. Each individual has a different number of observations (N_i) ranging between 1 and 13. The response variable $Y(t_{ij})$ is the hourly wage, with the time variable t_{ij} denoting the duration of the work experience (in years), $i = 1, \ldots, 888$ and $j = 1, \ldots, N_i$. We have two predictor variables. The first one is the *race* variable with three levels (*black, Hispanic and white*), hence generating two dummy variables – $(X^{(1)}(T), X^{(2)}(T)) = (1, 0)$, $(X^{(1)}(T), X^{(2)}(T)) = (0, 1)$ and $(X^{(1)}(T), X^{(2)}(T)) = (0, 0)$ standing for *black*, *Hispanic* and *white*, respectively. The second one is $X^{(3)}(T)$ (*hgc*) the highest grade completed by the individual. The covariates are standardized as in Section 3.2.1, so that the estimated coefficients have an equivalent scale.

We check the constancy of the coefficients for the signal and/or the variability function (Table 3.3) using the four tests in Section 3.1.1. The four tests lead to similar conclusions, with exception of the LRT test for testing constancy of the coefficient $\theta_0(T)$. Our conclusion is based on the L_1, L_2 and $L_{\max}$ tests results, since these tests perform better than the LRT test in the simulation study. The aforementioned tests indicate that all the covariates have non-varying coefficients with respect to both the signal and variability functions.

Further, we check the monotonicity or convexity of $\beta_0(T)$ in Table 3.4. Using both tests in Sections 3.1.2 and 3.1.3, we do not have enough arguments to reject the hypothesis that the coefficient function is non-decreasing or convex. This means, for *white* individuals who have lower *hgc* level, that as the duration of the work experience increases the median hourly wage is non-decreasing.

Table 3.3. Wages data. P-values for constancy.

Coefficients	LRT	L_1	L_2	$L_{\max}$
$\beta_0(T)$	0.1200	0.1750	0.2300	0.3500
$\beta_1(T)$	0.5000	0.3850	0.3400	0.3200
$\beta_2(T)$	0.8550	0.7150	0.7250	0.7350
$\beta_3(T)$	0.8750	0.8100	0.8100	0.8100
$\theta_0(T)$	0.0350	0.0500	0.1250	0.1700
$\theta_1(T)$	0.5400	0.1350	0.1600	0.2200
$\theta_2(T)$	0.0800	0.0950	0.1000	0.1000
$\theta_3(T)$	0.1950	0.0700	0.0650	0.0650

Table 3.4. Wages data. P-values for monotonicity and convexity of $\beta_0(T)$.

Null hypothesis	LRT	$d_{\min}$
Monotone increasing	0.6600	1.0000
Convex	0.8400	0.6250

3.4 Conclusion

We consider constancy checking for the coefficients in the general heteroscedastic model proposed in the previous chapter, using several types of tests. The coefficients in both the signal and variability functions (the variance of errors) are allowed to be more flexible via the varying coefficient models. We compare the performance of the LRT proposed in Chapter 2 and RS test of Wang et al. (2009) with the other three tests proposed in this chapter. These tests are applied on the signal (median function) as well as on the variability function. We apply the testing procedures in a simulation study and a data example. For the signal, the simulation study in Section 3.2.1 shows that the three tests (L_1, L_2 and $L_{\max}$) constructed in this chapter have a performance somewhat comparable to that of the other two tests, although none is dominating (cf the crossing power curves). For the variability function however, the proposed tests have better overall (size and power) performances.

Further, we propose two testing procedures to test for monotonicity or convexity.

Both tests perform well in the simulation study (Sections 3.2.2). Furthermore, we investigate a coefficient's convexity in the signal and its constancy in the variability function (Section 3.2.3), using the LRT test procedure in Section 3.1.4. This test can be generalized to any other shapes either in the signal or the variability functions. For instance, we can test monotonicity of a coefficient function in the signal and constancy of a coefficient function in the variability function, simultaneously.

Although the proposed testing procedures in this chapter have good performance in the simulation studies, it is of interest to prove their consistency. If no modeling bias is assumed, this can be shown similarly as Knight (1999) (see also the discussion in Section 2.5).

Part II

Variable selection in quantile varying coefficient models

Chapter 4

Variable selection in quantile varying coefficient models with heteroscedastic error

As in the previous chapter, we consider the location-scale varying coefficient model in Equation (1.7) with the variability function of the form given in Equation (3.1).

To get a nuanced picture for the relationship between a response and covariates, one can consider several regression quantiles. If the variability function is, however, of a homoscedastic structure, investigating several regression quantiles is time consuming. Under a homoscedastic model, the effect of the covariates is the same on the entire distribution of the response. The computational time is larger, especially when non-parametric estimations are considered. Therefore, in Chapter 2 (Gijbels et al., 2018), we study several structures of the variability function, such as power and exponential functions; proposing also a Likelihood-Ratio-Type test (LRT) to choose between two variability structures. Further, in Chapter 3 (Gijbels et al., 2017), we develop several constancy testing procedures in addition to the LRT for both the median and variability functions to find the best suitable model among a family of models. The LRT can also be used to do variable selection. However, the LRT can be time consuming and the type-I error rate might inflate when a large number of covariates are considered. This is because in the LRT, we have to compare various models with several combinations of covariates included.

Variable selection is a hot topic in the literature nowadays due to the application of high-dimensional data, where a lot of covariates are collected and stored. Several popular methods have first been introduced in the parametric mean regression (least

squares), including the Least absolute shrinkage and selection operator (Lasso) (Tibshirani, 1996), the Smoothly Clipped Absolute Deviation (SCAD) (Fan and Li, 2001), the Adaptive Lasso (Zou, 2006), the one-step sparse estimator (Zou and Li, 2008) and the elastic net (Zou and Zhang, 2009). A lot of research has also been done in the non-parametric regression setting (Wang and Xia, 2009; Xue, 2009; Noh and Park, 2010; Huang et al., 2012; Ma et al., 2013; Antoniadis et al., 2014; Wang and Lin, 2014; Wang et al., 2014; Yang et al., 2016; Li et al., 2017, to cite a few). A brief discussion on grouped regularization techniques and on a robust variable selection method is given by Gijbels et al. (2015).

In an attempt to have variable selection techniques which are robust to outliers and non-normal errors, robust regressions such as least absolute deviation and quantile regression have been proposed (Li and Zhu, 2008; Zou and Yuan, 2008; Wu and Liu, 2009b). For the non-parametric quantile regression setting, SCAD (Kai et al., 2011; Noh et al., 2012; Lee et al., 2014; Guo et al., 2017), Adaptive Lasso (Jiang et al., 2012; Tang et al., 2012, 2013b; Zhao et al., 2013), a smooth-threshold estimating equation (Wang and Sun, 2017), a Boosting approach (Fenske et al., 2011) and NonNegative Garrote (Lin et al., 2013) have been considered, among others.

In contrast to the above mentioned literature, apart from the location, the variability function is in addition investigated in this chapter. We extend the grouped Adaptive Lasso by Tang et al. (2013b) to our context where we use P-splines (gALassoP). Furthermore, we propose a NonNegative Garrote (NNG) estimator. We show the consistency of the gALassoP.

The remainder of the chapter is organized as follows. Section 4.1 deals with the estimation methods and the asymptotic results. Simulation studies are carried out in Section 4.2. The proposed variable selection techniques are applied on two data examples in Section 4.3. Section 4.4 concludes the results in this chapter. Finally, Section 4.5 presents theoretical details. The R-code to implement the methods of this chapter on a data set is deferred to Appendix A and is available at `http://ibiostat.be/online-resources/online-resources/longitudinal`.

4.1 Estimation Methods

We use the Adaptive He (AHe) approach (He, 1997; Andriyana et al., 2016; Andriyana and Gijbels, 2017; Gijbels et al., 2017, 2018) to estimate the median as well as the

variability function.

4.1.1 Step 1: estimation of the median

Assuming **H1** in Chapter 2, the median function of $Y(T)$ is given in Equation (2.2).

Grouped Adaptive Lasso

Using unpenalized B-splines, Tang et al. (2013b) propose a grouped Adaptive Lasso approach (gALassoB). A penalty (in L_1-norm) for each coefficient group, corresponding to a particular functional coefficient, is added to the objective function (i.e. Equation (4.1) with $\lambda_P = 0$). Here we combine it with the P-splines technique in Equation (2.6) and write the penalized objective function as

$$S(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left(Y_{ij} - \mathbf{U}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}\right) + \lambda_P \sum_{k=0}^{p} \omega_{1k} \|\mathbf{D}_{m_k}^{1} \boldsymbol{\alpha}_k\|_1 + \lambda_L \sum_{k=0}^{p} \omega_{2k} \|\boldsymbol{\alpha}_k\|_1, \quad (4.1)$$

where λ_L is a nonnegative regularization parameter that determines the sparsity of the solution; $\omega_{2k} = \|\hat{\boldsymbol{\alpha}}_k^B\|_1^{-\eta_L}$, with $\eta_L \geq 1$ and $k = 0, \ldots, p$, are the penalty weights of the coefficient groups; $\hat{\boldsymbol{\alpha}}_k^B$ are the coefficient estimates using unpenalized B-splines (i.e. when $\lambda_P = \lambda_L = 0$ in Equation (4.1)). Tang et al. (2013b) showed from their simulation study that the grouped Adaptive L_1- and L_2-penalties perform equivalently well in finite samples. The objective function with the L_1-norm penalty can be solved by a standard linear programming. Therefore, using the L_1-norm penalty is computationally more convenient than the L_2-norm. The data-driven smoothing parameters λ_P and λ_L are obtained using the following SIC-criterion (Schwarz, 1978; Tang et al., 2013b; Gijbels et al., 2018) on a two-dimensional grid:

$$\mathrm{SIC}(\lambda_P, \lambda_L) = \ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(Y_{ij} - \mathbf{U}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\alpha}}\right)\right) + \frac{\ln(N)}{2N} p_{\lambda_P,\lambda_L}, \quad (4.2)$$

where $N = \sum_{i=1}^{n} N_i$; $\hat{\boldsymbol{\alpha}}$ is the minimizers in Equation (4.1); and p_{λ_P,λ_L} is the effective degrees of freedom, taken to equal the size of the elbow set $\mathcal{E}_\lambda$ in Equation 2.12.

NonNegative Garrote

In a classical multiple linear regression model, Breiman (1995) proposes the NNG for subset regression. Here we use this methodology by first obtaining the initial estimates of $\beta_k(t)$ ($\hat{\beta}_k^{\text{init}}(t)$), the minimizers of the objective function in Equation (2.6) (P-splines). Then, using these estimates, the NonNegative Garrote shrinkage factors $\hat{\mathbf{c}} = (\hat{c}_0, \ldots, \hat{c}_p)^{\mathrm{T}}$ are obtained by solving the following optimization problem:

$$\min_{c_0,\ldots,c_p} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left(Y_{ij} - \sum_{k=0}^{p} c_k \hat{\beta}_k^{\text{init}}(t_{ij}) X_{ij}^{(k)}\right) + \lambda_{\text{NNG}} \sum_{k=0}^{p} c_k \text{ s.t. } c_k \geq 0 \ (k = 0, 1, \ldots, p). \tag{4.3}$$

The NonNegative Garrote estimator for $\beta_k(t)$ is given by

$$\hat{\beta}_k^{\text{NNG}}(t_{ij}) = \hat{c}_k \hat{\beta}_k^{\text{init}}(t_{ij}). \tag{4.4}$$

To obtain the tuning parameter λ_{NNG}, we use the SIC-criterion:

$$\text{SIC}(\lambda_{\text{NNG}}) = \ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(Y_{ij} - \sum_{k=0}^{p}\hat{c}_k\hat{\beta}_k^{\text{init}}(t_{ij})X_{ij}^{(k)}\right)\right) + \frac{\ln(N)}{2N} p_{\lambda_{\text{NNG}}}. \tag{4.5}$$

4.1.2 Step 2: estimation of the variability function

When the interest is in studying several quantiles, we have to estimate the variability function. From the model in Equation (1.7) and since $V(\mathbf{X}(T), T) \geq 0$, we have Equation (2.7). Therefore, assuming **H2** in Chapter 2, we can estimate $V(\cdot,\cdot)$ based on the following equation,

$$\begin{aligned} q_{0.5}\left(\ln|Y(T) - \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\beta}(T)| \,\middle|\, \mathbf{X}(T), T\right) &= \ln V(\mathbf{X}(T), T) \\ &= \sum_{k=0}^{p} \theta_k(T) X^{(k)}(T) = \mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\theta}(T), \end{aligned} \tag{4.6}$$

with $\theta_0(T) = \ln\{\gamma(T)\}$ and $\boldsymbol{\theta}(T) = (\theta_0(T), \ldots, \theta_p(T))^{\mathrm{T}}$. Then, $\theta_k(T)$ can be approximated in a basis (of size $m_k^{\text{V}} = u_k^{\text{V}} + \nu_k^{\text{V}}$) of normalized B-splines of degree ν_k^{V} with $u_k^{\text{V}} + 1$ quasi-uniform knots.

Similar variable selection procedures as for the median can be constructed for the variability function. For the gALassoP, the estimated functional coefficients are obtained by minimizing the following objective function, where $\mathbf{U}_{ij}^{\text{V}} =$

$\left(\left(\mathbf{U}^{\mathrm{V}}{}_{ij}^{(1)}\right)^{\mathrm{T}},\ldots,\left(\mathbf{U}^{\mathrm{V}}{}_{ij}^{(p)}\right)^{\mathrm{T}}\right)^{\mathrm{T}}_{m^{\mathrm{V}}_{tot}\times 1}$ with $\mathbf{U}^{\mathrm{V}}{}_{ij}^{(k)} = \mathbf{B}^{\mathrm{V}}_k(t_{ij},\nu^{\mathrm{V}}_k)X^{(k)}_{ij}$ and the weights $\omega^{\mathrm{V}}_{2k} = \|(\hat{\boldsymbol{\alpha}}^{\mathrm{V}})^B_k\|_1^{-\eta^{\mathrm{V}}_L}$ are based on the (unpenalized) B-splines estimates $(\hat{\boldsymbol{\alpha}}^{\mathrm{V}})^B_k$ (i.e when $\lambda^{\mathrm{V}}_P = \lambda^{\mathrm{V}}_L = 0$ in Equation (4.7)) with $\eta^{\mathrm{V}}_L \geq 1$:

$$S(\boldsymbol{\alpha}^{\mathrm{V}}) = \sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\Big(\ln(|Y_{ij} - \mathbf{X}^{\mathrm{T}}_{ij}\hat{\boldsymbol{\beta}}(t_{ij})|) - (\mathbf{U}^{\mathrm{V}}_{ij})^{\mathrm{T}}\boldsymbol{\alpha}^{\mathrm{V}}\Big) + \lambda^{\mathrm{V}}_P\sum_{k=0}^{p}\omega^{\mathrm{V}}_{1k}\|\mathbf{D}^{d^{\mathrm{V}}_k}_{m^{\mathrm{V}}_k}\boldsymbol{\alpha}^{\mathrm{V}}_k\|_1 + \lambda^{\mathrm{V}}_L\sum_{k=0}^{p}\omega^{\mathrm{V}}_{2k}\|\boldsymbol{\alpha}^{\mathrm{V}}_k\|_1, \tag{4.7}$$

where λ^{V}_L is a nonnegative regularization parameter.

For the NNG, the estimated functional coefficients are obtained by solving,

$$\min_{c^{\mathrm{V}}_0,\ldots,c^{\mathrm{V}}_p}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(\ln(|Y_{ij} - \mathbf{X}^{\mathrm{T}}_{ij}\hat{\boldsymbol{\beta}}(t_{ij})|) - \sum_{k=0}^{p}c^{\mathrm{V}}_k\hat{\theta}^{\mathrm{init}}_k(t_{ij})X^{(k)}_{ij}\right) + \lambda^{\mathrm{V}}_{\mathrm{NNG}}\sum_{k=0}^{p}c^{\mathrm{V}}_k$$
$$\text{s.t. } c^{\mathrm{V}}_k \geq 0 \ (k = 0,1,\ldots,p), \tag{4.8}$$

where $\hat{\theta}^{\mathrm{init}}_k(t_{ij})$ are the estimates obtained minimizing Equation (4.7) with $\lambda^{\mathrm{V}}_L = 0$ and $\lambda^{\mathrm{V}}_{\mathrm{NNG}} \geq 0$ is the regularization parameter.

The implementation of the estimation procedures can be written similarly as in Chapter 2. We use the Frisch-Newton interior point algorithm (Portnoy and Koenker, 1997), used in Tang et al. (2013b) after updating it to our procedures.

4.1.3 Asymptotic properties of the grouped Adaptive Lasso

In this section, we show that the gALassoP approach leads to a consistent variable selection procedure and the resulting estimator achieves the optimal convergence rate. We assume that only the first s covariates are relevant, i.e. $\beta_k(\cdot) \neq 0$ for $k = 0,1,\ldots,s$ and $\beta_k(\cdot) = 0$ for $k = s+1,\ldots,p$.

For the first step of our estimation procedure in Equation (4.1), where the $\epsilon(T)$ is assumed to have median zero given $(\mathbf{X}(T),T)$ (**H1**), we have the following main theorem for the gALassoP (see Section 4.5.2 for the proof). The extra notations and assumptions are stated in Section 4.5.1.

Theorem 4.1 (gALassoP) *Suppose Assumptions* **2.2**, **2.3**, **2.5**, **2.6** *and* **4.1** – **4.5** *hold. Furthermore, assume that* $u^{3/4}_{\max}\lambda_{P\max}n^{-1/2} \to 0$, $\lambda_L u^{1/2}_{\max}n^{-1/2} \to 0$ *and* $u^{-3/2}_{\max}\lambda_L \to \infty$*; as n tends to* ∞*. Then*

(a) $\hat{\beta}_k(\cdot) = 0$, $k = s+1, \ldots, p$, *with probability approaching 1;*

(b) $\|\hat{\beta}_k - \beta_k\|_2 = O_p\left(n^{-r_{\min}/(2r_{\min}+1)}\right)$, $k = 0, 1, \ldots, s$.

Theorem 4.1 **(a)** says that the gALassoP is consistent in variable selection. As provided in Theorem 4.1 **(b)**, the resulting estimator has an optimal global convergence rate for nonparametric regression (Stone, 1982).

The next theorem states the corresponding asymptotic behavior for the variability function. We assume there are s^{V} relevant covariates in the model, with $\theta_k(\cdot) \neq 0$ for $k = 0, 1, \ldots, s^{\mathrm{V}}$ and $\theta_k(\cdot) = 0$ for $k = s^{\mathrm{V}}+1, \ldots, p$. Relying on Theorem 4.1, we have Equation (2.13).

Assuming **H2**, we have the following theorem for the gALassoP (see Section 4.5.3 for the proof).

Theorem 4.2 (gALassoP) *Suppose Assumptions* **2.6**, **2.9** – **2.12**, **4.2**, **4.4**, **4.6** *and* **4.7** *hold. Furthermore, assume that* $(u^{V}_{\max})^{3/4}\lambda^{V}_{P\max}n^{-1/2} \to 0$, $\lambda^{V}_{L}(u^{V}_{\max})^{1/2}n^{-1/2} \to 0$ *and* $(u^{V}_{\max})^{-3/2}\lambda^{V}_{L} \to \infty$*; as* n *tends to* ∞*. Then,*

(a) $\hat{\theta}_k(\cdot) = 0$, $k = s^{V}+1, \ldots, p$, *with probability approaching 1;*

(b) $\|\hat{\theta}_k - \theta_k\|_2 = O_p(n^{-r^{V}_{\min}/(2r^{V}_{\min}+1)})$, $k = 0, 1, \ldots, s^{V}$.

Remark 4.1 *The same proof can be used to generalize the asymptotic properties in Theorems 4.1 and 4.2 to the grouped Adaptive* L_κ*-penalty with any* $\kappa \geq 1$*. Hence, the results in this Chapter can be extended to* L_κ*-penalty, but with the disadvantage of not being able to use linear programming.*

4.2 Simulation study

We investigate the performance of the gALassoP and NNG approaches defined in Section 4.1, using two settings: homoscedastic and heteroscedastic models. The two methods are also compared with the gALassoB in Tang et al. (2013b), the grouped SCAD (gSCADB) in Noh et al. (2012) and the grouped SCAD (gSCADP) when adding the penalization term in Equation (2.6) ($\lambda_P \sum_{k=0}^{p} \omega_{1k} \sum_{l=d_k+1}^{m_k} |\Delta^{d_k}\alpha_{kl}|$) to the objective function in Noh et al. (2012). R-software which uses Fortran code from the quantreg R-package (Koenker et al., 2017) is used for the gALassoP, NNG and gALassoB. Where as for the gSCADB and gSCADP a Matlab-based modeling system CVX is used.

Remark 4.2 *We can easily show that the gSCADB has the same convergence rate as the gALassoP by making use of Equation (6.5) in Noh et al. (2012) and Lemma 4.5. Similarly, it is easy to show that the gSCADP has the same convergence rate as the gALassoP by making use of Equation (6.5) in Noh et al. (2012), Lemma 4.5 and Equation* (4.11).

4.2.1 Homoscedastic model

We simulate from the following model (considered by Tang et al. (2013b)),

$$Y_{ij} = \beta_0(t_{ij}) + \sum_{k=1}^{23} \beta_k(t_{ij}) X_{ij}^{(k)} + 1.75\epsilon_{ij}.$$

The coefficients (also presented in Figure 4.1) are formulated as follows,

$$\beta_0(T) = 15 + 20\sin(\pi T/60) \quad \beta_1(T) = 2 - 3\cos(\pi(T-25)/15) \quad \beta_2(T) = 6 - 0.2T$$
$$\beta_3(T) = -4 + (20-T)^3/2000 \quad \beta_k(T) = 0 \text{ for } k = 4, \ldots, 23.$$

We use 1.75 for the variability function ($V(\mathbf{X}_{ij}, t_{ij})$) such that the Signal-to-Noise Ratio is approximately seven. The error term is generated from a transformed multivariate normal distribution as in Equation (2.14) (to fulfill assumptions **H1** and **H2** in Section 2.1), with $\zeta_{ij} = E_{ij} + Z_{ij}$, $E_{ij} \sim N(0,4)$ and Z_{ij} is generated from a Gaussian process with zero mean and a decayed exponential covariance

$$cov(Z_{ij}, Z_{lz}) = \begin{cases} 4\exp(-|t_{ij} - t_{lz}|) & \text{if } i = l, \\ 0 & \text{if } i \neq l. \end{cases}$$

The covariate $X_{ij}^{(1)}$ is generated from a Unif$[t_{ij}/10, 2 + t_{ij}/10]$, $X_{ij}^{(2)} \sim \mathrm{N}\left(0, \frac{1+X_{ij}^{(1)}}{2+X_{ij}^{(1)}}\right)$, $X_{ij}^{(3)} \sim \text{Bern}(0.6)$, and the rest are redundant variables, $X_{ij}^{(k)}$ for $k = 4, \ldots, 23$, having the same distribution as Z_{ij}.

The time variable T ranges from $1, 2, \ldots, 30$. We have an unbalanced design: each time point (excluding time 1) has 60% probability of being skipped. Then the actual time point is obtained by adding a Unif$[-0.5, 0.5]$.

We simulated 500 samples of size $n = 200$. B-splines of degree three with eight quasi-uniform knots and differencing order one are used for the gALassoP, NNG and

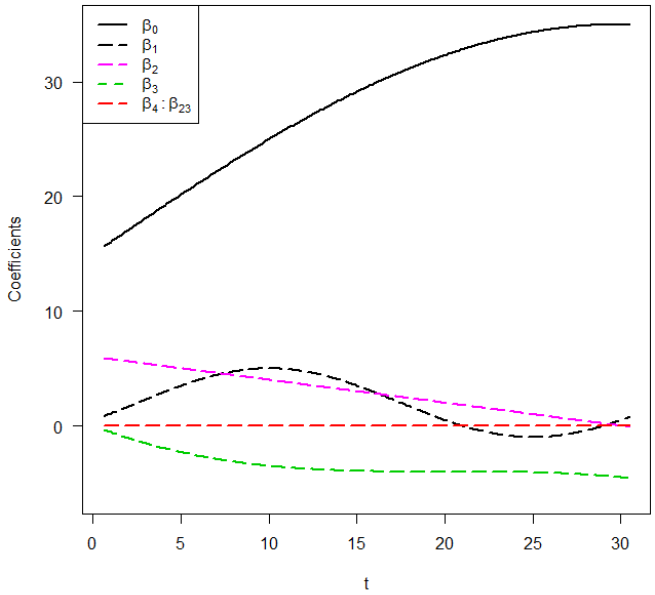


Figure 4.1. True coefficients of the median.

gSCADP. Where as for the gALassoB and gSCADB, the number of knots are chosen adaptively from a grid between four and eight (similar as in Tang et al., 2013b). In all the methods, we have two tuning parameters for smoothness and sparsity. Hence, we search over a two-dimensional grid of the same size for all methods to make it comparable. We set $\eta_P = 1$ and $\eta_L = 2$ (as in Tang et al., 2013b). The cut-off criterion to keep the covariates is $\|\boldsymbol{\alpha}_k\|_1 > 10^{-6}$ for the gALassoB, gALassoP, gSCADB and gSCADP, and is $c_k > 10^{-6}$ for the NNG.

We measure the ability of the methods to find the true model (oracle), those containing the true model (good), models with one extra variable (+1), those that remove one true variable (-1). We further look at the median number of selected variables that are truly relevant (Med.r), the median number of selected variables that are redundant (Med.z) and the median computing time in seconds (Med.t) (see Table 4.1).

The procedures have equivalent performance in selecting the significant variables for both the signal and variability functions (Table 4.2). However, the NNG outperforms the other procedures in terms of selecting the true model and/or the computing time.

We also compare the methods by the root approximated integrated square errors

Table 4.1. Measures of the performance.

Measure	Definition
Oracle %	the percentage of replications that the exact true model is selected
Good %	the percentage of replications that the truly relevant variables are included in the model
+1 %	the percentage of replications that one extra variable is included in the model
-1 %	the percentage of replications that one important variable is removed from the model
Med.r	the median number of selected variables that are truly relevant
Med.z	the median number of redundant variables selected
Med.t	the median time in seconds

defined as follows:

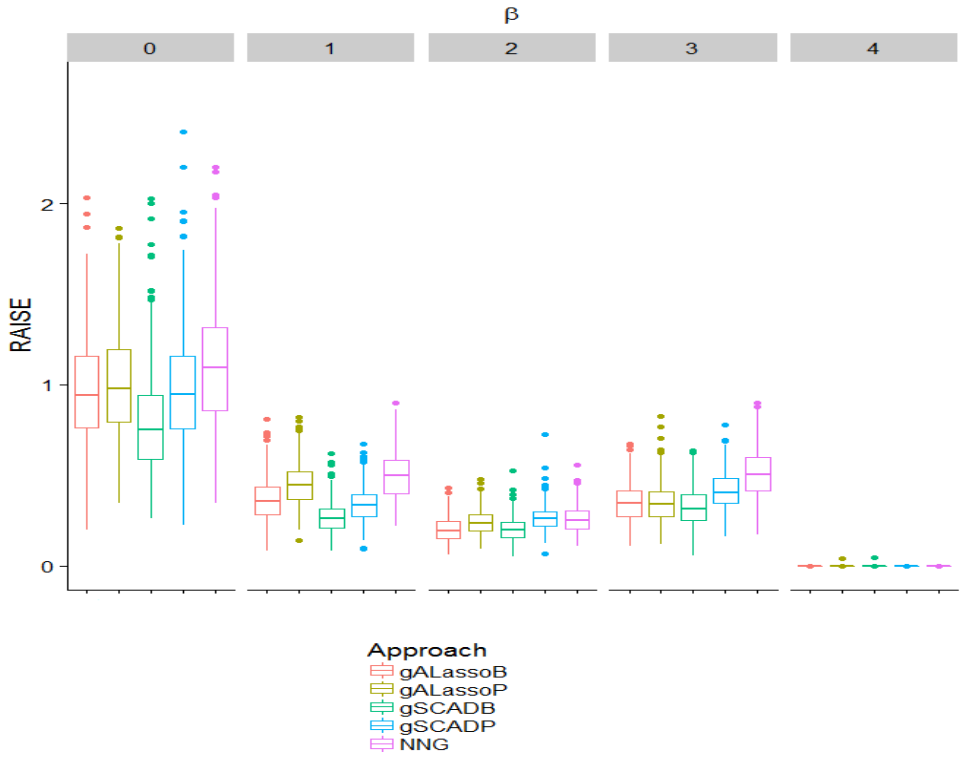
$$RAISE(\hat{\beta}_k^{(i)}(\cdot)) = \left(\frac{1}{301}\sum_{a=1}^{301}[\hat{\beta}_k^{(i)}(t_a) - \beta_k^{(i)}(t_a)]^2\right)^{\frac{1}{2}}, \text{ for simulation } i,$$

$$RAISE(\hat{\theta}^{(i)}(\cdot)) = \left(\frac{1}{301}\sum_{a=1}^{301}[\hat{\theta}_k^{(i)}(t_a) - \theta_k^{(i)}(t_a)]^2\right)^{\frac{1}{2}}, \text{ for simulation } i,$$

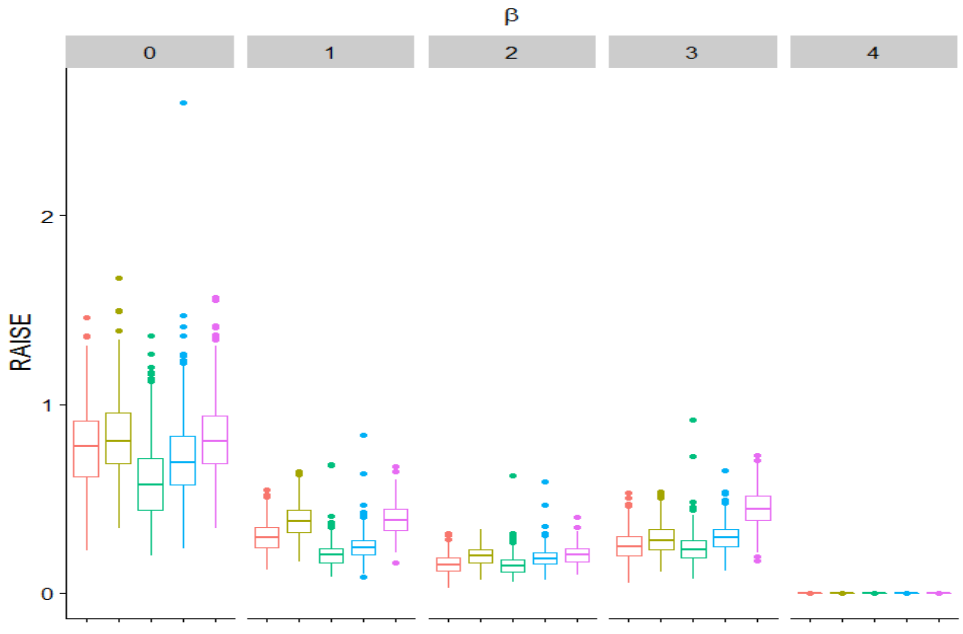
where $t_a = a/10, a = 5, 6, \ldots, 305$.

Figure 4.2 (a) shows that all the methods have a similar efficiency in estimating the coefficients of the median. However, the gSCADB and gSCADP have bad performance compared to the others for the variability function (Figure 4.3 (a)), which might be surprising. Note that the estimated coefficients are close to zero for the irrelevant covariates ($X^{(4)}(T)$ and $X^{(1)}(T)$, respectively, for the median and the variability functions), which supports Theorems 4.1 (a) and 4.2 (a). The boxplots for the other irrelevant variables (not presented) are also similar (close to zero).

Furthermore, we look at the performance of the methods when increasing the sample size to $n = 400$. From Table 4.3 we conclude that, for the signal, the oracle percentages of the gALassoP and gALassoB improved, as expected. From Figures 4.2 (b) and 4.3 (b), we note that all the methods improve in terms of estimation accuracy. The relative median computing time to estimate the median function for the NNG, compared to that of the gALassoB and gSCADB, respectively, is 0.159 and 0.008. The grouped SCAD approach (whether or not P-splines smoothness penalty is considered) has bad estimation for the variability function (Figure 4.3).

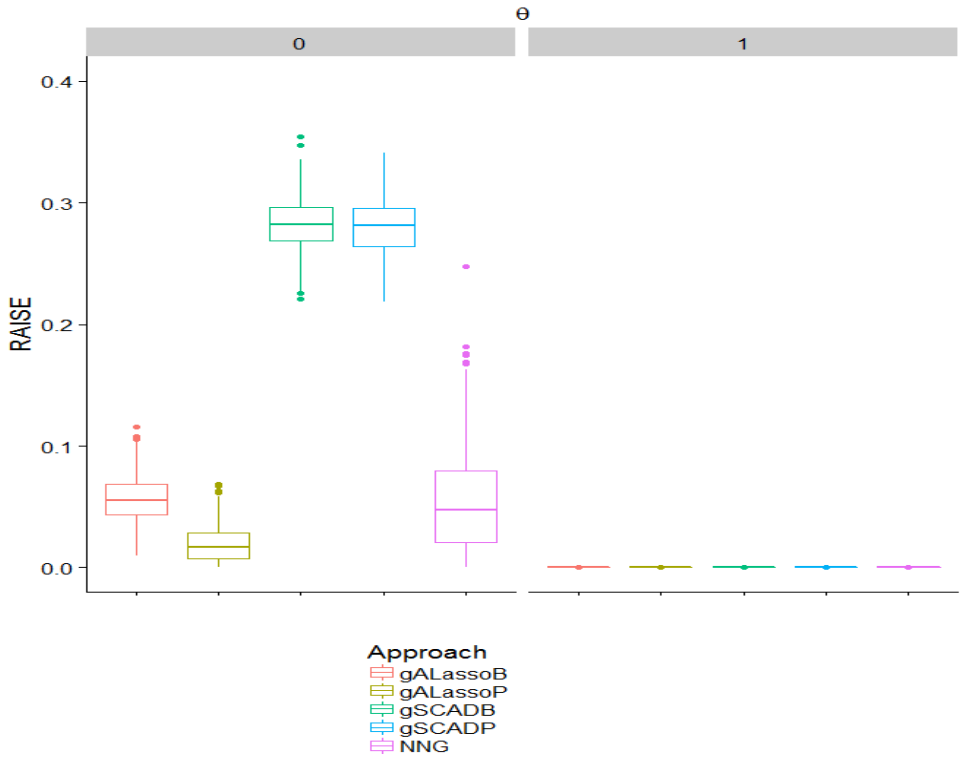


(a) $n = 200$

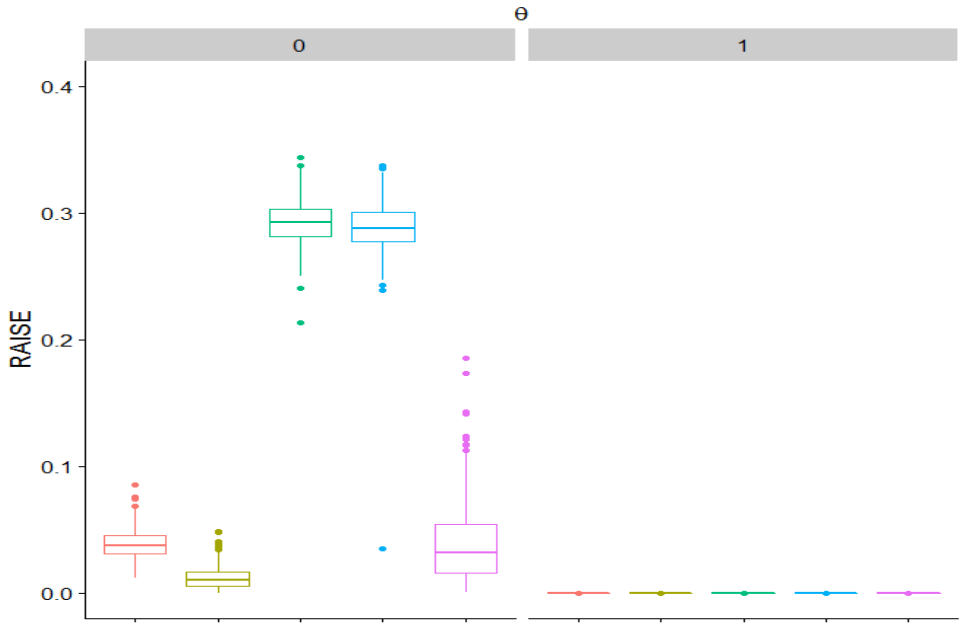


(b) $n = 400$

Figure 4.2. $RAISE(\hat{\beta}_k^{(i)}(\cdot))$ for the homoscedastic model.



(a) $n = 200$



(b) $n = 400$

Figure 4.3. $RAISE(\hat{V}^{(i)}(\cdot,\cdot))$ for the homoscedastic model.

Table 4.2. Performance of the variable selection techniques for the homoscedastic model with $n = 200$.

	Median					Variability				
Method	gAL_P	gAL_B	NNG	gS_B	gS_P	gAL_P	gAL_B	NNG	gS_B	gS_P
Oracle %	98.4	99.4	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0
Good %	100.0	100.0	100.0	100.0	100.0	–	–	–	–	–
+1 %	1.6	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-1 %	0.0	0.0	0.0	0.0	0.0	–	–	–	–	–
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	–	–	–	–	–
Med.z	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Med.t	36.1 (35.7, 36.6)	28.9 (28.3, 29.6)	4.6 (4.5, 4.7)	563.6 (554.4, 572.2)	604.9 (594.4, 614.3)	36.3 (35.4, 37.1)	26.3 (25.6, 26.9)	4.0 (3.9, 4.1)	581.8 (573.1, 592.5)	620.8 (611.4, 633.3)

NOTE: gAL_P: gALassoP; gAL_B: gALassoB; gS_B: gSCADB; gS_P: gSCADP; the numbers in brackets are the first and third quartiles.

Table 4.3. Performance of the variable selection techniques for the homoscedastic model with $n = 400$.

	Median					Variability				
Method	gAL_P	gAL_B	NNG	gS_B	gS_P	gAL_P	gAL_B	NNG	gS_B	gS_P
Oracle %	100.0	99.8	100.0	99.8	99.8	100.0	100.0	100.0	100.0	100.0
Good %	100.0	100.0	100.0	100.0	100.0	–	–	–	–	–
+1 %	0.0	0.2	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.0
-1 %	0.0	0.0	0.0	0.0	0.0	–	–	–	–	–
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	–	–	–	–	–
Med.z	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Med.t	77.1 (76.1, 77.8)	57.2 (56.5, 58.0)	8.1 (8.0, 8.3)	2054.9 (2019.2, 2082.4)	2147.9 (2099.7, 2181.3)	66.3 (64.6, 67.9)	50.3 (49.4, 51.3)	6.9 (6.8, 7.1)	2102.0 (2067.2, 2132.9)	2194.1 (2151.6, 2234.2)

NOTE: gAL_P: gALassoP; gAL_B: gALassoB; gS_B: gSCADB; gS_P: gSCADP; the numbers in brackets are the first and third quartiles.

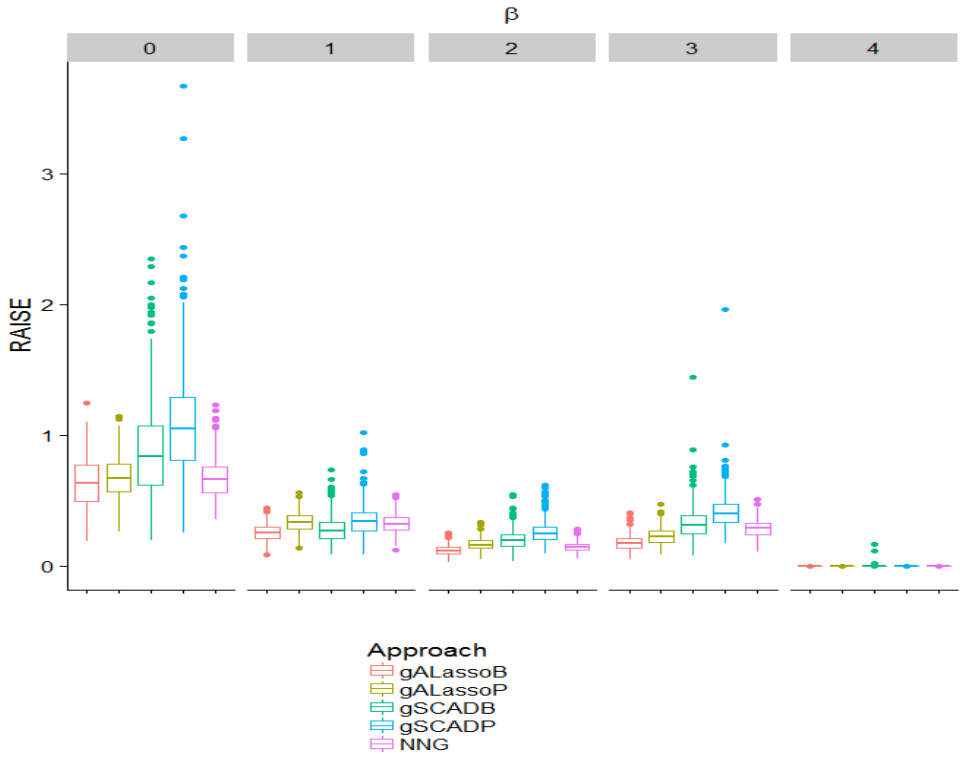
4.2.2 Heteroscedastic model

In this section we consider a similar setting as in Section 4.2.1 with $n = 200$, but now with a heteroscedastic error $\left(V(\mathbf{X}(T), T) = \exp\left(\frac{\beta_3(T)}{10} X^{(6)}(T)\right)\right)$. For the signal, the conclusion is similar to that of the previous section, indicating that the methods have similar performance with respect to variable selection and estimation accuracy (see Table 4.4 and Figure 4.4 (a), respectively). For the variability function, the gSCADB and gSCADP could not identify the important variable. The estimation accuracy of the gSCADB, gSCADP and gALassoB is also bad for the important variable in the variability function (Figure 4.4 (b)). Tang et al. (2013b) also indicate in one of their simulation setting that the gSCAD shows bad performance, in a mean regression setting. Again, the NNG outperforms the other methods with respect to the execution time. The relative median computing time to estimate the median function for the NNG, compared to that of the gALassoB and gSCADB, respectively, is 0.146 and 0.008. The gALassoB has bad estimation performance for the coefficient of the important variable in the variability function, which calls for the importance of considering the penalized B-splines.

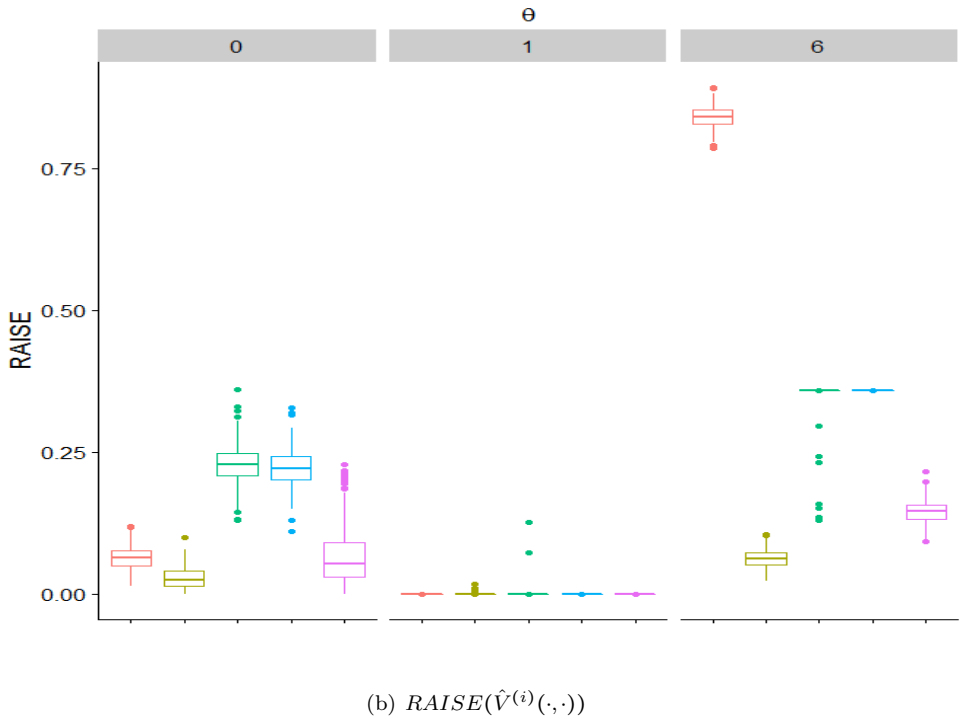
Table 4.4. Performance of the variable selection techniques for the heteroscedastic model with $n = 200$.

	Median					Variability				
Method	gAL_P	gAL_B	NNG	gS_B	gS_P	gAL_P	gAL_B	NNG	gS_B	gS_P
Oracle %	100.0	100.0	100.0	98.6	100.0	99.0	99.8	100.0	1.0	0.0
Good %	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	1.4	0.0
+1 %	0.0	0.0	0.0	0.6	0.0	1.0	0.2	0.0	0.4	0.0
-1 %	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.6	100.0
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	1 (1, 1)	1 (1, 1)	1 (1, 1)	0 (0, 0)	0 (0, 0)
Med.z	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Med.t	37.1 (36.6, 37.7)	32.8 (31.5, 34.6)	4.8 (4.7, 4.9)	572.0 (562.9, 580.2)	605.6 (596.3, 615.9)	36.7 (35.9, 37.5)	28.6 (27.6, 30.0)	4.1 (4.0, 4.1)	612.6 (603.4, 623.6)	641.0 (631.2, 652.4)

NOTE: gAL_P: gALassoP; gAL_B: gALassoB; gS_B: gSCADB; gS_P: gSCADP; the numbers in brackets are the first and third quartiles.



(a) $RAISE(\hat{\beta}_k^{(i)}(\cdot))$



(b) $RAISE(\hat{V}^{(i)}(\cdot,\cdot))$

Figure 4.4. *RAISE* for the heteroscedastic model.

4.3 Data examples

In this section, we study three data examples; containing two, 22 and 27 covariates, respectively, for Grunfeld, KTB and Intego data.

4.3.1 Grunfeld data

This data example contains annual measurements for 11 US manufacturing firms over 20 years, for the years 1935–1954. The response variable is gross investment at year T. Each firm has the same number of measurements $N_i = 20$. There are two covariates in the data set, the market value of the firm $X^{(1)}(T)$ and the stock of plant and equipment (capital) $X^{(2)}(T)$. See AER R-package (Kleiber and Zeileis, 2008) for more information on the data. The data is first studied by Grunfeld (1958). It is also considered by Andriyana (2015), where a general heteroscedastic model is recommended.

The covariates are standardized as in Section 3.2.1, so that they have equivalent scale. B-splines of degree three with six quasi-uniform knots and differencing order one are used for the gALassoP, NNG and gSCADP. Where as for the gALassoB and gSCADB, the number of knots are chosen adaptively from a grid between four and six.

Table 4.5 shows that, using all the procedures discussed in Section 4.2, the two covariates are selected in the signal part. For the variability function, all the procedures other than the gSCADB select both the covariates (Table 4.6). To assess the selection stability of the methods, we obtain 200 bootstrap samples from the data. For each bootstrap, we resample the firms $(Y(t_{ij}), X^{(1)}(t_{ij}), X^{(2)}(t_{ij}), j = 1, \ldots, N_i)$ with replacement and apply the variable selection methods to the bootstrap sample. We see from Table 4.5 that all the methods are stable for the signal. However, the gSCADB is not stable for the variability function (Table 4.6). Using the NNG, $X^{(2)}(T)$ is selected 194 times and 177 times, respectively, for the median and the variability functions. The third row of Tables 4.5 and 4.6 shows that, using the NNG, the models selected in the bootstrap samples agree with the models selected in the observed data 97% and 83.5% of times for the signal and the variability functions, respectively.

Table 4.5. Grunfeld data. Frequencies that the covariates are selected in the observed sample (O) and in the bootstrap samples (Boot) for the median function.

	gAL_P		gAL_B		NNG		gS_B		gS_P	
	O	Boot	O	Boot	O	Boot	O	Boot	O	Boot
$X^{(1)}$	1	200	1	200	1	200	1	200	1	200
$X^{(2)}$	1	200	1	200	1	194	1	200	1	200
AgreeP	–	100.0	–	100.0	–	97.0	–	100.0	–	100.0

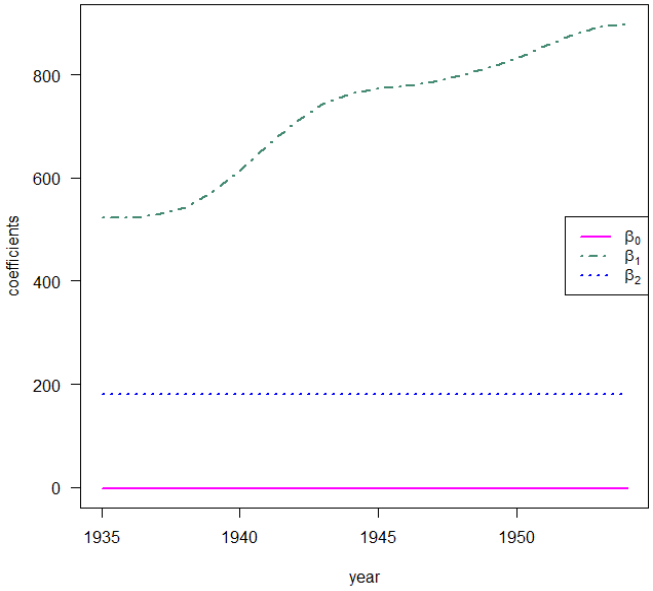
NOTE: gAL_P: gALassoP; gAL_B: gALassoB; gS_B: gSCADB; gS_P: gSCADP; AgreeP: the percentage that the model selected by the bootstrap sample agrees with the model selected by the observed sample.

Table 4.6. Grunfeld data. Frequencies that the covariates are selected in the observed sample (O) and in the bootstrap samples (Boot) for the variability function.

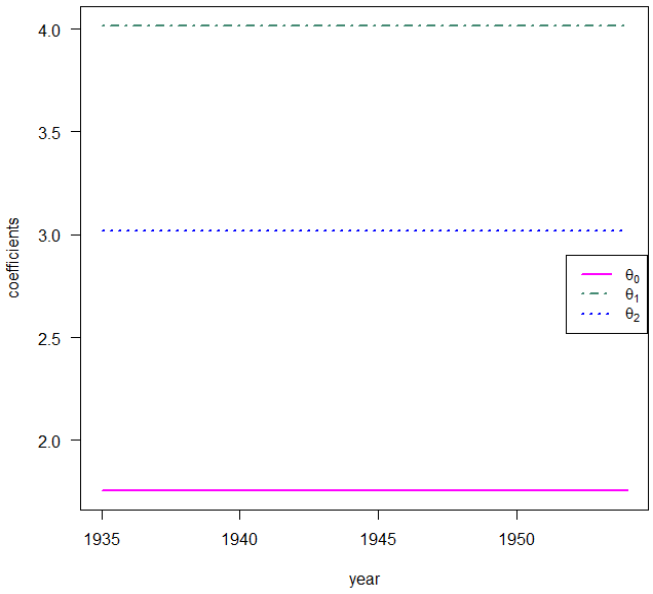
	gAL_P		gAL_B		NNG		gS_B		gS_P	
	O	Boot	O	Boot	O	Boot	O	Boot	O	Boot
$X^{(1)}(T)$	1	199	1	131	1	189	0	145	1	200
$X^{(2)}(T)$	1	192	1	191	1	177	1	176	1	200
AgreeP	–	95.5	–	63.0	–	83.5	–	22.0	–	100.0

NOTE: gAL_P: gALassoP; gAL_B: gALassoB; gS_B: gSCADB; gS_P: gSCADP; AgreeP: the percentage that the model selected by the bootstrap sample agrees with the model selected by the observed sample.

Even though the covariates are significant for the variability function, Figure 4.5 (b) shows that the estimated coefficients (obtained minimizing Equation (4.7) with $\lambda_L^{\mathrm{v}} = 0$) are constant over time. Hence, the model can be further simplified by checking the constancy of the coefficients using the testing procedures proposed in Chapter 3 or using step 1 of Tang et al. (2012).



(a) Median



(b) Variability

Figure 4.5. Grunfeld data. The estimated coefficient functions of the standardized covariates.

4.3.2 KTB data

This data example is obtained from the German Continental Deep Drill program (KTB). It contains 68 covariates of petrophysical properties and geochemical compositions from drill cuttings measured at each of the 5922 depth points down to a final depth of 9.1 km (with a distance of 1 to 2m between the cuttings). The aim of the program is to study the properties and processes of the upper 10 km of the continental crust. The response variable is the content of cataclastic rocks (the sum of three cataclastic components, namely cataclastic gneiss, cataclastic metabasite and cataclasites) in volume percent with logarithmic scale. We consider the 22 covariates (see Table 4.7) as in Winter et al. (2002), with $X^{(0)}(T) \equiv 1$. The covariates are standardized as in Section 3.2.1. Winter et al. (2002) pointed out that the composition of the cataclastic shear-zones highly depends on the lithology (and thus the depth). Hence, we allow the coefficients to vary with the depth. This data set is also investigated by Antoniadis et al. (2012b), where they also use the varying coefficient models comparing four types of variable selection techniques in a mean regression setting.

To analyze the data set, B-splines of degree three with 16 quasi-uniform knots (similar as in Antoniadis et al., 2012b) and differencing order one are used for the gALassoP, NNG and gSCADP. Where as for the gALassoB and gSCADB, the number of knots are chosen adaptively from a grid of 6-16.

As in the previous data example, we conduct variable selection using all the methods. For the median function all the methods other than the gSCADB and gSCADP select 21 covariates (Table 4.8). As shown by Antoniadis et al. (2012b), less covariates are selected for mean regression. This might be because our point of interest is now on the median of the response. For the variability function, the NNG selects 12 covariates, whereas the gALassoP selects 19 covariates. Using P-splines estimation (minimizing Equations (2.6) and (4.7) with $\lambda_L^{\mathrm{V}} = 0$, respectively, for the median and the variability functions), we plot the estimated coefficient functions for the first four covariates in Figure 4.6. We see in Figure 4.6 (a) that $\hat{\beta}_3(T)$ is almost around zero.

To quantify the effect of leaving out irrelevant variables from the model, we calculate the SIC-criterion when the estimation is done based on the selected set of variables with P-splines. For the median function, the SIC values when only $X^{(3)}(T)$ is removed, removing the irrelevant variables using the NonNegative Garrote from Antoniadis et al. (2012b), removing the irrelevant variables from Antoniadis et al. (2012b) other than $X^{(3)}(T)$, and using the full model are presented in Table 4.9. As indicated from the table, the model with the set of selected variables when only

Table 4.7. KTB data. Covariates.

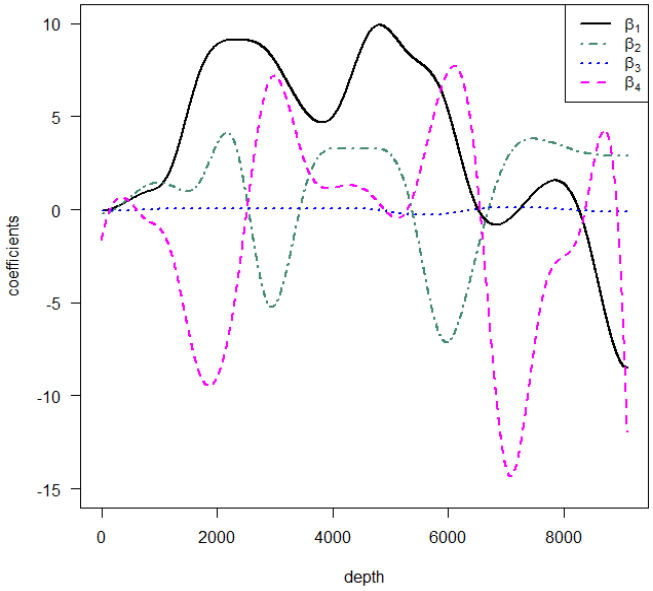
Label	description	Label	description
$X^{(1)}(T)$	crystal water (H_2O in weight %)	$X^{(12)}(T)$	rubidium (Rb, in ppm)
$X^{(2)}(T)$	carbon (C in weight %)	$X^{(13)}(T)$	zirkonium (Zr, in ppm)
$X^{(3)}(T)$	quartz (QRZ in weight %)	$X^{(14)}(T)$	chronium (Cr, in ppm)
$X^{(4)}(T)$	SiO_2 content (in weight %)	$X^{(15)}(T)$	vanadium (V, in ppm)
$X^{(5)}(T)$	Al_2O_3 content (in weight %)	$X^{(16)}(T)$	copper (Cu, in ppm)
$X^{(6)}(T)$	Fe_2O_3 content (in weight %)	$X^{(17)}(T)$	potassium (K, in weight %)
$X^{(7)}(T)$	MgO content (in weight %)	$X^{(18)}(T)$	thorium^{232} (Th, in ppm)
$X^{(8)}(T)$	CaO content (in weight %)	$X^{(19)}(T)$	uranium^{238} (U, in ppm)
$X^{(9)}(T)$	Na_2O content (in weight %)	$X^{(20)}(T)$	magnetic susceptibility (dimensionless)
$X^{(10)}(T)$	sulfur (S, in weight %)	$X^{(21)}(T)$	density (in $Mg \cdot m^{-3}$)
$X^{(11)}(T)$	strontium (Sr, in ppm)	$X^{(22)}(T)$	thermal conductivity (in $W \cdot m^{-1} \cdot K^{-1}$)

$X^{(3)}(T)$ is removed has the lowest SIC value. Similarly, for the variability function, the SIC values when removing the irrelevant variables using the NNG in this chapter, removing the irrelevant variables using the gALassoP, and using the full model are presented at the bottom of Table 4.9. The table indicates that the model with the set of selected variables resulting from the NNG has the lowest SIC. The estimated median log of the content of cataclastic rocks, using all the methods, is plotted in Figure 4.7. The content is increasing up to about 7000 meters with a small peak around 2000 meters and then slightly decreasing. All the methods respect the pattern of the evolution of the cataclastic rocks content over the depth which is also supported by the comparable SIC values in Table 4.8. For the variability function, we see that even though the NNG selects much less covariates, it has a smaller SIC value compared to that of the gALassoP and gALassoB.

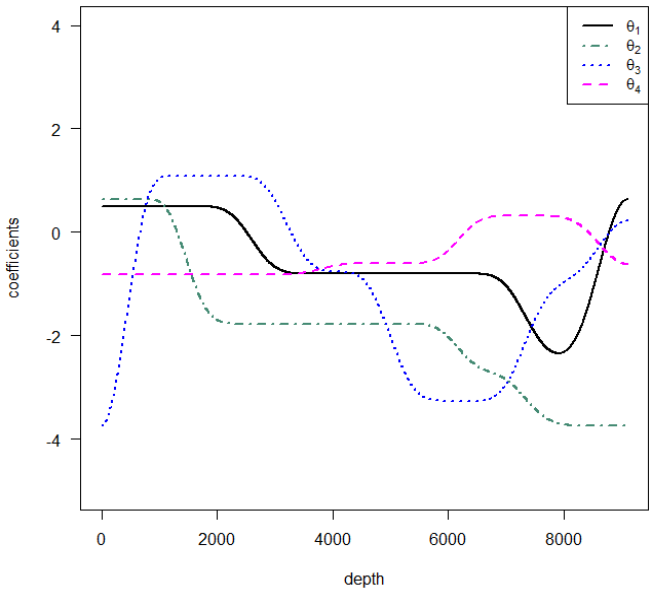
Table 4.8. KTB data. Selected covariates.

	Median					Variability				
	gAL_P	gAL_B	NNG	gS_B	gS_P	gAL_P	gAL_B	NNG	gS_B	gS_P
$X^{(1)}(T)$	x	x	x	x	x	x	x	x	x	x
$X^{(2)}(T)$	x	x	x	x	x	x	x	x	x	x
$X^{(3)}(T)$				x	x	x			x	x
$X^{(4)}(T)$	x	x	x	x	x	x	x	x	x	x
$X^{(5)}(T)$	x	x	x	x	x	x	x	x	x	x
$X^{(6)}(T)$	x	x	x	x	x	x	x	x	x	x
$X^{(7)}(T)$	x	x	x	x	x	x	x		x	x
$X^{(8)}(T)$	x	x	x	x	x	x	x	x	x	x
$X^{(9)}(T)$	x	x	x	x	x	x	x		x	x
$X^{(10)}(T)$	x	x	x	x	x	x	x		x	x
$X^{(11)}(T)$	x	x	x	x	x	x	x		x	x
$X^{(12)}(T)$	x	x	x	x	x	x	x	x	x	x
$X^{(13)}(T)$	x	x	x	x	x	x	x		x	x
$X^{(14)}(T)$	x	x	x	x	x		x	x	x	x
$X^{(15)}(T)$	x	x	x	x	x	x	x	x	x	x
$X^{(16)}(T)$	x	x	x	x	x		x		x	x
$X^{(17)}(T)$	x	x	x	x	x		x		x	x
$X^{(18)}(T)$	x	x	x	x	x	x	x		x	x
$X^{(19)}(T)$	x	x	x	x	x	x	x	x	x	x
$X^{(20)}(T)$	x	x	x	x	x	x	x		x	x
$X^{(21)}(T)$	x	x	x	x	x	x	x	x	x	x
$X^{(22)}(T)$	x	x	x	x	x	x	x	x	x	x
SIC	-1.31	-1.30	-1.44	-1.47	-1.30	-0.42	0.70	-0.73	-0.94	-0.95
NS	21	21	21	22	22	19	21	12	22	22

NOTE: gAL_P: gALassoP; gAL_B: gALassoB; gS_B: gSCADB; gS_P: gSCADP; SIC: SIC-criterion in Equations (4.2) or (4.5); NS: number of selected covariates.



(a) Median

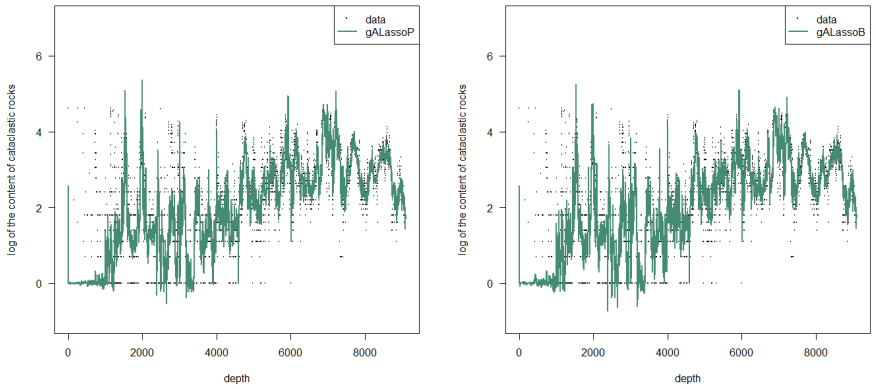


(b) Variability

Figure 4.6. KTB data. The estimated coefficient functions of the standardized covariates.

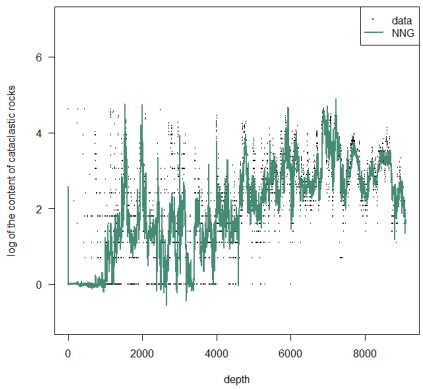
Table 4.9. KTB data. The SIC values based on the P-splines estimation of the selected covariates from the different methods.

Method	SIC
For the median function	
NNG	-1.294
NNG of Antoniadis et al. (2012b)	-1.289
NNG of Antoniadis et al. (2012b) including $X^{(3)}(T)$	-1.284
The full model	-1.285
For the variability function	
NNG	-0.451
gALassoP	-0.448
The full model	-0.444

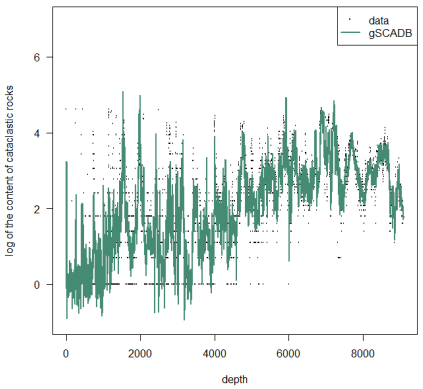


(a)

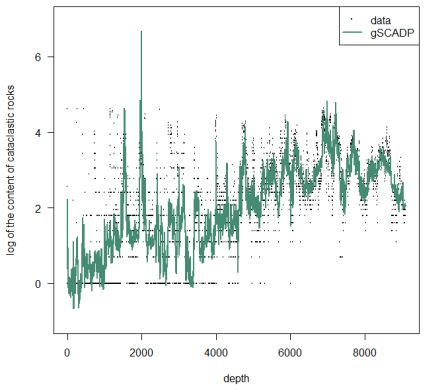
(b)



(c)



(d)



(e)

Figure 4.7. KTB data. The estimated median log of the content of cataclastic rocks.

4.3.3 Intego data

This data set is obtained from Intego, a Belgian general practice-based morbidity registration network at the Department of General Practice of KU Leuven. Ninety-seven general practitioners of 55 practices, evenly spread throughout Flanders, Belgium, collaborate in the Intego project. The data contains annual measurements of 6529 patients over the years 2000–2015. Each patient has different number of measurements. The response is eGFR (the estimated glomerular filtration rate) in mL/min/1.73 m^2/year. This response variable, which is calculated with the modification of diet in the renal disease equation, is used to measure the kidney function of the patients. A rapid annual decline of kidney function is defined as $\geq$ 3mL/min/1.73 m^2/year (Vaes et al., 2015). We have 27 covariates (see Table 4.10) consisting of gender, age, weight, height, blood pressure, lab tests, presence of morbidities and prescribed medications. The main aim of the project is to explore the association between blood pressure value and eGFR over time (year).

To analyze the data set, B-splines of degree three with eight quasi-uniform knots and differencing order one (as in Section 4.2) are used for the gALassoP, NNG and gSCADP. Where as for the gALassoB and gSCADB, the number of knots are chosen adaptively from a grid of between four and eight. The covariates are standardized as in Section 3.2.1.

We conduct the variable selection using all the techniques used in Section 4.2. For the median function, as indicated in Table 4.11, the NNG chooses only 13 covariates, while having the same SIC as the gALassoP and gALassoB. The other methods select over 20 covariates. Based on the selected set of variables by the NNG, the blood pressure values (systolic and diastolic) are not selected as important covariates. For the variability function, again the NNG selects only one covariate. Figure 4.8 shows the estimated functional coefficients of the first five covariates. The figure indicates that indeed $\hat{\beta}_5(T)$, the estimated functional coefficient of the diastolic blood pressure, is around zero. As shown in Figure 4.8 (a), age has a negative association with the kidney function (median eGFR), which is varying over time.

We plot the estimated median function of eGFR for the maximum values of the covariates, i.e. for female patients with the highest age, weight, height, blood pressure, lab tests, and presence of morbidities and prescribed medications. Figure 4.9 shows that the B-splines based methods have wiggly functional estimates. Based on the NNG, we see that the eGFR value is decreasing over time for these specific type of patients.

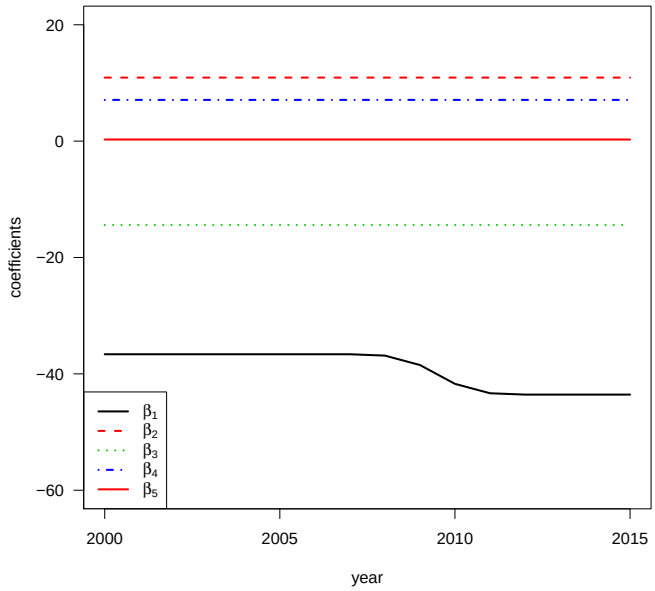
Table 4.10. Intego data. Covariates.

Label	description	Label	description
$X^{(1)}(T)$	age	$X^{(15)}(T)$	heart failure (1 = yes)
$X^{(2)}(T)$	weight	$X^{(16)}(T)$	hypercholesterolemia (1 = yes)
$X^{(3)}(T)$	height	$X^{(17)}(T)$	ischemic heart disease (1 = yes)
$X^{(4)}(T)$	systolic blood pressure	$X^{(18)}(T)$	stroke (1 = yes)
$X^{(5)}(T)$	diastolic blood pressure	$X^{(19)}(T)$	urinary disease (1 = yes)
$X^{(6)}(T)$	glucose lab test	$X^{(20)}(T)$	arterosclerosis (1 = yes)
$X^{(7)}(T)$	cholesterol lab test	$X^{(21)}(T)$	hypertension (1 = yes)
$X^{(8)}(T)$	ldl cholesterol lab test	$X^{(22)}(T)$	diuretic medication (1 = yes)
$X^{(9)}(T)$	hdl cholesterol lab test	$X^{(23)}(T)$	beta blockers medication (1 = yes)
$X^{(10)}(T)$	triglycerides lab test	$X^{(24)}(T)$	calcium blockers medication (1 = yes)
$X^{(11)}(T)$	hemoglobin lab test	$X^{(25)}(T)$	renin agents medication (1 = yes)
$X^{(12)}(T)$	uric acid lab test	$X^{(26)}(T)$	lipid agents medication (1 = yes)
$X^{(13)}(T)$	potassium lab test	$X^{(27)}(T)$	gender (1 = Female)
$X^{(14)}(T)$	diabetes (1 = yes)		

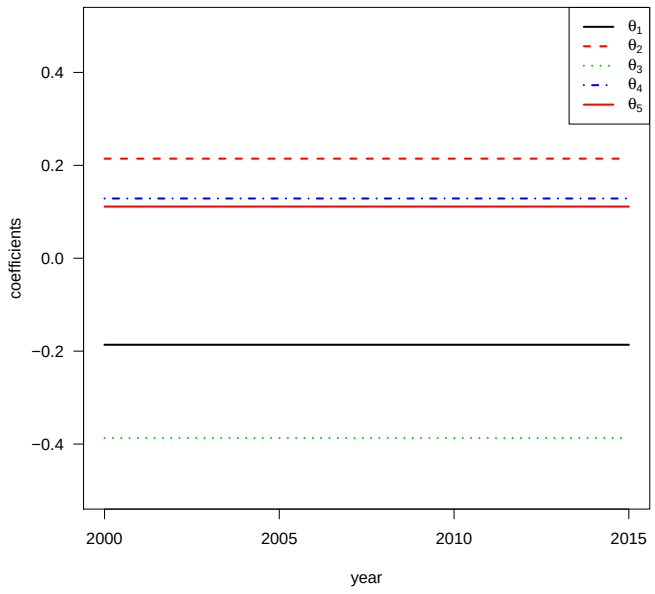
Table 4.11. Intego data. Selected covariates.

	Median					Variability				
	gAL_P	gAL_B	NNG	gS_B	gS_P	gAL_P	gAL_B	NNG	gS_B	gS_P
$X^{(1)}(T)$	x	x	x	x	x				x	x
$X^{(2)}(T)$	x	x	x	x	x				x	x
$X^{(3)}(T)$	x	x	x	x	x				x	x
$X^{(4)}(T)$	x	x		x	x				x	x
$X^{(5)}(T)$	x	x		x	x				x	x
$X^{(6)}(T)$	x	x		x	x				x	x
$X^{(7)}(T)$	x	x	x	x	x	x	x		x	x
$X^{(8)}(T)$	x	x		x	x	x	x		x	x
$X^{(9)}(T)$	x	x	x	x	x				x	x
$X^{(10)}(T)$	x	x	x	x	x	x	x		x	x
$X^{(11)}(T)$	x	x	x	x	x		x	x	x	x
$X^{(12)}(T)$	x	x	x	x	x				x	x
$X^{(13)}(T)$	x	x	x	x	x				x	x
$X^{(14)}(T)$	x	x		x	x				x	x
$X^{(15)}(T)$	x	x		x	x				x	x
$X^{(16)}(T)$										x
$X^{(17)}(T)$	x				x					x
$X^{(18)}(T)$	x	x		x	x				x	x
$X^{(19)}(T)$	x	x	x	x	x				x	x
$X^{(20)}(T)$		x		x	x				x	x
$X^{(21)}(T)$									x	x
$X^{(22)}(T)$	x	x	x	x	x				x	x
$X^{(23)}(T)$				x						x
$X^{(24)}(T)$				x	x				x	x
$X^{(25)}(T)$	x	x	x	x	x					x
$X^{(26)}(T)$				x						x
$X^{(27)}(T)$	x	x	x	x	x				x	x
SIC	1.84	1.84	1.84	-23.35	-27.34	-0.84	3.64	-0.86	-27.24	-31.05
NS	21	21	13	24	23	3	4	1	22	27

NOTE: gAL_P: gALassoP; gAL_B: gALassoB; gS_B: gSCADB; gS_P: gSCADP; SIC: SIC-criterion in Equations (4.2) or (4.5); NS: number of selected covariates.



(a) Median



(b) Variability

Figure 4.8. Intego data. The estimated coefficient functions of the standardized covariates.

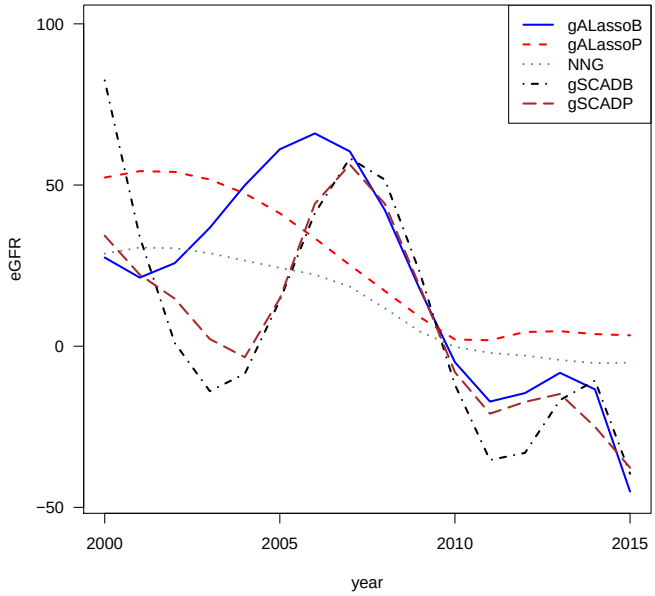


Figure 4.9. Intego data. The estimated median eGFR.

4.4 Conclusion

We propose two variable selection methods – grouped Adaptive Lasso using P-splines and NonNegative Garrote. We show the consistency of the grouped Adaptive Lasso with respect to variable selection and convergence rate. For a linear model, Zou (2006) has shown, using least squares, that the NNG estimator can be viewed as a special case of the Adaptive Lasso (ALasso) estimator. Based on this fact, the author has shown that the NNG has an equivalent asymptotic behavior as the ALasso. Here we can not proceed similarly, as we have grouped coefficients (i.e. replacing c_k by $\frac{\|\boldsymbol{\alpha}_k\|_1}{\|\hat{\boldsymbol{\alpha}}_k^{\text{init}}\|_1}$ in Equation (4.3) will not end up in Equation (4.1), with $\hat{\boldsymbol{\alpha}}_k^{\text{init}}$ an initial estimate obtained minimizing Equation (4.1) with $\lambda_L = 0$).

The simulation study in Section 4.2 shows that both the grouped Adaptive Lasso and NonNegative Garrote perform equivalently well, for the median function, compared to the gALassoB in Tang et al. (2013b) and the grouped SCAD (whether or not a P-spline smoothness penalty is added) in Noh et al. (2012). When a heteroscedastic model is considered in the simulation study, the two methods outperform in terms of

estimating the functional coefficients of the variability function. The gALassoB has the worst performance in estimating the functional coefficient of the relevant variable in the variability function. The two proposed methods also outperform the grouped SCAD in terms of selecting the relevant variable in the variability function. The NNG outperforms in terms of the computing time in all settings. The methods are illustrated on three data examples.

The model, we get in this chapter, can be further simplified by checking the constancy of the coefficients using the testing procedures proposed in Chapter 3 or using step 1 of Tang et al. (2012), leading to a partially linear varying coefficient model.

By using the AHe approach we are able to select variables not only for the median function but for the entire distribution, through investigating the variability function, with less computing time. In the simulation study, we see that our variable selection methods have identified properly the important variables in terms of both the signal and variability functions. Hence, both the gALassoP and NNG are recommended to do variable selection and estimation.

4.5 Assumptions, auxiliary results and proofs

4.5.1 Notations and Assumptions

We first introduce some notations:

$$\mathbf{Z} = (\mathbf{U}_{11}, \ldots, \mathbf{U}_{1N_1}, \ldots, \mathbf{U}_{n1}, \ldots, \mathbf{U}_{nN_n})^{\mathrm{T}}_{N\times 1}; \quad \psi_\tau(u) = \tau - I(u<0);$$

$$\mathbf{H}^2 = \mathbf{Z}^{\mathrm{T}}\mathbf{Z}; \quad \mathbf{z}_{ij} = \mathbf{H}^{+}\mathbf{U}_{ij};$$

$$\mathbf{Z}^{\mathrm{V}} = \left(\mathbf{U}^{\mathrm{V}}_{11}, \ldots, \mathbf{U}^{\mathrm{V}}_{1N_1}, \ldots, \mathbf{U}^{\mathrm{V}}_{n1}, \ldots, \mathbf{U}^{\mathrm{V}}_{nN_n}\right)^{\mathrm{T}}_{N\times 1} \quad (\mathbf{H}^{\mathrm{V}})^2 = (\mathbf{Z}^{\mathrm{V}})^{\mathrm{T}}\mathbf{Z}^{\mathrm{V}}$$

$$\hat{\boldsymbol{\theta}}^{*} = \mathbf{H}^{\mathrm{V}}\hat{\boldsymbol{\alpha}}^{\mathrm{V}} \quad \tilde{\boldsymbol{\theta}}^{*} = \mathbf{H}^{\mathrm{V}}\tilde{\boldsymbol{\alpha}}^{\mathrm{V}}$$

$$\hat{\boldsymbol{\alpha}} = ((\hat{\boldsymbol{\alpha}}_0)^{\mathrm{T}}, (\hat{\boldsymbol{\alpha}}_1)^{\mathrm{T}}, \ldots, (\hat{\boldsymbol{\alpha}}_p)^{\mathrm{T}})^{\mathrm{T}}_{m_{tot}\times 1}; \quad \boldsymbol{\theta}^{*} = \mathbf{H}^{\mathrm{V}}\boldsymbol{\alpha}^{\mathrm{V}};$$

$$\tilde{\boldsymbol{\alpha}} = ((\tilde{\boldsymbol{\alpha}}_0)^{\mathrm{T}}, (\tilde{\boldsymbol{\alpha}}_1)^{\mathrm{T}}, \ldots, (\tilde{\boldsymbol{\alpha}}_p)^{\mathrm{T}})^{\mathrm{T}}_{m_{tot}\times 1}; \quad \tilde{\boldsymbol{\beta}}^{*} = \mathbf{H}\tilde{\boldsymbol{\alpha}};$$

$$R_{ij} = \textstyle\sum_{k=0}^{p}(\tilde{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}_k(t_{ij}, \nu_k) - \beta_k(t_{ij}))X_{ij}^{(k)}; \quad \boldsymbol{\beta}^{*} = \mathbf{H}\boldsymbol{\alpha};$$

$$\hat{\boldsymbol{\beta}}^{*} = \mathbf{H}\hat{\boldsymbol{\alpha}};$$

where $\mathbf{H}^+$ denotes the Moore inverse of $\mathbf{H}$, and $\tilde{\boldsymbol{\alpha}}_k$ are the coefficients of the best possible spline approximation of $\beta_k(T)$ (defined in Lemma 4.1). Let us define $\boldsymbol{\chi}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iN_i})^{\mathrm{T}}$ as the $N_i \times (p+1)$ design matrix on the ith subject, and $\boldsymbol{\chi} = (\boldsymbol{\chi}_1^{\mathrm{T}}, \ldots, \boldsymbol{\chi}_n^{\mathrm{T}})^{\mathrm{T}}$.

Assumptions:

4.1. $\beta_k \in \mathcal{H}_{r_k}$, $k = 0, 1, \ldots, p$, for some $r_k > 3/2$. Let ς_k denote the Hölder constant for the kth varying coefficient function β_k (see Definition 2.2 in Section 2.6.1).

4.2. For all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, N_i\}$, the random design vector $\mathbf{X}_{ij}$ is bounded in probability. The eigenvalues of the matrix $N^{-1}\boldsymbol{\chi}^{\mathrm{T}}\boldsymbol{\chi}$ are bounded away from zero and infinity.

4.3. $\lim_{n\to\infty} u_{\max}^2 n^{\delta^M - 1} = 0$ for some $\delta^M \in]0, 1[$.

4.4. $\max_{i=1,\ldots,n} N_i < \infty$.

4.5. Let $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iN_i})^{\mathrm{T}}$. For each i, $0 < E[\psi_{0.5}(\boldsymbol{\epsilon}_i)\psi_{0.5}(\boldsymbol{\epsilon}_i)^{\mathrm{T}}] < \infty$.

4.6. $\theta_k \in \mathcal{H}_{r_k^{\mathrm{V}}}$, $k = 0, 1, \ldots, p$, for some $r_k^{\mathrm{V}} > 3/2$. Let ς_k^{V} denote the Hölder constant for the kth varying coefficient function θ_k.

4.7. Let $\boldsymbol{\epsilon}_i^{\mathrm{V}} = (\epsilon_{i1}^{\mathrm{V}}, \ldots, \epsilon_{iN_i}^{\mathrm{V}})^{\mathrm{T}}$. For each i, $0 < E[\psi_{0.5}(\boldsymbol{\epsilon}_i^{\mathrm{V}})\psi_{0.5}(\boldsymbol{\epsilon}_i^{\mathrm{V}})^{\mathrm{T}}] < \infty$.

4.5.2 Proof of Theorem 4.1

The proof of Theorem 4.1 (a) is similar to that of Tang et al. (2013b), the essential difference is that here we have to incorporate the smoothing penalty term. The proof in Theorem 4.1 (b) is similar to that in Chapter 2, the difference is that here we give a proof for the median function and we have in addition the regularization term.

First, we need the following lemmas. Lemma 4.1 (a) is taken directly from Corollary 6.21 of Schumaker (1981), Lemma 4.1 (b) immediately follows from Lemma 4.1 (a), Lemma 4.2 is a basic inequality, Lemmas 4.3 and 4.4 are taken, respectively, from proof of Theorem 2.1 and Lemma 3.4 in He and Shi (1994), the arguments used to prove Lemma 4.5 are similar to those used in the proof of Lemma 2.4, and Lemma 4.6 is derived from Lemma 2 of Tang et al. (2013b). Hence, we omit the proofs.

Lemma 4.1 *Suppose Assumptions* **2.2**, **2.6** *and* **4.1** *hold. For a constant $W_{4,k}$ that depends on only ν_k and c_k, there exists a spline coefficient vector $\tilde{\boldsymbol{\alpha}} = (\tilde{\boldsymbol{\alpha}}_0, \ldots, \tilde{\boldsymbol{\alpha}}_p)$*

such that,

(a) $sup_{t\epsilon\mathcal{T}}|\beta_k(t) - \tilde{\boldsymbol{\alpha}}_k^T\mathbf{B}_k(t,\nu_k)| \le W_{4,k}u_{\max}^{-r_{\min}}$, $k = 0,\ldots,p$.

(b) $\|\tilde{\boldsymbol{\alpha}}_k\|_1 \neq 0$ *if* $k \in \{0,\ldots,s\}$, $\tilde{\boldsymbol{\alpha}}_k = \mathbf{0}_{m_k}$ *if* $k = s+1,\ldots,p$; *with* $\mathbf{0}_{m_k}$ *a vector of* 0 *of length* m_k.

Lemma 4.2 *For a vector* $\mathbf{a}$ *with dimension* m, $\|\mathbf{a}\|_1 \le m^{1/2}\|\mathbf{a}\|_2$.

Lemma 4.3 *Let* $\hat{\boldsymbol{\alpha}}^B$ *be the minimizer of Equation* (2.6) *with* λ_P *equals zero. Suppose Assumptions* **2.2**, **2.3**, **2.5**, **2.6** *and* **4.1** – **4.5** *hold. Then* $\|\hat{\boldsymbol{\alpha}}^B - \tilde{\boldsymbol{\alpha}}\|_2 = O_p(n^{-1/2}u_{\max})$.

Lemma 4.4 *Under Assumption* **4.3**, *there exists a positive constant* $W_{5,k}$ *depending only on* b, B, ν_k *and* ς_k, *such that, except on an event whose probability tends to zero with* n, *for all* $\beta_k \in \mathcal{H}_{r_k}$ *and* $\hat{\beta}_k$ *minimizer of Equation* (4.1),

$$\|\hat{\beta}_k - \beta_k\|_2^2 \le W_{5,k}\left\{(u_{\max})^{-2r_{\min}} + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}(\hat{\beta}_k(t_{ij}) - \beta_k(t_{ij}))^2\right\}.$$

Lemma 4.5 *Suppose Assumptions* **2.3**, **2.5**, **2.6**, **4.1** *and* **4.2** *hold, and* $\boldsymbol{\xi} \in \mathbb{R}^{m_{tot}}$.

(a) *For any sequence* $\{L_n\}$ *satisfying* $1 \le L_n \le (u_{\max})^{\delta_0/10}$, *for some* $0 < \delta_0 < (r_{\min} - 1/2)/(2r_{\min}+1)$,

$$\sup_{\|\boldsymbol{\xi}\|_2\le 1}(u_{\max})^{-1}\left|\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\Big[\rho_{0.5}\Big\{\epsilon_{ij} - L_n(u_{\max})^{1/2}\mathbf{z}_{ij}^T\boldsymbol{\xi} - R_{ij}\Big\} - \rho_{0.5}\Big\{\epsilon_{ij} - R_{ij}\Big\}\right.$$
$$+L_n(u_{\max})^{1/2}\mathbf{z}_{ij}^T\boldsymbol{\xi}(0.5 - I(\epsilon_{ij} < 0)) - E_\epsilon\Big[\rho_{0.5}\Big\{\epsilon_{ij} - L_n(u_{\max})^{1/2}\mathbf{z}_{ij}^T\boldsymbol{\xi} - R_{ij}\Big\}$$
$$\left.-\rho_{0.5}\Big\{\epsilon_{ij} - R_{ij}\Big\}\Big]\Big]\right| = o_P(1),$$

where E_ϵ *is the conditional expectation given* $(\mathbf{X}_{ij}, t_{ij})$ *for* $i = 1,\ldots,n; j = 1,\ldots,N_i$.

(b) *For any* $\omega > 0$, *there exists* $L := L_\omega$ *(sufficiently large) such that as* $n \to \infty$,

$$P\left\{(u_{\max})^{-1}\left(\inf_{\|\boldsymbol{\xi}\|_2=1}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}E_\epsilon\Big[\rho_{0.5}\Big\{\epsilon_{ij} - L(u_{\max})^{1/2}\mathbf{z}_{ij}^T\boldsymbol{\xi} - R_{ij}\Big\} - \rho_{0.5}\Big\{\epsilon_{ij} - R_{ij}\Big\}\Big]\right.\right.$$
$$\left.\left. - L(u_{\max})^{1/2}\left\|\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbf{z}_{ij}(0.5 - I(\epsilon_{ij} < 0))\right\|_2\right) > 1\right\} > 1 - \omega.$$

Lemma 4.6 *Suppose Assumptions* **2.5**, **2.6**, **4.1**, **4.2** *and* **4.4** *hold. Then, the eigenvalues of* $u_{\max}\mathbf{H}^2/n$ *are uniformly bounded away from* 0 *and* ∞ *in probability.*

Now, we need to prove the following lemma, using similar lines as that of Lemma 2.5.

Lemma 4.7 *Suppose Assumptions* **2.3**, **2.5**, **2.6**, **4.1**, **4.2** *and* **4.4** *hold. Furthermore, assume that* $(u_{\max})^{3/4}\lambda_{P\max}n^{-1/2} \to 0$ *and* $(u_{\max})^{1/2}\lambda_L n^{-1/2} \to 0$ *as* n *tends to* ∞. *Then,* $\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} (\hat{\beta}_k(t_{ij}) - \tilde{\beta}_k(t_{ij}))^2 = O_P(u_{\max})$.

Proof 4.1 (Proof of Lemma 4.7) *Using Lemma 4.5, for any* $\omega > 0$, *there exists* $L := L_\omega$ *such that as* $n \to \infty$,

$$P\left\{(u_{\max})^{-1}\left(\inf_{\|\boldsymbol{\xi}\|_2 = L(u_{\max})^{1/2}} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left\{\epsilon_{ij} - \mathbf{z}_{ij}^T\boldsymbol{\xi} - R_{ij}\right\} - \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left\{\epsilon_{ij} - R_{ij}\right\}\right) > 1\right\} > 1 - \omega. \tag{4.9}$$

We have:

$$\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left\{\epsilon_{ij} - \mathbf{z}_{ij}^T(\hat{\boldsymbol{\beta}}^* - \tilde{\boldsymbol{\beta}}^*) - R_{ij}\right\} = \inf_{\|\boldsymbol{\xi}\|_2 \in \mathbb{R}^{m_{tot}}} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left\{\epsilon_{ij} - \mathbf{z}_{ij}^T\boldsymbol{\xi} - R_{ij}\right\}. \tag{4.10}$$

From the proof of Lemma 2.3 in Andriyana et al. (2014), assuming $\|\boldsymbol{\beta}^* - \tilde{\boldsymbol{\beta}}^*\|_2 = L(u_{\max})^{1/2}$ *and by Lemma 4.6, we have:*

$$\begin{aligned}\lambda_P \sum_{k=0}^{p} \omega_{1k}(\|\mathbf{D}_k^{d_k}\boldsymbol{\alpha}_k\|_1 - \|\mathbf{D}_k^{d_k}\tilde{\boldsymbol{\alpha}}_k\|_1) &\le \sum_{k=1}^{p} 4^{d_k/2} m_{\max}^{3/4} \lambda_{P\max} \|\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}\|_2 \\ &= o_P(u_{\max}),\end{aligned} \tag{4.11}$$

where $m_{\max} = \max(m_0, \ldots, m_p)$.

By Lemma 4.2 and the triangle inequality

$$\lambda_L \sum_{k=0}^{p} \omega_{2k}(\|\boldsymbol{\alpha}_k\|_1 - \|\tilde{\boldsymbol{\alpha}}_k\|_1) \le \lambda_L \sum_{k=0}^{p} \omega_{2k}(\|\boldsymbol{\alpha}_k - \tilde{\boldsymbol{\alpha}}_k\|_1) \le \lambda_L \sum_{k=0}^{p} \omega_{2k} m_k^{1/2}(\|\boldsymbol{\alpha}_k - \tilde{\boldsymbol{\alpha}}_k\|_2).$$

By Lemmas 4.1 (b) and 4.3, for $k = 0, \ldots, s$ *and* n *large enough, there exist a positive constant* $c_L \ge 0$ *such that* $\|\hat{\boldsymbol{\alpha}}_k^B\|_1 \ge c_L$ *in probability and therefore* $\omega_{2k} \le c_L^{-\eta_L}$. *Assuming* $\|\boldsymbol{\beta}^* - \tilde{\boldsymbol{\beta}}^*\|_2 = L(u_{\max})^{1/2}$ *and by Lemma 4.6,*

$$\lambda_L \sum_{k=0}^{p} \omega_{2k}(\|\boldsymbol{\alpha}_k\|_1 - \|\tilde{\boldsymbol{\alpha}}_k\|_1) \le p c_L^{-\eta_l} \lambda_L u_{\max}^{3/2} n^{-1/2}.$$

Now, assuming $\lambda_L u_{\max}^{1/2} n^{-1/2} \longrightarrow 0$ *and by the fact that* p *is finite, we have*

$$\lambda_L \sum_{k=0}^{p} \omega_{2k} (\|\boldsymbol{\alpha}_k\|_1 - \|\tilde{\boldsymbol{\alpha}}_k\|_1) = o_P(u_{\max}). \tag{4.12}$$

By Equations (4.9), (4.11) and (4.12),

$$P\left(\inf_{\|\boldsymbol{\beta}^* - \tilde{\boldsymbol{\beta}}^*\|_2 = L(u_{\max}^V)^{1/2}} S(\boldsymbol{\alpha}) - S(\tilde{\boldsymbol{\alpha}}) > 0 \right) \to 1. \tag{4.13}$$

By convexity of $S(\boldsymbol{\alpha})$*, Equation* (4.10) *and since*

$$\sum_{k=0}^{p} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(\hat{\beta}_k(t_{ij}) X_{ij}^{(k)} - \tilde{\beta}_k(t_{ij}) X_{ij}^{(k)} \right)^2 = \|\hat{\boldsymbol{\beta}}^* - \tilde{\boldsymbol{\beta}}^*\|_2^2,$$

Equation (4.13) *implies that*

$$\begin{aligned} P\left(\sum_{k=0}^{p} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(\hat{\beta}_k(t_{ij}) X_{ij}^{(k)} - \tilde{\beta}_k(t_{ij}) X_{ij}^{(k)} \right)^2 \leq L^2 u_{\max} \right) \\ = P(\|\hat{\boldsymbol{\beta}}^* - \tilde{\boldsymbol{\beta}}^*\|_2 \leq L(u_{\max})^{1/2}) > 1 - \omega, \end{aligned}$$

which proves the lemma.

Proof 4.2 (Proof of Theorem 4.1 (b)) *First, by the fact that* $2ab \leq a^2 + b^2$ *for any* $(a, b) \in \mathbb{R}^2$*, we have*

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\hat{\beta}_k(t_{ij}) - \beta_k(t_{ij}) \right)^2 \\ &\qquad \leq \frac{2}{n} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\hat{\beta}_k(t_{ij}) - \tilde{\beta}_k(t_{ij}) \right)^2 + 2 \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\tilde{\beta}_k(t_{ij}) - \beta_k(t_{ij}) \right)^2. \end{aligned}$$

By Lemma 4.7, we have that

$$\frac{2}{n} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\hat{\beta}_k(t_{ij}) - \tilde{\beta}_k(t_{ij}) \right)^2 = \frac{2}{n} O_P(u_{\max}) = O_P\left(u_{\max} n^{-1} \right). \tag{4.14}$$

By Lemma 4.1 (a), we have that

$$2\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\tilde{\beta}_k(t_{ij})-\beta_k(t_{ij})\right)^2 \le 2\left(\max_{k=0,\ldots,p}(W_{4,k})\right)^2 (u_{\max})^{-2r_{\min}} = O_P\left((u_{\max})^{-2r_{\min}}\right). \tag{4.15}$$

Combining Equations (4.14) *and* (4.15) *leads to*

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{\beta}_k(t_{ij})-\beta_k(t_{ij})\right)^2 &= O_P\left(u_{\max}n^{-1}+(u_{\max})^{-2r_{\min}}\right)\\ &= O_P\left(n^{-2r_{\min}/(2r_{\min}+1)}\right).\end{aligned} \tag{4.16}$$

Using Equation (4.16) *and Lemma 4.4, the proof of the theorem is now complete.*

Proof 4.3 (Proof of Theorem 4.1 (a)) *From the proof of Theorem 1 (a) of Tang et al. (2013b), it is sufficient to prove that*

$$\left\|\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\psi_{0.5}(Y_{ij}-\mathbf{U}_{ij}^T\hat{\boldsymbol{\alpha}})\mathbf{U}_{ij}^{(k)}\right\|_2 \le \lambda_P\omega_{1k}+\lambda_L\omega_{2k},\ \text{for } k=s+1,\ldots,p.$$

We now check the above inequality holds:

From Tang et al. (2013b), we have that $\|\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\psi_{0.5}(Y_{ij}-\mathbf{U}_{ij}^T\hat{\boldsymbol{\alpha}})\mathbf{U}_{ij}^{(k)}\|_2 = O_P(n^{1/2})$. *By Lemmas 4.1 (b) and 4.3,* $\|\hat{\boldsymbol{\alpha}}_k^B\|_2 = O_P(n^{-1/2}u_{\max})$ *for* $k=s+1,\ldots,p$, *therefore by Lemma 4.2 there exist, for large* n, *a positive constant* e_k *such that* $\|\hat{\boldsymbol{\alpha}}_k^B\|_1 \le u_k^{1/2}\|\hat{\boldsymbol{\alpha}}_k^B\|_2 \le e_k n^{-1/2}u_{\max}^{3/2}$. *Hence, for* $\eta_L \ge 1$, $\lambda_L\omega_{2k} = \lambda_L\|\hat{\boldsymbol{\alpha}}_k^B\|_1^{-\eta_L} \ge \lambda_L\|\hat{\boldsymbol{\alpha}}_k^B\|_1^{-1} \ge \lambda_L e_k^{-1}n^{1/2}u_{\max}^{-3/2}$.

As $Range(\hat{\beta}_k^B(T)) \ge 0$, $\lambda_P\omega_{1k} \ge 0$. *Hence, assuming* $\lambda_L u_{\max}^{-3/2} \longrightarrow \infty$, *the proof is complete.*

4.5.3 Proof of Theorem 4.2

First, we need the following lemmas. Lemma 4.8 (a) is taken directly from Corollary 6.21 of Schumaker (1981), Lemma 4.8 (b) follows from Lemma 4.8 (a), the arguments used to prove Lemma 4.9 are similar to those used in the proof of Theorem 2.1 in He and Shi (1994), Lemma 4.10 is derived from Lemma 3.4 of He and Shi (1994) and Lemma 4.11 can be proved as in Lemma 2 of Tang et al. (2013b). Hence, we omit the proofs.

Lemma 4.8 *Suppose Assumptions* **2.6**, **2.12** *and* **4.6** *hold. For a constant* $W_{1,k}^V$ *that depends only on* ν_k^V *and* ς_k^V, *there exists a spline coefficient vector* $\tilde{\boldsymbol{\alpha}}^V$ *such that,*

(a) $sup_{t\in\mathcal{T}}|\theta_k(t) - (\tilde{\boldsymbol{\alpha}}_k^V)^T\mathbf{B}_k^V(t,\nu_k^V)| \le W_{1,k}^V(u_{\max}^V)^{-r_{\min}^V}$, $k = 0,\ldots,p$.

(b) $\|\tilde{\boldsymbol{\alpha}}_k^V\|_1 \neq 0$ *if* $k \in \{0,\ldots,s^V\}$, $\tilde{\boldsymbol{\alpha}}_k^V = \mathbf{0}_{m_k^V}$ *if* $k = s^V+1,\ldots,p$.

Lemma 4.9 *Let* $(\hat{\boldsymbol{\alpha}}^V)^B$ *be the minimizer of Equation* (4.7) *with* $\lambda_P^V = \lambda_L^V = 0$. *Suppose Assumptions* **2.6**, **2.9** – **2.12**, **4.2**, **4.4**, **4.6** *and* **4.7** *hold. Then* $\|(\hat{\boldsymbol{\alpha}}^V)^B - \tilde{\boldsymbol{\alpha}}^V\|_2 = O_p(n^{-1/2}u_{\max}^V)$.

Lemma 4.10 *Under Assumption* **2.10**, *there exists a positive constant* $W_{2,k}^V$ *depending only on* b, B, ν_k^V *and* ς_k^V, *such that, except on an event whose probability tends to zero with* n, *for all* $\hat{\theta}_k$ *minimizer of Equation* (4.7) *and* $\theta_k \in \mathcal{H}_{r_k^V}$,

$$\|\hat{\theta}_k - \theta_k\|_2^2 \le W_{2,k}^V\left\{(u_{\max}^V)^{-2r_{\min}^V} + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}(\hat{\theta}_k(t_{ij}) - \theta_k(t_{ij}))^2\right\}.$$

Lemma 4.11 *Suppose Assumptions* **2.6**, **2.11**, **4.2**, **4.4** *and* **4.6** *hold. Then the eigenvalues of* $u_{\max}^V(\mathbf{H}^V)^2/n$ *are uniformly bounded away from* 0 *and* ∞ *in probability.*

Now, we need to prove the following lemma, which is a preliminary result for our main theorem.

Lemma 4.12 *Suppose Assumptions* **2.6**, **2.9**, **2.11**, **4.2**, **4.4** *and* **4.6** *hold. Furthermore, assume that* $(u_{\max}^V)^{3/4}\lambda_{P\max}^V n^{-1/2} \to 0$ *and* $(u_{\max}^V)^{1/2}\lambda_L^V n^{-1/2} \to 0$; *as* n *tends to* ∞. *Then,* $\sum_{i=1}^n \frac{1}{N_i}\sum_{j=1}^{N_i}(\hat{\theta}_k(t_{ij}) - \tilde{\theta}_k(t_{ij}))^2 = O_P(u_{\max}^V)$.

Proof 4.4 (Proof of Lemma 4.12) *From the proof of Lemma 2.5, Equation (2.13), assuming* $n^{-2r_{\min}^V/(2r_{\min}^V+1)} \to 0$ *and for any* $\omega > 0$, *there exists* $L := L_\omega$ *such that as* $n \to \infty$,

$$P\Big(\inf_{\|\boldsymbol{\theta}^*-\tilde{\boldsymbol{\theta}}^*\|_2 = L(u_{\max}^V)^{1/2}} Q(\boldsymbol{\alpha}^V) > u_{\max}^V\Big) > 1-\omega, \tag{4.17}$$

with

$$Q(\boldsymbol{\alpha}^V) = \sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\left\{\rho_{0.5}\left(\ln|Y_{ij} - \mathbf{X}_{ij}^T\hat{\boldsymbol{\beta}}(t_{ij})| - \sum_{k=0}^{p}\sum_{l=1}^{m}\alpha_{kl}^V B_{kl}^V(t_{ij},\nu_k^V)X_{ij}^{(k)}\right)\right.$$
$$\left. -\rho_{0.5}\left(\ln|Y_{ij} - \mathbf{X}_{ij}^T\hat{\boldsymbol{\beta}}(t_{ij})| - \sum_{k=0}^{p}\sum_{l=1}^{m}\tilde{\alpha}_{kl}^V B_{kl}^V(t_{ij},\nu_k^V)X_{ij}^{(k)}\right)\right\}.$$

From the proof of Lemma 2.3 in Andriyana et al. (2014), assuming $\|\hat{\boldsymbol{\theta}}^* - \tilde{\boldsymbol{\theta}}^*\|_{L_2} =$

$L(u_{\max}^V)^{1/2}$ *and by Lemma 4.11, we have:*

$$\lambda_P^V \sum_{k=0}^{p} \omega_{1k}^V (\|\mathbf{D}_k^{d_k^V} \hat{\boldsymbol{\alpha}}_k^V\|_1 - \|\mathbf{D}_k^{d_k^V} \tilde{\boldsymbol{\alpha}}_k^V\|_1) = o_P(u_{\max}^V). \tag{4.18}$$

Similar as in the proof of Lemma 4.7, by Lemmas 4.2, 4.8 (b), 4.9 and 4.11 and assuming $\|\hat{\boldsymbol{\theta}}^* - \tilde{\boldsymbol{\theta}}^*\|_2 = L(u_{\max}^V)^{1/2}$,

$$\lambda_L^V \sum_{k=0}^{p} \omega_{2k}^V (\|\hat{\boldsymbol{\alpha}}_k^V\|_1 - \|\tilde{\boldsymbol{\alpha}}_k^V\|_1) = o_P(u_{\max}^V). \tag{4.19}$$

By Equations (4.17), (4.18) and (4.19),

$$P\left(\inf_{\|\boldsymbol{\theta}^* - \tilde{\boldsymbol{\theta}}^*\|_{L_2} = L(u_{\max}^V)^{1/2}} S(\boldsymbol{\alpha}^V) - S(\tilde{\boldsymbol{\alpha}}^V) > 0 \right) \to 1. \tag{4.20}$$

By convexity of $S(\boldsymbol{\alpha}^V)$*, the fact that* $S(\hat{\boldsymbol{\alpha}}^V) - S(\boldsymbol{\alpha}^V) \leq 0$ *and since*

$$\sum_{k=0}^{p} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(\hat{\theta}_k(t_{ij}) X_{ij}^{(k)} - \tilde{\theta}_k(t_{ij}) X_{ij}^{(k)} \right)^2 = \|\hat{\boldsymbol{\theta}}^* - \tilde{\boldsymbol{\theta}}^*\|_2^2,$$

Equation (4.20) *implies that*

$$\begin{aligned} P\left(\sum_{k=0}^{p} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left(\hat{\theta}_k(t_{ij}) X_{ij}^{(k)} - \tilde{\theta}_k(t_{ij}) X_{ij}^{(k)} \right)^2 \leq L^2 u_{\max}^V \right) \\ = P(\|\hat{\boldsymbol{\theta}}^* - \tilde{\boldsymbol{\theta}}^*\|_2 \leq L(u_{\max}^V)^{1/2}) > 1 - \omega, \end{aligned}$$

which proves the lemma.

Proof 4.5 (Proof of Theorem 4.2 (b)) *Using similar lines as the proof of Theorem 4.1 (b) and using Lemmas 4.8 (a) and 4.12, we have*

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\hat{\theta}_k(t_{ij}) - \theta_k(t_{ij}) \right)^2 &= O_P\left(u_{\max}^V n^{-1} + (u_{\max}^V)^{-2r_{\min}^V} \right) \\ &= O_P\left(n^{-2r_{\min}^V/(2r_{\min}^V+1)} \right). \end{aligned} \tag{4.21}$$

Using Equation (4.21) *and Lemma 4.10, the proof of the theorem is complete.*

Proof 4.6 (Proof of Theorem 4.2 (a)) *From the proof of Theorem 1 (a) of Tang*

et al. (2013b), for $k = s^V + 1, \ldots, p$, it is sufficient to prove that

$$\left\| \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \psi_{0.5}(\ln|Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}} \hat{\boldsymbol{\beta}}(T)| - \left(\mathbf{U}_{ij}^{V}\right)^{T} \hat{\boldsymbol{\alpha}}^{V}) \mathbf{U}^{V(k)}_{ij} \right\|_2 \leq \lambda_P^V \omega_{1k}^V + \lambda_L^V \omega_{2k}^V.$$

We have from Tang et al. (2013b) that

$$\| \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \psi_{0.5}(\ln|Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}} \hat{\boldsymbol{\beta}}(T)| - \left(\mathbf{U}_{ij}^{V}\right)^{T} \hat{\boldsymbol{\alpha}}^{V}) \mathbf{U}^{V(k)}_{ij} \|_2 = O_P(n^{1/2}).$$

By Lemmas 4.8 (b) and 4.9, $\|(\hat{\boldsymbol{\alpha}}^V)_k^B\|_2 = O_P(n^{-1/2} u_{\max}^V)$ for $k = s^V + 1, \ldots, p$, therefore by Lemma 4.2 there exist a positive constant e_k^V such that $\|(\hat{\boldsymbol{\alpha}}^V)_k^B\|_1 \leq (u_k^V)^{1/2} \|(\hat{\boldsymbol{\alpha}}^V)_k^B\|_2 \leq e_k^V n^{-1/2} (u_{\max}^V)^{3/2}$. Hence, for $\eta_L^V \geq 1$, $\lambda_L^V \omega_{2k}^V = \lambda_L^V \|(\hat{\boldsymbol{\alpha}}^V)_k^B\|_1^{-\eta_L^V} \geq \lambda_L^V \|(\hat{\boldsymbol{\alpha}}^V)_k^B\|_1^{-1} \geq \lambda_L^V (e_k^V)^{-1} n^{1/2} (u_{\max}^V)^{-3/2}$.

As Range$((\hat{\beta}_k^V)^B(T)) \geq 0$, $\lambda_P^V \omega_{2k}^V \geq 0$. Hence, assuming $\lambda_L^V (u_{\max}^V)^{-3/2} \longrightarrow \infty$, the proof is complete.

Chapter 5

A two-stage screening and variable selection method in ultrahigh-dimensional quantile varying coefficient models with heteroscedastic error

As in Chapters 3 and 4, we consider the location-scale varying coefficient model in Equation (1.7) with the variability function of the form given in Equation (3.1).

For a given quantile level τ $(0 < \tau < 1)$, we define the set of active/relevant variables in the model in Equation (1.7) as

$$M_* = \{k : q_\tau\,(Y(T)|\mathbf{X}(T), T) \text{ functionally depends on } X^{(k)}(T)\},$$

for $k = 1, \ldots, p$. We denote $s = |M_*|$ the number of relevant variables in the model.

The previous chapter and the literature there in focus on situation where the number of covariates p is finite or p is growing slower than the sample size. Recently, there is an increased interest in ultrahigh-dimensional data with p much larger than the sample size. This type of problem arises, for instance, in studying tumor identification, finance, etc. Existing penalization methods for ultrahigh-dimensional parametric mean regression (least squares) include the Lasso in Meinshausen and Bühlmann (2006), Boosting in Bühlmann (2006), the Dantzig selector in Candes and Tao (2007), Sure Independence Screening (SIS) in Fan and Lv (2008) and forward selection in Wang (2009). Several authors consider ultrahigh-dimensional nonparametric mean regression (Huang et al., 2010; Wei et al., 2011; Xue and Qu, 2012; Cheng et al., 2016; Chen

et al., 2017, to cite a few).

A lot of work has also been done for ultrahigh-dimensional linear quantile regression models, including but not limited to the Lasso (Belloni and Chernozhukov, 2011), the Adaptive Lasso (Zheng et al., 2013), nonconvex penalties (Wang et al., 2012), composite quasi-likelihood (Bradic et al., 2011) and the Dantzig-type penalty (Park et al., 2017). In an ultrahigh-dimensional nonparametric quantile regression setting, the Adaptive Lasso (Honda et al., 2017) and variable screening (He et al., 2013) have been considered, among others. Tang et al. (2013a) propose a two-stage approach for quantile varying coefficient models.

This chapter has novel contributions to the existing literature. We propose a two-stage approach, that requires less computing time, for dimension reduction and variable selection, in ultrahigh-dimensional varying coefficient models. In addition to the location, we study the influence of covariates on the the scale (variability function) of the response. Further, a P-splines penalty is added to the objective function in order to avoid the problem of overfitting by B-splines estimation. In the first stage, we propose the Sure Independence Screening (SIS) to reduce the size of the model from ultrahigh-dimension to a model that has size close to the true model, which contains the true model as a submodel. In the second stage, variable selection is done by the NonNegative Garrote (NNG) on the reduced set of variables, in order to exclude the remaining irrelevant covariates. We compare the proposed two-stage method with the existing two-stage method of Tang et al. (2013a).

The remainder of the chapter is organized as follows. Sections 5.1 to 5.3 present the screening and variable selection methods considered in the thorough numerical comparison. Simulation studies are carried out in Section 5.4. The estimation methods are applied to a data example in Section 5.5. Finally, we discuss the main findings in Section 5.6. The R-code to implement the methods of this chapter on the data example is deferred to Appendix A and is available at `http://ibiostat.be/online-resources/online-resources/longitudinal`.

5.1 Methods

Assuming **H1**, the conditional median function of $Y(T)$ is given in Equation (2.2).

From the model in Equation (1.7) and since $V(\mathbf{X}, T) \geq 0$, we have Equation (2.7). Therefore, from Equation (3.1) and assuming **H2** in Chapter 2, $V(\cdot, \cdot)$ can be esti-

mated based on Equation (4.6).

Our variable selection and screening method is a two-stage procedure. We compare the proposed procedures with the two-stage method of Tang et al. (2013a). The method of Tang et al. (2013a) is based on B-splines estimation, with the grouped Lasso in the first (dimension reduction) stage and the grouped Adaptive Lasso in the second (variable selection) stage. We compare the grouped Lasso first stage, with the Sure Independence Screening first stage. In the second stage, we compare the grouped Adaptive Lasso based on B-splines (gALassoB), the grouped Adaptive Lasso using P-splines (gALassoP) and the NonNegative Garrote. Here at the second stage, we expect the gALassoP and NNG to perform better than the gALassoB, as in the previous chapter. Further, the NNG is expected to outperform in computing time.

In the subsequent subsections, we describe the screening and variable selection technique only for the median function. When the interest is in the variability function, we need to replace $(Y_{ij}, u_k, \nu_k, m_k)$ by $(\ln|Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\beta}}(T)|, u_k^{\mathrm{V}}, \nu_k^{\mathrm{V}}, m_k^{\mathrm{V}})$ for $k = 0, \ldots, p$, where $\hat{\boldsymbol{\beta}}(T)$ are the estimated functional coefficients from the median.

All the optimization problems are implemented in R-software and solved with the Frisch-Newton interior point algorithm (Portnoy and Koenker, 1997) of Tang et al. (2013a).

5.2 Dimension reduction – Stage 1

In this section, we describe the two first-stage methods considered in the numerical comparison. In the first stage, the dimension of the set of covariates is reduced to a moderate size, ideally including the true set of active/relevant covariates.

5.2.1 Grouped Lasso

In this method, regularization with a grouped Lasso penalty on the spline coefficients is done via a penalty on the L_1-norm of the $\boldsymbol{\alpha}_k$ ($\|\boldsymbol{\alpha}_k\|_1$), inducing sparsity as in Tang et al. (2013a). The resulting estimator is

$$\tilde{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} = \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left(Y_{ij} - \mathbf{U}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}\right) + \lambda_L \sum_{k=0}^{p} \|\boldsymbol{\alpha}_k\|_1,$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_p^{\mathrm{T}})^{\mathrm{T}}$; $\tilde{\boldsymbol{\alpha}} = (\tilde{\boldsymbol{\alpha}}_0^{\mathrm{T}}, \ldots, \tilde{\boldsymbol{\alpha}}_p^{\mathrm{T}})^{\mathrm{T}}$; λ_L is a nonnegative regularization parameter that determines the sparsity of the solution. We denote the estimated set of active variables as $\hat{M}_{\mathrm{Lasso}} = \{k : \|\tilde{\boldsymbol{\alpha}}_k\|_1 > 0, k = 1, \ldots, p\}$.

From Theorem 1 of Tang et al. (2013a), the estimator in this stage is sparse with the number of nonzero groups at the order close to the true size, and is consistent in estimation at the nearly optimal nonparametric rate. They show, however, that the estimator in this stage is not consistent in variable selection.

The smoothing parameter $u = u_0 = \ldots = u_k$ and the regularization parameter λ_L are chosen in a data-driven way, by minimizing the following SIC-criterion (Schwarz, 1978; Tang et al., 2013b; Gijbels et al., 2018) on a two-dimensional grid:

$$\mathrm{SIC}(u, \lambda_L) = \ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(Y_{ij} - \mathbf{U}_{ij}^{\mathrm{T}}\tilde{\boldsymbol{\alpha}}\right)\right) + \frac{\ln(N)}{2N}p_{u,\lambda_L}, \tag{5.1}$$

where $N = \sum_{i=1}^{n} N_i$; and p_{u,λ_L} is the effective degrees of freedom, taken to equal the size of the elbow set $\mathcal{E}_\lambda$ in Equation (2.12). The degree of the B-splines is fixed.

5.2.2 Sure Independence Screening

As an alternative to the grouped Lasso, we propose the Sure Independence Screening as a first dimension reduction, in this section. The SIS allows to rapidly reduce the dimension of the covariate space (p) to a moderate scale (He et al., 2013). Further, it has an advantage in computational time over the grouped Lasso. We estimate the marginal contribution of each covariate (while including the intercept in the model) individually with P-splines,

$$\begin{aligned}(\tilde{\boldsymbol{\alpha}}_0^{\mathrm{T}}, \tilde{\boldsymbol{\alpha}}_k^{\mathrm{T}}) = \arg\min_{(\boldsymbol{\alpha}_0^{\mathrm{T}}, \boldsymbol{\alpha}_k^{\mathrm{T}})} \sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\Bigg(Y_{ij} - \sum_{l\in\{0,k\}} X_{ij}^{(l)}\boldsymbol{\alpha}_l^{\mathrm{T}}\mathbf{B}_l(t_{ij}, \nu_l)\Bigg) \\ + \lambda_P \sum_{l\in\{0,k\}} \omega_{1l}\|\mathbf{D}_{m_l}^{d_l}\boldsymbol{\alpha}_l\|_1,\end{aligned} \tag{5.2}$$

where $\omega_{1l} = \mathrm{Range}(\tilde{\beta}_l^B(T))^{-\eta_P}$ with $\mathrm{Range}(\tilde{\beta}_l^B(T))$ is the range of quantile regression estimators obtained using unpenalized B-splines (i.e. the solution of Equation (5.2) with $\lambda_P = 0$), $\eta_P \geq 0$; $\lambda_P > 0$ is the smoothing parameter; and d_l is the differencing order in the penalty term (which is set to 1 as in the previous chapters). We take a fixed number of knots (eight knots are used in the simulation study as in Section 4.2), but choose the smoothing parameter λ_P data-driven by minimizing the SIC in

Equation (5.1).

Defining the marginal quantile utility as $\tilde{f}_{nk}(t_{ij}) = \tilde{\boldsymbol{\alpha}}_0^{\mathrm{T}}\mathbf{B}(t_{ij},\nu) + X_{ij}^{(k)}\tilde{\boldsymbol{\alpha}}_k^{\mathrm{T}}\mathbf{B}(t_{ij},\nu) - \hat{q}_{0.5}(Y_{ij})$, the SIS is based on the magnitude of the estimated marginal component $\hat{f}_{nk}^m = n^{-1}\sum_{i=1}^n N_i^{-1}\sum_{j=1}^{N_i}\tilde{f}_{nk}(t_{ij})^2$, where $\hat{q}_{0.5}(Y_{ij})$ is the sample median of Y_{ij}, for $i = 1,\ldots,n$ and $j = 1,\ldots,N_i$. We rank the variables based on their $\hat{f}_{nk}^m$-value and keep the top $[n/\ln(n)]$ variables (as in He et al. (2013)). We denote the estimated set of active variables as $\hat{M}_{SIS}$.

5.3 Variable selection – Stage 2

In the second stage, we start from the selected set of the first stage and do a further dimension reduction with a variable selection technique. Below we describe the different variable selection techniques that are considered in the simulation study.

5.3.1 Grouped Adaptive Lasso with B-splines after grouped Lasso

In this method, we conduct the grouped Adaptive Lasso in Tang et al. (2013b) on the reduced model in Section 5.2.1. This procedure is based on unpenalized B-splines and denoted as $\mathrm{gALassoB}_{\mathrm{Lasso}}$. We solve

$$\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(Y_{ij} - \mathbf{U}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}\right) + \lambda_L \sum_{k\in\hat{M}_{\mathrm{Lasso}}}\omega_{2k}\|\boldsymbol{\alpha}_k\|_1, \tag{5.3}$$

where $\lambda_L > 0$ is a regularization parameter that determines the sparsity of the solution; $\omega_{2k} = \|\tilde{\boldsymbol{\alpha}}_k\|_1^{-\eta_L}$, for $k \in \hat{M}_{\mathrm{Lasso}}$, are the penalty weights of the estimated coefficient group in Section 5.2.1, with $\eta_L \geq 1$. We use the same number of knots as obtained in Section 5.2.1. Then, the SIC in Equation (5.1) is used to obtain a data-driven λ_L.

5.3.2 Grouped Adaptive Lasso with P-splines after grouped Lasso

This procedure is denoted as $\text{gALassoP}_{\text{Lasso}}$. In order to enforce smooth estimates, we add a P-splines penalty to the previous optimization problem (Equation (5.3)):

$$\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left(Y_{ij} - \mathbf{U}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}\right) + \lambda_P \sum_{k \in \hat{M}_{\text{Lasso}}} \omega_{1k} \|\mathbf{D}_m^1 \boldsymbol{\alpha}_k\|_1 + \lambda_L \sum_{k \in \hat{M}_{\text{Lasso}}} \omega_{2k} \|\boldsymbol{\alpha}_k\|_1, \tag{5.4}$$

where λ_P and λ_L are the smoothing and regularization parameters, respectively; $\omega_{1k} = \text{Range}(\tilde{\beta}_k(T))^{-\eta_P}$, with $\text{Range}(\tilde{\beta}_k(T))$ the range of quantile regression estimators obtained in Section 5.2.1; and $\omega_{2k} = \|\tilde{\boldsymbol{\alpha}}_k\|_1^{-\eta_L}$ the penalty weights of the estimated coefficient group from Section 5.2.1. The tuning parameters λ_P and λ_L are chosen data-driven, by minimizing the SIC in Equation (5.1).

5.3.3 NonNegative Garrote after grouped Lasso

The procedure is denoted as $\text{NNG}_{\text{Lasso}}$. On the reduced set of covariates (from Section 5.2.1), we apply the NonNegative Garrote. The NonNegative Garrote procedure is based on the initial estimates of the $\beta_k(t)$ ($\hat{\beta}_k^{\text{init}}(t)$). These initial estimates are the P-splines estimates, i.e. the minimizers of Equation (5.4) with $\lambda_L = 0$. Then, using these estimates, the NonNegative Garrote shrinkage factors $\hat{\mathbf{c}} = (\hat{c}_k : k \in \hat{M}_{\text{Lasso}})$ are obtained by solving the following optimization problem:

$$\min_{c_k : k \in \hat{M}_{\text{Lasso}}} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left(Y_{ij} - \sum_{k \in \hat{M}_{\text{Lasso}}} c_k \hat{\beta}_k^{\text{init}}(t_{ij}) X_{ij}^{(k)}\right) + \lambda_{\text{NNG}} \sum_{k \in \hat{M}_{\text{Lasso}}} c_k$$
$$\text{s.t. } c_k \geq 0 \ (k \in \hat{M}_{\text{Lasso}}), \tag{5.5}$$

where $\lambda_{\text{NNG}} > 0$ is a regularization parameter (chosen data-driven by minimizing the SIC in Equation (5.1)). Finally, the covariates corresponding to $\hat{c}_k > 0$ are retained.

5.3.4 Grouped Adaptive Lasso with P-splines after SIS

The procedure is denoted as $\text{gALassoP}_{\text{SIS}}$ and is based on the SIS-selected set in the first stage and the grouped Adaptive Lasso with a P-splines penalty as in Equation (5.4) in the second stage, where $\tilde{\beta}$ and $\tilde{\boldsymbol{\alpha}}$ in Equation (5.4) are now the minimizers of Equation (5.4) with $\lambda_P = \lambda_L = 0$.

5.3.5 NonNegative Garrote after SIS

We refer to the procedure, denoted as NNG_{SIS}, as the method that is based on a first stage SIS and the NNG in the second stage (minimizing Equation (5.5), replacing $\hat{M}_{\text{Lasso}}$ by $\hat{M}_{\text{SIS}}$).

5.4 Simulation study

In this section, we investigate the performance of the five variable selection procedures described in Section 5.3 on two simulation settings: a heteroscedastic model with uncorrelated covariates (setting 1) and a homoscedastic model with correlated covariates (setting 2).

5.4.1 Setting 1: heteroscedastic model

We consider the following heteroscedastic model:

$$Y_{ij} = \beta_0(t_{ij}) + \sum_{k=1}^{p} \beta_k(t_{ij}) X_{ij}^{(k)} + 2.1 \exp\left(\frac{\beta_3(t_{ij})}{10} X_{ij}^{(6)}\right) \epsilon_{ij}.$$

The coefficients (also presented in Figure 4.1), considered by Tang et al. (2013b), are formulated as follows,

$$\beta_0(T) = 15 + 20\sin(\pi T/60) \qquad \beta_1(T) = 2 - 3\cos(\pi(T-25)/15) \qquad \beta_2(T) = 6 - 0.2T$$

$$\beta_3(T) = -4 + (20-T)^3/2000 \qquad \beta_k(T) = 0 \text{ for } k = 4, \ldots, p.$$

The error term is generated from a transformed multivariate normal distribution as in Equation (2.14) (to fulfill assumptions **H1** and **H2**), with $\zeta_{ij} = E_{ij} + Z_{ij}$, with $E_{ij} \sim N(0,4)$ and Z_{ij} is generated from a Gaussian process with zero mean and an exponential decaying covariance

$$cov(Z_{ij}, Z_{lz}) = \begin{cases} 4\exp(-|t_{ij} - t_{lz}|) & \text{if } i = l, \\ 0 & \text{if } i \neq l. \end{cases}$$

The covariates are generated from the same distribution as Z_{ij}. The time variable T ranges from $1, 2, \ldots, 30$. Each time point (excluding time 1) has 80% probability of

being skipped, therefore creating an unbalanced design. Then, the actual time point is obtained by adding a Unif$[-0.5, 0.5]$.

We generate 500 data sets of size $n = 50$ and consider $p = 100, 200$ and 400. The choice of the covariates, coefficient functions and error structure results in a Signal-to-Noise Ratio of approximately seven.

We use a grid of length five for the smoothing parameters u and λ_P, and a grid of length ten for the regularization parameter λ_L everywhere throughout the variable selection procedure. B-splines of degree $\nu_k = 3$ are used. When eight knots are used and $p = 100$, $m_{tot} = 1010$ parameters (α_{kl}) should be estimated based on $N \approx 355$ observations. As in the previous chapter, we set $\eta_P = 1$ and $\eta_L = 2$.

We measure the performance of the methods based on the same criteria as in Chapter 4 (see Table 4.1).

Comparison of the first and second stage methods

For the median function, at the first stage both the Lasso and SIS have an oracle percentage that is zero (Table 5.1); note that the SIS always selects a fixed amount of variables $(n/\ln(n))$, so it is expected that too many variables are selected. However, they both preserve the important variables (good percentage close to 100). In the second stage, the gALassoP$_{\text{Lasso}}$ has the highest oracle percentage (93%). The other procedures, except the gALassoP$_{\text{SIS}}$, perform also well.

Table 5.2 shows the results for the variability function. In that table we present the result of Stage 1 for each procedure, since the result from the median (more precisely $\hat{\beta}_k$) is used to obtain the estimated active set for the variability function. The SIS has good performance in terms of preserving the relevant variables. The NNG$_{\text{SIS}}$ has the best performance at the second stage with an oracle percentage of 94.8%. The NNG$_{\text{Lasso}}$ also performs well (oracle percentage of 72%). The NNG$_{\text{SIS}}$ outperforms all the methods in terms of the computing time, when focusing on the median or the variability function.

Comparison of tuning parameter selection criteria

In Section 5.2.1, we discuss a SIC to choose the smoothing and regularization parameters in the first and second stages. Tang et al. (2013a) propose an Extended Bayesian

Table 5.1. Setting 1. Performance for the median function using SIC for $p = 100$ with data-driven choice of number of knots.

	Stage 1		Stage 2				
Method	Lasso	SIS	$\text{gALassoB}_{\text{Lasso}}$	$\text{gALassoP}_{\text{Lasso}}$	$\text{NNG}_{\text{Lasso}}$	$\text{gALassoP}_{\text{SIS}}$	NNG_{SIS}
Oracle %	0.0	0.0	87.6	93.0	67.2	3.6	67.6
Good %	100.0	94.2	100.0	100.0	100.0	94.2	94.2
+1 %	0.0	0.0	11.4	6.4	12.8	7.8	18.0
-1 %	0.0	5.8	0.0	0.0	0.0	5.8	5.8
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)
Med.z	17 (14, 21)	9 (9, 9)	0 (0, 0)	0 (0, 0)	0 (0, 1)	4 (2, 5)	0 (0, 1)
	–	–	41.6	36.5	41.5	6.8	5.8
Med.t	–	–	(39.5,	(34.6,	(39.3,	(6.6,	(5.7,
	–	–	43.9)	38.8)	43.6)	7.3)	6.3)

NOTE: The numbers in brackets are the first and third quartiles.

Table 5.2. Setting 1. Performance for the variability function using SIC for $p = 100$ with data-driven choice of number of knots.

	$\text{gALassoB}_{\text{Lasso}}$		$\text{gALassoP}_{\text{Lasso}}$		$\text{NNG}_{\text{Lasso}}$		$\text{gALassoP}_{\text{SIS}}$		NNG_{SIS}	
Method	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Oracle %	42.6	8.8	0.0	14.4	10.4	72.0	0.0	21.0	0.0	94.8
Good %	58.0	9.0	21.6	15.0	74.8	72.2	97.8	94.0	98.4	95.2
+1 %	5.8	0.2	0.0	0.4	3.0	0.2	0.0	22.8	0.0	0.4
-1 %	42.0	91.0	78.4	85.0	25.2	27.8	2.2	6.0	1.6	4.8
Med.r	1(0,1)	0(0,0)	0(0,0)	0(0,0)	1(0,1)	1(0,1)	1(1,1)	1(1,1)	1(1,1)	1(1,1)
Med.z	0(0,0)	0(0,0)	0(0,0)	0(0,0)	9(0,15)	0(0,0)	11(11,11)	2(1,3)	11(11,11)	0(0,0)
	–	37.2	–	32.4	–	35.0	–	5.5	–	4.7
Med.t	–	(35.0,	–	(30.8,	–	(33.2,	–	(5.3,	–	(4.3,
	–	38.6)	–	34.1)	–	36.3)	–	5.9)	–	4.9)

NOTE: S1: Stage 1; S2: Stage 2; the numbers in brackets are the first and third quartiles.

Information Criterion (EBIC) for tuning parameter selection in the second stage:

$$\text{EBIC}(\lambda_P, \lambda_L) = \ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\rho_{0.5}\left(Y_{ij} - \mathbf{U}_{ij}^{\mathrm{T}}\hat{\boldsymbol{\alpha}}\right)\right) + \frac{\ln(N)}{2N}p_{\lambda_P,\lambda_L} + \vartheta\frac{\ln(|\hat{M}_1|)}{2N}p_{\lambda_P,\lambda_L},$$

where $\hat{M}_1$ is the selected set from the first stage (either $\hat{M}_{\text{SIS}}$ or $\hat{M}_{\text{Lasso}}$); $\hat{\boldsymbol{\alpha}}$ are the estimates from the second stage; and $0 \leq \vartheta \leq 1$ (we take $\vartheta = 0.5$ as in Tang et al. (2013a)).

We compare the performance of the SIC and EBIC in this paragraph. Tables 5.3 and 5.4 summarize the results when the EBIC is used in the second stage. All procedures improve slightly, except for the $\text{NNG}_{\text{Lasso}}$ in the variability function, when the EBIC is used to choose the tuning parameters.

Table 5.3. Setting 1. Performance for the median function using EBIC for $p = 100$ with data-driven choice of number of knots.

	Stage 1		Stage 2				
Method	Lasso	SIS	$\text{gALassoB}_{\text{Lasso}}$	$\text{gALassoP}_{\text{Lasso}}$	$\text{NNG}_{\text{Lasso}}$	$\text{gALassoP}_{\text{SIS}}$	NNG_{SIS}
Oracle %	0.0	0.0	95.6	97.8	72.6	3.8	74.2
Good %	100.0	95.6	100.0	100.0	100.0	94.2	95.6
+1 %	0.0	0.0	4.2	2.0	12.2	9.4	14.4
-1 %	0.0	4.4	0.0	0.0	0.0	5.8	4.4
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)
Med.z	17 (14, 21)	9 (9, 9)	0 (0, 0)	0 (0, 0)	0 (0, 1)	4 (2, 5)	0 (0, 0)
	–	–	32.2	31.2	30.3	6.8	3.1
Med.t	–	–	(30.5,	(30.2,	(29.2,	(6.6,	(3.1,
	–	–	34.1)	32.7)	31.7)	7.3)	3.2)

NOTE: The numbers in brackets are the first and third quartiles.

Choice of the number of knots

In order to reduce the computing time for the Lasso in stage 1, we now fix the number of knots to $u + 1 = 5$ (as in Tang et al., 2013a) rather than selecting it data-driven. Then, the number of B-spline coefficients is $m_{tot} = 707$ (when $p = 100$). We compare again the two information criteria, for both the median (Tables 5.5 and 5.7) and variability functions (Tables 5.6 and 5.8). The tables show that again the procedures, generally, improve when the EBIC is used. When it comes to whether the number of knots should be chosen in a data-driven way or not, we conclude from Tables 5.3 and 5.7 for the median and Tables 5.4 and 5.8 for the variability function, that fixing the number of knots does not worsen the performance, in general. The $\text{gALassoP}_{\text{SIS}}$ and NNG_{SIS} are performing better when the number of knots is fixed at 5.

Table 5.4. Setting 1. Performance for the variability function using EBIC for $p = 100$ with data-driven choice of number of knots.

	$\text{gALassoB}_{\text{Lasso}}$		$\text{gALassoP}_{\text{Lasso}}$		$\text{NNG}_{\text{Lasso}}$		$\text{gALassoP}_{\text{SIS}}$		NNG_{SIS}	
Method	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Oracle %	0.0	14.2	0.0	13.0	0.0	64.2	0.0	24.2	0.0	95.2
Good %	17.6	14.6	20.4	13.2	66.2	64.4	97.8	91.4	98.6	95.4
+1 %	0.6	0.4	0.2	0.2	0.0	0.0	0.0	25.0	0.0	0.2
-1 %	82.4	85.4	79.6	86.8	33.8	35.6	2.2	8.6	1.4	4.6
Med.r	0(0,0)	0(0,0)	0(0,0)	0(0,0)	1(0,1)	1(0,1)	1(1,1)	1(1,1)	1(1,1)	1(1,1)
Med.z	0(0,0)	0(0,0)	0(0,0)	0(0,0)	10(0,16)	0(0,0)	11(11,11)	1(0,3)	11(11,11)	0(0,0)
	–	30.6	–	28.6	–	26.8	–	5.5	–	3.0
Med.t	–	(28.6,	–	(27.0,	–	(25.8,	–	(5.3,	–	(2.9,
	–	32.4)	–	30.5)	–	28.4)	–	5.9)	–	3.1)

NOTE: S1: Stage 1; S2: Stage 2; the numbers in brackets are the first and third quartiles.

Table 5.5. Setting 1. Performance for the median function using SIC for $p = 100$ with $u = 4$.

	Stage 1		Stage 2				
Method	Lasso	SIS	$\text{gALassoB}_{\text{Lasso}}$	$\text{gALassoP}_{\text{Lasso}}$	$\text{NNG}_{\text{Lasso}}$	$\text{gALassoP}_{\text{SIS}}$	NNG_{SIS}
Oracle %	0.0	0.0	91.8	92.0	64.4	31.0	77.0
Good %	100.0	95.6	100.0	100.0	100.0	95.6	96.0
+1 %	0.0	0.0	7.8	7.4	18.8	33.8	14.2
-1 %	0.0	4.4	0.0	0.0	0.0	4.4	4.0
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)
Med.z	14 (11, 18)	9 (9, 9)	0 (0, 0)	0 (0, 0)	0 (0, 1)	1 (0, 2)	0 (0, 0)
	–	–	5.1	6.1	5.2	4.8	4.1
Med.t	–	–	(4.9,	(5.7,	(5.1,	(4.6,	(3.9,
	–	–	5.3)	6.4)	5.4)	5.6)	4.5)

NOTE: The numbers in brackets are the first and third quartiles.

Influence of the number of covariates

In the bottom part of Tables 5.7 and 5.8, the performance of the variable selection procedures are presented for $p = 200$, using $u = 4$ and the EBIC for the second stage.

Table 5.6. Setting 1. Performance for the variability function using SIC for $p = 100$ with $u = 4$.

	$\text{gALassoB}_{\text{Lasso}}$		$\text{gALassoP}_{\text{Lasso}}$		$\text{NNG}_{\text{Lasso}}$		$\text{gALassoP}_{\text{SIS}}$		NNG_{SIS}	
Method	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Oracle %	0.0	3.4	0.0	8.0	0.0	48.2	0.0	36.6	0.0	97.2
Good %	4.0	3.6	11.8	8.2	48.8	48.6	99.8	94.0	99.8	97.4
+1 %	0.0	0.2	0.0	0.2	0.0	0.2	0.0	23.8	0.0	0.2
-1 %	96.0	96.4	88.2	91.8	51.2	51.4	0.2	6.0	0.2	2.6
Med.r	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,1)	0(0,1)	1(1,1)	1(1,1)	1(1,1)	1(1,1)
Med.z	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,13)	0(0,0)	11(11,11)	1(0,2)	11(11,11)	0(0,0)
	–	4.6	–	4.5	–	4.5	–	4.6	–	3.7
Med.t	–	(4.4,	–	(4.3,	–	(4.2,	–	(4.4,	–	(3.6,
	–	4.9)	–	4.8)	–	4.7)	–	4.8)	–	3.9)

NOTE: S1: Stage 1; S2: Stage 2; the numbers in brackets are the first and third quartiles.

The $\text{gALassoP}_{\text{Lasso}}$ also here performs best when it comes to variable selection for the median and bad for the variability function. We can conclude again that the NNG_{SIS} performs well for the median as well as the variability function. When increasing the number of covariates to $p = 400$ ($N \approx 355 < m_{tot} = 2807$), the conclusion is similar with slightly worse performance for all procedures (see Tables 5.9 and 5.10).

Conclusion

Under Setting 1, the criterion EBIC improves the performance of the procedures compared to the SIC. The EBIC resulted in an improvement whether a fixed or data-driven selected number of knots is used. Further, fixing the number of knots, generally, does not worsen the performance of the procedures. Then, using $u = 4$ and the EBIC, the procedures are compared when $p = 100$, 200, and 400. In all cases, the NNG_{SIS} has good performance for both the median and variability functions, while using significantly less computing time. There is a slight reduction in the performance for all procedures as p increases.

In the previous chapter, the gALassoB, gALassoP and NNG perform equivalently well in terms of variable selection. In this chapter, after the Lasso first stage, the gALassoP perform the best compared to the others for the median function. Where

Table 5.7. Setting 1. Performance for the median function using EBIC with $u = 4$.

	Stage 1		Stage 2				
Method	Lasso	SIS	$\text{gALassoB}_{\text{Lasso}}$	$\text{gALassoP}_{\text{Lasso}}$	$\text{NNG}_{\text{Lasso}}$	$\text{gALassoP}_{\text{SIS}}$	NNG_{SIS}
For $p = 100$							
Oracle %	0.0	0.0	91.8	94.8	72.4	43.0	80.8
Good %	100.0	95.6	100.0	100.0	100.0	96.0	95.6
+1 %	0.0	0.0	7.8	4.8	15.2	29.6	11.4
-1 %	0.0	4.4	0.0	0.0	0.0	4.0	4.4
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)
Med.z	14 (11, 18)	9 (9, 9)	0 (0, 0)	0 (0, 0)	0 (0, 0)	1 (0, 2)	0 (0, 0)
	–	–	5.1	6.1	5.4	4.6	2.7
Med.t	–	–	(4.9,	(5.7,	(5.2,	(4.4,	(2.6,
	–	–	5.3)	6.4)	5.6)	5.1)	2.8)
For $p = 200$							
Oracle %	0.0	0.0	88.6	91.4	73.6	40.0	75.0
Good %	100.0	92.4	100.0	100.0	100.0	90.8	92.4
+1 %	0.0	0.0	10.8	8.0	8.8	27.0	13.6
-1 %	0.0	7.6	0.0	0.0	0.0	9.2	7.6
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)
Med.z	25 (20, 31.25)	9 (9, 9)	0 (0, 0)	0 (0, 0)	0 (0, 1)	1 (0, 2)	0 (0, 0)
	–	–	63.6	50.4	49.7	15.1	5.2
Med.t	–	–	(59.1,	(44.7,	(46.3,	(13.3,	(5.1,
	–	–	67.9)	53.9)	53.2)	16.5)	5.3)

NOTE: The numbers in brackets are the first and third quartiles.

as after the SIS first stage, the NNG perform better than the gALassoP for both the median and variability functions.

5.4.2 Setting 2: homoscedastic model

We investigate the performance of the proposed methods, when the covariates are correlated. The data are generated from the following model (longitudinal version of

Table 5.8. Setting 1. Performance for the variability function using EBIC with $u = 4$.

	$\text{gALassoB}_{\text{Lasso}}$		$\text{gALassoP}_{\text{Lasso}}$		$\text{NNG}_{\text{Lasso}}$		$\text{gALassoP}_{\text{SIS}}$		NNG_{SIS}	
Method	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
For $p = 100$										
Oracle %	0.0	3.4	0.0	7.6	0.0	51.4	0.0	41.4	0.0	96.2
Good %	4.0	3.6	13.4	7.8	52.0	51.6	99.6	91.8	99.8	96.4
+1 %	0.0	0.2	0.0	0.2	0.0	0.2	0.0	21.8	0.0	0.2
-1 %	96.0	96.4	86.6	92.2	48.0	48.4	0.4	8.2	0.2	3.6
Med.r	0(0,0)	0(0,0)	0(0,0)	0(0,0)	1(0,1)	1(0,1)	1(1,1)	1(1,1)	1(1,1)	1(1,1)
Med.z	0(0,0)	0(0,0)	0(0,0)	0(0,0)	7(0,13)	0(0,0)	11(11,11)	1(0,2)	11(11,11)	0(0,0)
	–	4.6	–	4.6	–	4.7	–	4.4	–	2.7
Med.t	–	(4.4,	–	(4.3,	–	(4.4,	–	(4.3,	–	(2.6,
	–	4.9)	–	4.9)	–	4.9)	–	4.6)	–	2.7)
For $p = 200$										
Oracle %	0.0	0.4	0.0	1.4	0.0	23.0	0.0	31.8	0.0	92.2
Good %	0.4	0.4	1.6	1.4	24.2	23.0	97.6	84.6	99.0	92.8
+1 %	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.2	0.0	0.6
-1 %	99.6	99.6	98.4	98.6	75.8	77.0	2.4	15.4	1.0	7.2
Med.r	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,0)	1(1,1)	1(1,1)	1(1,1)	1(1,1)
Med.z	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,8)	0(0,0)	11(11,11)	1(0,2)	11(11,11)	0(0,0)
	–	63.7	–	46.9	–	45.2	–	13.4	–	5.2
Med.t	–	(59.3,	–	(40.1,	–	(40.1,	–	(12.8,	–	(5.1,
	–	68.7)	–	50.8)	–	49.2)	–	17.6)	–	5.3)

NOTE: S1: Stage 1; S2: Stage 2; the numbers in brackets are the first and third quartiles.

Tang et al. (2013a)):

$$Y_{ij} = \beta_0(t_{ij}) + \sum_{k=1}^{p} \beta_k(t_{ij}) X_i^{(k)} + \epsilon_{ij}.$$

Table 5.9. Setting 1: Performance for the median function using EBIC for $p = 400$ with $u = 4$.

	Stage 1		Stage 2				
Method	Lasso	SIS	$\text{gALassoB}_{\text{Lasso}}$	$\text{gALassoP}_{\text{Lasso}}$	$\text{NNG}_{\text{Lasso}}$	$\text{gALassoP}_{\text{SIS}}$	NNG_{SIS}
Oracle %	0.0	0.0	86.4	89.0	66.2	32.0	70.0
Good %	100.0	86.8	100.0	100.0	99.8	86.6	86.8
+1 %	0.0	0.0	13.2	10.0	6.0	32.6	12.6
-1 %	0.0	12.6	0.0	0.0	0.2	12.8	12.6
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)
Med.z	44 (37, 51)	9 (9, 9)	0 (0, 0)	0 (0, 0)	0 (0, 2)	1 (0, 2)	0 (0, 0)
	–	–	213.8	271.0	265.3	10.6	9.8
Med.t	–	–	(201.7,	(259.1,	(252.2,	(10.4,	(9.6,
	–	–	233.7)	287.0)	283.1)	10.8)	9.9)

NOTE: The numbers in brackets are the first and third quartiles.

Table 5.10. Setting 1: Performance for the variability function using EBIC for $p = 400$ with $u = 4$.

	$\text{gALassoB}_{\text{Lasso}}$		$\text{gALassoP}_{\text{Lasso}}$		$\text{NNG}_{\text{Lasso}}$		$\text{gALassoP}_{\text{SIS}}$		NNG_{SIS}	
Method	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Oracle %	0.0	0.0	0.0	0.0	0.0	7.4	0.0	28.6	0.0	89.2
Good %	0.0	0.0	0.0	0.0	9.4	7.6	97.0	83.6	96.6	89.6
+1 %	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.2	0.0	0.4
-1 %	100.0	100.0	100.0	100.0	90.6	92.4	3.0	16.4	3.4	10.4
Med.r	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,0)	1(1,1)	1(1,1)	1(1,1)	1(1,1)
Med.z	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,0)	11(11,11)	1(0,2)	11(11,11)	0(0,0)
	–	223.5	–	265.2	–	252.4	–	10.9	–	9.6
Med.t	–	(205.7,	–	(246.8,	–	(235.4,	–	(10.7,	–	(9.4,
	–	239.6)	–	281.2)	–	271.0)	–	11.1)	–	9.8)

NOTE: S1: Stage 1; S2: Stage 2; the numbers in brackets are the first and third quartiles.

The coefficient functions are given by

$$\beta_0(T) = 15 + 20\sin(\pi T/2) \quad \beta_1(T) = 2 - 3\cos(\pi(6T-5)/3) \quad \beta_2(T) = 6 - 6T$$

$$\beta_3(T) = 3 \qquad \beta_k(T) = 0 \text{ for } k = 4, \ldots, p.$$

The covariates $\left(X_i^{(1)}, \ldots, X_i^{(p)}\right)$ are generated from a multivariate normal $N(\mathbf{0}, \mathbf{\Sigma})$-distribution with $\mathbf{\Sigma}_{jk} = 0.5^{|j-k|}$ for $j, k = 1, \ldots, p$. The error ϵ_{ij} and the time variable t_{ij} are generated as in Section 5.4.1. To ensure that the time variable is between zero

and one, we standardize it as in Equation (3.9). We consider $p = 100$ and 400.

All results in this section are based on B-splines of degree 3 with $u = 4$ and the EBIC in stage 2 is used to select the tuning parameters.

Table 5.11 shows that all the procedures are performing well for the median function, except for the gALassoP$_{\text{SIS}}$. We see under this setting that at stage 1 the grouped Lasso has an oracle percentage of 52.4 and the median number of redundant variables selected is zero. Even though the oracle percentage of SIS is zero as expected (because of the $n/\ln(n)$ cutoff point), the good percentage is 85.8. The NNG$_{\text{Lasso}}$ performs very well, even when $p = 400$. The NNG$_{\text{SIS}}$ is able to select the true model 83.2% and 73.2% of times, respectively, for $p = 100$ and $p = 400$. Here also we see that the NNG$_{\text{SIS}}$ outperforms in terms of computating time: the relative computing time (for $p = 400$) compared to the gALassoP$_{\text{SIS}}$, NNG$_{\text{Lasso}}$, gALassoP$_{\text{Lasso}}$ and gALassoB$_{\text{Lasso}}$, respectively, is 0.93, 0.05, 0.04, and 0.05.

The results of variable selection in the variability function are given in Table 5.12. All the procedures perform very good, except for the gALassoP$_{\text{SIS}}$. The true model is selected by the NNG$_{\text{SIS}}$ 99.2% and 97% of times, respectively, for $p = 100$ and $p = 400$.

Conclusion

In Setting 2, we see that all the procedures have good performance for both the median and variability functions, except for the gALassoP$_{\text{SIS}}$. However, the NNG$_{\text{SIS}}$ again outperforms in the computing time, especially when $p = 400$.

5.5 Data example: UK Employment data

In this section we illustrate the performance of the methods on the UK Employment data. The data set contains company accounts from Datastream International which provides accounts records of employment and remuneration (i.e. wage bill) for 140 United Kingdom quoted companies whose main activity is manufacturing from 1976 to 1984. The companies have unbalanced number of measurements ranging from seven to nine. The data set is available in the plm R-package (Croissant and Millo, 2008) named EmplUK. More details on the data can be found in the data appendix of Arellano and Bond (1991). The response variable $Y(t_{ij})$ is the logarithm of number of employees in company i at time t_{ij} . The covariates are $X^{(1)}(t_{ij})$ the average annual

Table 5.11. Setting 2. Performance for the median function using EBIC with $u = 4$.

	Stage 1		Stage 2				
Method	Lasso	SIS	$\text{gALassoB}_{\text{Lasso}}$	$\text{gALassoP}_{\text{Lasso}}$	$\text{NNG}_{\text{Lasso}}$	$\text{gALassoP}_{\text{SIS}}$	NNG_{SIS}
For $p = 100$							
Oracle %	52.4	0.0	92.0	92.8	94.0	56.4	83.2
Good %	95.6	85.8	93.6	93.4	95.6	85.8	85.8
+1 %	29.4	0.0	1.6	0.6	1.6	18.4	2.4
-1 %	4.2	13.2	5.8	6.0	4.2	13.2	13.2
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)
Med.z	0 (0, 1)	9 (9, 9)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 1)	0 (0, 0)
	–	–	4.8	5.0	4.9	3.4	2.7
Med.t	–	–	(4.7,	(4.9,	(4.8,	(3.2,	(2.7,
	–	–	5.0)	5.1)	5.0)	3.5)	2.8)
For $p = 400$							
Oracle %	14.4	0.0	92.2	94.6	92.4	51.2	73.2
Good %	97.2	76.4	96.0	96.0	97.2	76.6	76.4
+1 %	26.0	0.0	3.2	1.2	4.4	18.4	3.0
-1 %	2.6	21.4	3.6	3.6	2.6	21.2	21.4
Med.r	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)	3 (3, 3)
Med.z	2 (1, 3)	9 (9, 9)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 2)	0 (0, 0)
	–	–	192.8	241.5	193.3	10.8	10.0
Med.t	–	–	(185.7,	(231.7,	(185.0,	(10.6,	(9.8,
	–	–	202.1)	252.7)	202.8)	11.0)	10.2)

NOTE: The numbers in brackets are the first and third quartiles.

wage per employee in the company, $X^{(2)}(t_{ij})$ the capital defined as the book value of gross fixed assets, and $X^{(3)}(t_{ij})$ an index of value-added output at constant cost. This data set has also been used by Andriyana and Gijbels (2017). In addition to the three original covariates, $p - 3$ artificial variables are added. The artificial variables are independent of the original covariates and have the same distribution as Z_{ij} in Section 5.4.1.

Table 5.12. Setting 2. Performance for the variability function using EBIC with $u = 4$.

	$\text{gALassoB}_{\text{Lasso}}$		$\text{gALassoP}_{\text{Lasso}}$		$\text{NNG}_{\text{Lasso}}$		$\text{gALassoP}_{\text{SIS}}$		NNG_{SIS}	
Method	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
For $p = 100$										
Oracle %	98.0	100.0	98.8	100.0	97.8	99.8	0.0	68.2	0.0	99.2
+1 %	1.8	0.0	1.0	0.0	2.0	0.2	0.0	19.4	0.0	0.8
Med.z	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,0)	12(12,12)	0(0,1)	12(12,12)	0(0,0)
	–	4.7	–	4.4	–	4.4	–	3.5	–	2.8
Med.t	–	(4.4,	–	(4.2,	–	(4.2,	–	(3.4,	–	(2.7,
	–	5.0)	–	4.7)	–	4.6)	–	3.6)	–	2.9)
For $p = 400$										
Oracle %	97.8	100.0	99.6	100.0	94.2	100.0	0.0	53.4	0.0	97.0
+1 %	1.6	0.0	0.0	0.0	4.2	0.0	0.0	25.0	0.0	1.8
Med.z	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,0)	0(0,0)	12(12,12)	0(0,1)	12(12,12)	0(0,0)
	–	210.2	–	255.1	–	189.8	–	11.0	–	9.9
Med.t	–	(194.4,	–	(237.7,	–	(172.5,	–	(10.8,	–	(9.7,
	–	224.2)	–	272.5)	–	207.6)	–	11.3)	–	10.2)

NOTE: S1: Stage 1; S2: Stage 2; the numbers in brackets are the first and third quartiles.

Five different values for the number of covariates are considered: $p = 50, 100, 200, 400,$ and 1000. The covariates are standardized as in Section 3.2.1, so that they have equivalent scale. To analyze the data set, B-splines of degree three with five quasi-uniform knots are used (as in Section 5.4).

Table 5.13 shows the number of artificial variables included in the model. For all values of p, all the methods perform well. Even when the number of covariates is 1000 ($m_{tot} = 8000$ parameters to be estimated), the faster NNG_{SIS} selects only 2 and 3 artificial variables, respectively, for the median and the variability functions. Using the NNG_{SIS}, $X_{ij}^{(2)}$ is kept in the model for the median. Where as for the variability function, none of the original variables are retained.

Table 5.13. UK Employment data. Number of artificial variables selected in the model.

p	$gALassoB_{Lasso}$	$gALassoP_{Lasso}$	NNG_{Lasso}	$gALassoP_{SIS}$	NNG_{SIS}
For Median					
50	0	0	0	0	0
100	0	0	0	1	1
200	0	0	0	0	1
400	0	0	0	0	4
1000	0	0	0	16	2
For Variability function					
50	0	0	0	2	3
100	0	0	0	1	0
200	0	0	0	1	2
400	0	0	0	5	1
1000	0	0	0	9	3

5.6 Conclusion

For both theoretical and computational purpose, Tang et al. (2013a) propose a two-stage approach for cross sectional data when the number of coefficients to be estimated (m_{tot}) is much bigger than n. They use the grouped Lasso for the first stage and the grouped Adaptive Lasso for the second stage. In this chapter, we propose another two-stage approach (NNG_{SIS}) when we have longitudinal observations (when the number of coefficients to be estimated is much bigger than the number of observations N). In the first stage, we consider the Sure Independence Screening to reduce the size of a model from ultrahigh-dimensional to a model which has size close to the true model. In the second stage, we apply the NonNegative Garrote (proposed in Chapter 4) to the reduced model, to have a method which is consistent in variable selection. We further adapt the approach of Tang et al. (2013a) in the second stage where we penalize the B-splines ($gALassoP_{Lasso}$). Two other variable selection techniques ($gALassoP_{SIS}$ and

$\text{NNG}_{\text{Lasso}}$) are also investigated.

From the simulation study, we observe that if we are interested in only the median function, the $\text{gALassoP}_{\text{Lasso}}$ performs best. However, if one is interested in the other quantiles (and therefore the variability function) in addition to the median, then the NNG_{SIS} performs very good in terms of selection in both the median and variability functions. The NNG_{SIS} further outperforms all the considered methods in the computing time. In general, NNG_{SIS} is recommended to do variable selection and estimation in an ultrahigh-dimensional setting.

In addition, we compare two types of information criteria for tuning parameter selection in the second stage. The procedures perform well when the EBIC is used. The number of knots can be fixed, since the procedures did not show worse performance when the number of knots is fixed in the simulation study. The $\text{gALassoP}_{\text{SIS}}$ is performing better (moderate performance for both the median and variability functions) when the number of knots is fixed at 5.

In the homoscedastic model with correlated covariates, all the methods perform better, especially the $\text{NNG}_{\text{Lasso}}$. This is due to the fact that the number of irrelevant variables included in the model by the grouped Lasso (in the first stage) is small.

The proposed procedures are illustrated on a data example. Although the NNG_{SIS} has good performance in the simulation study, it is of interest to show its consistency.

Chapter 6

Discussion and further research

This dissertation considers varying coefficient models in a quantile regression setting while allowing for heteroscedasticity. In addition of being able to investigate the entire distribution of the response, quantile regression is robust to a heavy-tailed distributions. Here, we focus on longitudinal data. For such data, it is more realistic to consider the coefficients (in the varying coefficient model) of each covariate to vary with 'time'. Several smoothing techniques have been considered in the literature to estimate the varying coefficients. As in Andriyana et al. (2014), the P-splines smoothing technique is used in this thesis. This technique has a computational advantage and eases the interpretation of the estimates by avoiding wiggliness of the B-splines.

In practice, not all data have a homoscedastic nature, as has been seen in CD4, Grunfeld, KTB and Intego data. To avoid this strong assumption, we allow for heteroscedasticity throughout the thesis. In order to avoid the crossing of the estimated quantile regression curves (when several quantiles are of interest) and to estimate the variability function, we adapt the approach of He (1997) to the context of the varying coefficient models. A Likelihood-Ratio-Type (LRT) test is developed in Chapter 2, to identify the structure of the heteroscedasticity (the variability function or the variance of errors). The identification of the variability function is important, as the estimation of the quantiles other than the median depend on the specific variability function. For the CD4 data, a simple heteroscedastic variability structure (that only depends on time) is recommended based on the LRT test.

The Wages data in Section 3.3 is found to have constant functional coefficients for both the median and variability functions. Considering a varying coefficient model for such

data will only complicate the interpretation and might be too complex. Hence, based on the varying coefficient model (with a very flexible variability function referred to as V_4, the full model) introduced in Chapter 2, testing procedures that check the constancy of the functional coefficients are developed in Chapter 3. Further, a testing procedure that checks for monotonicity or convexity of a functional coefficient is proposed. We also propose a testing procedure that can simultaneously check the shape of the functional coefficients in both the signal and variability functions.

The Intego data in Section 4.3.3 has several covariates. It is difficult to interpret the association of the blood pressure values with the eGFR (the measure of the kidney function of the patients) considering all the covariates. Hence, in Chapter 4, two variable selection techniques (the grouped Adaptive Lasso using P-splines (gALassoP) and the NonNegative Garrote (NNG)) are proposed.

The number of covariates p is assumed to be finite in Chapter 4 and allowed to grow slower than the sample size. However, due to an advancement in technology to identify and store a huge amount of information, there is a high demand in applications for ultahigh-dimensional models (when the number of coefficients to be estimated is much bigger than the number of observations). In Chapter 5, a two-stage approach (NNG_{SIS}) is proposed. This approach is illustrated on the UK Employement data, where in addition to the three original covariates several hundreds of artificial ones are generated and added to the data.

In Chapter 2, we consider an exponential varying coefficient type of variability function to be the full model. Alternatively, one could also consider a single-index type of modeling (e.g. Wu et al., 2010; Lian et al., 2015) where the variability function is $V(\mathbf{X}(T), T) = g(\mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\theta}(T))$, with an unknown link function $g(\cdot) \geq 0$. To estimate the variability function we would then need to go through two steps: in the first step the single index $\mathbf{X}^{\mathrm{T}}(T)\boldsymbol{\theta}(T)$ would be estimated with g assumed to be approximatively linear, and in the second step the unknown link function is estimated. This type of variability function can be considered when one wants to be sure in identifying the correct heteroscedastic model for the data at hand.

Another type of hypothesis testing (for the shape of the functional coefficients) to consider in the context of the varying coefficient models is the change point detection or piecewise testing. Such hypothesis is important for the functional coefficient $\beta_1(T)$ in the Intego data, which is varying in the interval around year 2010 and everywhere else it is constant up to a shift (Figure 4.8 (a)). In mean regression, Lebarbier (2005) and Kolar et al. (2009), among others, propose how to estimate the jump locations of

a piecewise constant function. Several authors have been focusing on linear quantile regression models, for instance, Zhang et al. (2014) and Lee et al. (2017). Using a score-type statistic, Zhang et al. (2014) propose a testing procedure to check for the existence of a threshold effect (piecewise constant). Lee et al. (2017) develop L_1-penalized estimators of both the sparse regression coefficients and the parameter for the change point, in an ultrahigh-dimensional setting. It would be interesting to study such hypothesis in quantile varying coefficient models.

In this dissertation, we focus on a longitudinal data setting. Quantile regression has recently emerged as a useful alternative tool to the classical Cox model for analyzing censored data (Ying et al., 1995; McKeague et al., 2001; Portnoy, 2003; Wang and Wang, 2009; De Backer et al., 2018, to name a few). Since quantile regression can be used to directly model the survival time (e.g. time to cure from a disease), it relaxes the proportional hazards assumption of the Cox model, and it is easy to interpret. When the survival time is more prone to random right censoring (e.g. the patients leave the study before being cured), it is more appropriate to consider the lower quantiles than the mean of the survival time. He et al. (2013) consider feature screening for an ultrahigh-dimensional varying coefficient model for a response which is subject to random right censoring. They investigate the diffuse large-B-cell lymphoma microarray data that contains the survival times and the gene expression measurements of 7399 genes for 240 patients, where nearly half of the survival time data are censored. It would be interesting to adapt the approach of He et al. (2013) to flexible screening with the P-splines technique. Further, considering the NNG and NNG_{SIS}, respectively, in Chapters 4 and 5, for this type of data would be of interest. Checking the structure of the variability function (to see if only investigating the median survival time suffices or not) as well as testing for the shape of the functional coefficients is also nice to consider.

It is interesting to consider quantile regression to investigate spatial data, which arise in various research areas including econometrics, epidemiology, environmental science and image analysis. For instance when investigating the ground level ozone concentration, which often shows high degrees of heterogeneity across different regions, studying several quantiles might be more appropriate than investigating only the mean. Further, adding spatial smoothing to this, captures the spatial trend (e.g. of the ozone concentration) and provides a way to extrapolate our estimators to the location at which observations are not available. Assuming that the quantile function is linear in time and smooth over the space, Reich (2012) and Das and Ghosal (2017) study spatio-temporal quantile varying coefficient models using Bayesian anal-

ysis. For mean regression, assuming that the observations are randomly distributed over two-dimensional domains in space, Mu et al. (2018) consider bivariate splines over triangulations to represent the coefficient functions in spatially varying coefficient models. The authors also propose a penalized approach to balance the goodness of fit and smoothing. They also construct a nonstationary bootstrap test to check whether some of the functional coefficients vary over the space or not. Extending the results of Mu et al. (2018) to quantile regression, might therefore be very useful when investigating spatial data.

Last but not least, implementing our proposed methodologies to another type of flexible models like additive models, where the functional coefficients are allowed to vary with some different variables, might be an interesting topic to consider. One example is the Prostate Cancer Data considered in Antoniadis et al. (2012a), where the interest is in studying the correlation between the level of prostate-specific antigen and a number of clinical measures. For such data, we might be interested to investigate if each clinical measure has a functional relation with the prostate-specific antigen. Antoniadis et al. (2012a) have shown, using mean regression, that some of the clinical measures, like the cancer volume and the prostate weight, have varying association with the prostate-specific antigen.

Bibliography

Ahkim, M., Gijbels, I., and Verhasselt, A. (2017). Shape testing in varying coefficient models. *TEST*, 26(2):429–450.

Ahkim, M. and Verhasselt, A. (2018). Testing for constancy in varying coefficient models. *Communications in Statistics-Theory and Methods*, 47(4):890–911.

Andriyana, Y. (2015). *P-splines quantile regression in varying coefficient models*. PhD Dissertation KU Leuven, Belgium. ISBN 978-90-8649-791-1.

Andriyana, Y. and Gijbels, I. (2017). Quantile regression in heteroscedastic varying coefficient models. *AStA Advances in Statistical Analysis*, 101(2):151–176.

Andriyana, Y., Gijbels, I., and Verhasselt, A. (2014). P-splines quantile regression estimation in varying coefficient models. *TEST*, 23(1):153–194.

Andriyana, Y., Gijbels, I., and Verhasselt, A. (2016). Quantile regression in varying-coefficient models: non-crossing quantile curves and heteroscedasticity. *Statistical Papers*, DOI: 10.1007/s00362-016-0847-7. Available from: `http://link.springer.com/article/10.1007/s00362-016-0847-7`.

Andriyana, Y., Ibrahim, M. A., Gijbels, I., and Verhasselt, A. (2018). *QRegVCM: Quantile Regression in Varying-Coefficient Models*. R package version 1.2.

Antoniadis, A., Gijbels, I., and Lambert-Lacroix, S. (2014). Penalized estimation in additive varying coefficient models using grouped regularization. *Statistical Papers*, 55(3):727–750.

Antoniadis, A., Gijbels, I., and Verhasselt, A. (2012a). Variable Selection in Additive Models Using P-Splines. *Technometrics*, 54(4):425–438.

Antoniadis, A., Gijbels, I., and Verhasselt, A. (2012b). Variable Selection in Varying-Coefficient Models Using P-Splines. *Journal of Computational and Graphical Statistics*, 21(3):638–661.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277–297.

Belloni, A. and Chernozhukov, V. (2011). l_1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, 39(1):82–130.

Bollaerts, K., Eilers, P. H., and Aerts, M. (2006). Quantile regression with monotonicity restrictions using P-splines and the L1-norm. *Statistical Modelling*, 6(3):189–207.

Bowman, A., Jones, M., and Gijbels, I. (1998). Testing Monotonicity of Regression. *Journal of Computational and Graphical Statistics*, 7(4):489–500.

Bradic, J., Fan, J., and Wang, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *Journal of the Royal Statistical Society, Series B*, 73(3):325–349.

Breiman, L. (1995). Better Subset Regression Using the Nonnegative Garrote. *Technometrics*, 37(4):373–384.

Bühlmann, P. (2006). Boosting for high-dimensional linear models. *The Annals of Statistics*, 34(2):559–583.

Cai, Z. and Xu, X. (2008). Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models. *Journal of the American Statistical Association*, 103(484):1595–1608.

Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351.

Chaudhuri, P. (1991). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, 39(2):246–269.

Chen, Y., Bai, Y., and Fung, W. (2017). Structural identification and variable selection in high-dimensional varying-coefficient models. *Journal of Nonparametric Statistics*, 29(2):258–279.

Cheng, M.-Y., Honda, T., and Zhang, J.-T. (2016). Forward Variable Selection for Sparse Ultra-High Dimensional Varying Coefficient Models. *Journal of the American Statistical Association*, 111(515):1209–1221.

Claeskens, G., Krivobokova, T., and Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika*, 96(3):529–544.

Croissant, Y. and Millo, G. (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software*, 27(2):1–43. Available from: `http://www.jstatsoft.org/v27/i02/`.

Das, P. and Ghosal, S. (2017). Analyzing ozone concentration by Bayesian spatio-temporal quantile regression. *Environmetrics*, 28(4):e2443–n/a.

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data. Monographs on statistics and applied probability 62*. Chapman & Hall/CRC, New York.

De Backer, M., Ghouch, A. E., and Van Keilegom, I. (2018). An Adapted Loss Function for Censored Quantile Regression. *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2018.1469996. Available from: `https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1469996`.

de Boor, C. (2001). *A Practical Guide to Splines, revised Edition. Applied Mathematical Sciences 27*. Springer-Verlag, New York.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121.

Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society, Series B*, 70(5):849–911.

Feng, X. and Zhu, L. (2016). Estimation and Testing of Varying Coefficients in Quantile Regression. *Journal of the American Statistical Association*, 111(513):266–274.

Fenske, N., Kneib, T., and Hothorn, T. (2011). Identifying Risk Factors for Severe Childhood Malnutrition by Boosting Additive Quantile Regression. *Journal of the American Statistical Association*, 106(494):494–510.

Fortin, M., Daigle, G., Ung, C.-H., Bégin, J., and Archambault, L. (2007). A variance-covariance structure to take into account repeated measurements and heteroscedasticity in growth modeling. *European Journal of Forest Research*, 126(4):573–585.

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

Ghosal, S., Sen, A., and van der Vaart, A. W. (2000). Testing monotonicity of regression. *The Annals of statistics*, 28(4):1054–1082.

Gijbels, I., Ibrahim, M. A., and Verhasselt, A. (2017). Shape testing in quantile varying coefficient models with heteroscedastic error. *Journal of Nonparametric Statistics*, 29(2):391–406.

Gijbels, I., Ibrahim, M. A., and Verhasselt, A. (2018). Testing the heteroscedastic error structure in quantile varying coefficient models. *The Canadian Journal of Statistics*, 46(2):246–264.

Gijbels, I. and Verhasselt, A. (2010). P-splines regression smoothing and difference type of penalty. *Statistics and Computing*, 20(4):499–511.

Gijbels, I., Verhasselt, A., and Vrinssen, I. (2015). Variable selection using P-splines. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(1):1–20.

Grant, M. C. and Boyd, S. P. (2012). *The CVX Users' Guide*. CVX Research, Inc.

Grunfeld, Y. (1958). *The determinant of corporate investment*. PhD Dissertation Department of Economic, University of Chicago, USA.

Guo, C., Yang, H., and Lv, J. (2017). Robust variable selection in high-dimensional varying coefficient models based on weighted composite quantile regression. *Statistical Papers*, 58(4):1009–1033.

Hastie, T. and Tibshirani, R. (1993). Varying-Coefficient Models. *Journal of the Royal Statistical Society, Series B*, 55(4):757–796.

He, X. (1997). Quantile Curves without Crossing. *The American Statistician*, 51(2):186–192.

He, X. and Ng, P. (1999). COBS: qualitatively constrained smoothing via linear programming. *Computational Statistics*, 14(3):315–337.

He, X., Ng, P., and Portnoy, S. (1998). Bivariate quantile smoothing splines. *Journal of the Royal Statistical Society, Series B*, 60(3):537–550.

He, X. and Shi, P. (1994). Convergence rate of b-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics*, 3(3-4):299–308.

He, X. and Shi, P. (1998). Monotone B-Spline Smoothing. *Journal of the American Statistical Association*, 93(442):643–650.

He, X., Wang, L., and Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics*, 41(1):342–369.

Honda, T. (2004). Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inference*, 121(1):113–125.

Honda, T., Ing, C.-K., and Wu, W.-Y. (2017). Adaptively weighted group lasso for semiparametric quantile regression models. Discussion Papers 2017-04, Graduate School of Economics, Hitotsubashi University. Available from: `https://EconPapers.repec.org/RePEc:hit:econdp:2017-04`.

Huang, J., Horowitz, J. L., and Wei, F. (2010). Variable selection in nonparametric additive models. *The Annals of statistics*, 38(4):2282–2313.

Huang, J., Wei, F., and Ma, S. (2012). Semiparametric Regression Pursuit. *Statistica Sinica*, 22(4):1403–1426.

Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128.

Ibrahim, M. A. and Verhasselt, A. (2018a). A two-stage screening and variable selection method in ultrahigh-dimensional quantile varying coefficient models with heteroscedastic error. *Submitted*.

Ibrahim, M. A. and Verhasselt, A. (2018b). Variable selection in quantile varying coefficient models with heteroscedastic error. *Submitted*.

Jiang, X., Jiang, J., and Song, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica*, 22(4):1479–1506.

Kai, B., Li, R., and Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *The Annals of Statistics*, 39(1):305–332.

Karlsson, M. and Lindmark, A. (2014). truncSP: An R Package for Estimation of Semi-Parametric Truncated Linear Regression Models. *Journal of Statistical Software*, 57(14):1–19.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., and Rinaldo, C. R. (1987). The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology*, 126(2):310–318.

Kim, M.-O. (2006). Quantile Regression with Shape-Constrained Varying Coefficients. *Sankhyā: The Indian Journal of Statistics*, 68(3):369–391.

Kim, M.-O. (2007). Quantile regression with varying coefficients. *The Annals of Statistics*, 35(1):92–108.

Kleiber, C. and Zeileis, A. (2008). *Applied Econometrics with R*. Springer-Verlag, New York. Available from: `https://CRAN.R-project.org/package=AER`.

Knight, K. (1999). Asymptotics for L_1-estimators of regression parameters under heteroscedasticity. *The Canadian Journal of Statistics*, 27(3):497–507.

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

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

Koenker, R. (2011). Additive models for quantile regression: Model selection and confidence bandaids. *Brazilian Journal of Probability and Statistics*, 25(3):239–262.

Koenker, R. and Bassett, Jr., G. (1978). Regression Quantiles. *Econometrica*, 46(1):33–50.

Koenker, R. and Ng, P. (2003). SparseM: A Sparse Matrix Package for R *. *Journal of Statistical Software*, 8(6):1–9.

Koenker, R. and Ng, P. (2005). A Frisch-Newton Algorithm for Sparse Quantile Regression. *Acta Mathematicae Applicatae Sinica, English Series*, 21(2):225–236.

Koenker, R., Ng, P., and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, 81(4):673–680.

Koenker, R., Portnoy, S., Ng, P. T., Zeileis, A., Grosjean, P., and Ripley, B. D. (2017). *quantreg: Quantile Regression*. R package version 5.33.

Kolar, M., Song, L., and Xing, E. P. (2009). Sparsistent Learning of Varying-coefficient Models with Structural Changes. *Proceedings of theAdvances in Neural Information Processing Systems*, 22:1006–1014.

Lebarbier, É. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing*, 85(4):717–736.

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., Liao, Y., Seo, M. H., and Shin, Y. (2017). Oracle Estimation of a Change Point in High Dimensional Quantile Regression. *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2017.1319840. Available from: `http://dx.doi.org/10.1002/cjs.11346`.

Li, J., Li, Y., and Zhang, R. (2017). B spline variable selection for the single index models. *Statistical Papers*, 58(3):691–706.

Li, N., Xu, X., and Liu, X. (2011). Testing the constancy in varying-coefficient regression models. *Metrika*, 74(3):409–438.

Li, Y. and Zhu, J. (2008). L_1-norm quantile regression. *Journal of Computational and Graphical Statistics*, 17(1):163–185.

Lian, H., Liang, H., and Carroll, R. J. (2015). Variance function partially linear single-index models. *Journal of the Royal Statistical Society, Series B*, 77(1):171–194.

Lin, C.-Y., Bondell, H., Zhang, H. H., and Zou, H. (2013). Variable selection for non-parametric quantile regression via smoothing spline analysis of variance. *Stat*, 2(1):255–268.

Ma, S., Song, Q., Wang, L., et al. (2013). Simultaneous variable selection and estimation in semiparametric modeling of longitudinal/clustered data. *Bernoulli*, 19(1):252–274.

McKeague, I. W., Subramanian, S., and Sun, Y. (2001). Median regression and the missing information principle. *Journal of Nonparametric Statistics*, 13(5):709–727.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics*, 34(3):1436–1462.

Mu, J., Wang, G., and Wang, L. (2018). Estimation and inference in spatially varying coefficient models. *Environmetrics*, 29(1):e2485–n/a.

Murnane, R. J., Willett, J. B., and Boudett, K. P. (1999). Do male dropouts benefit from obtaining a GED, postsecondary education, and training? *Evaluation Review*, 23(5):475–503.

Noh, H., Chung, K., and Van Keilegom, I. (2012). Variable selection of varying coefficient models in quantile regression. *Electronic Journal of Statistics*, 6:1220–1238.

Noh, H. S. and Park, B. U. (2010). Sparse varying coefficient models for longitudinal data. *Statistica Sinica*, 20(3):1183–1202.

Park, S., He, X., and Zhou, S. (2017). Dantzig-type penalization for multiple quantile regression with high dimensional covariates. *Statistica Sinica*, 27(4):1619–1638.

Portnoy, S. (2003). Censored Regression Quantiles. *Journal of the American Statistical Association*, 98(464):1001–1012.

Portnoy, S. and Koenker, R. (1997). The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators. *Statistical Science*, 12(4):279–300.

Reich, B. J. (2012). Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *Journal of the Royal Statistical Society, Series C*, 61(4):535–553.

Reich, B. J., Bondell, H. D., and Wang, H. J. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, 11(2):337–352.

Schnabel, S. K. and Eilers, P. H. (2013). Simultaneous estimation of quantile curves using quantile sheets. *AStA Advances in Statistical Analysis*, 97(1):77–87.

Schumaker, L. (1981). *Spline Functions: Basic Theory*. Wiley, New York.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Stone, C. J. (1982). Optimal Global Rates of Convergence for Nonparametric Regression. *The Annals of Statistics*, 10(4):1040–1053.

Tang, Y., Song, X., Wang, H. J., and Zhu, Z. (2013a). Variable selection in high-dimensional quantile varying coefficient models. *Journal of Multivariate Analysis*, 122:115–132.

Tang, Y., Wang, H. J., and Zhu, Z. (2013b). Variable selection in quantile varying coefficient models with longitudinal data. *Computational Statistics & Data Analysis*, 57(1):435–449.

Tang, Y., Wang, H. J., Zhu, Z., and Song, X. (2012). A unified variable selection approach for varying coefficient models. *Statistica Sinica*, 22(2):601–628.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288.

Vaes, B., Beke, E., Truyers, C., Elli, S., Buntinx, F., Verbakel, J. Y., Goderis, G., and Van Pottelbergh, G. (2015). The correlation between blood pressure and kidney function decline in older people: a registry-based cohort study. *BMJ open*, 5(6):e007571.

Van Keilegom, I. and Wang, L. (2010). Semiparametric modeling and estimation of heteroscedasticity in regression analysis of cross-sectional data. *Electronic Journal of Statistics*, 4:133–160.

Wang, H. (2009). Forward Regression for Ultra-High Dimensional Variable Screening. *Journal of the American Statistical Association*, 104(488):1512–1524.

Wang, H. and Xia, Y. (2009). Shrinkage Estimation of the Varying Coefficient Model. *Journal of the American Statistical Association*, 104(486):747–757.

Wang, H. J. and Wang, L. (2009). Locally Weighted Censored Quantile Regression. *Journal of the American Statistical Association*, 104(487):1117–1128.

Wang, H. J., Zhu, Z., and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *The Annals of Statistics*, 37(6B):3841–3866.

Wang, J. C. and Meyer, M. C. (2011). Testing the monotonicity or convexity of a function using regression splines. *The Canadian Journal of Statistics*, 39(1):89–107.

Wang, K. and Lin, L. (2014). Variable selection in robust semiparametric modeling for longitudinal data. *Journal of the Korean Statistical Society*, 43(2):303–314.

Wang, K. and Sun, X. (2017). Efficient parameter estimation and variable selection in partial linear varying coefficient quantile regression model with longitudinal data. *Statistical Papers*, DOI: 10.1007/s00362-017-0970-0. Available from: `https://doi.org/10.1007/s00362-017-0970-0`.

Wang, L., Wu, Y., and Li, R. (2012). Quantile Regression for Analyzing Heterogeneity in Ultra-high Dimension. *Journal of the American Statistical Association*, 107(497):214–222.

Wang, L., Xue, L., Qu, A., and Liang, H. (2014). Estimation and model selection in generalized additive partial linear models for correlated data with diverging number of covariates. *The Annals of Statistics*, 42(2):592–624.

Wei, F., Huang, J., and Li, H. (2011). Variable selection and estimation in high-dimensional varying-coefficient models. *Statistica Sinica*, 21(4):1515–1540.

Winter, H., Adelhardt, S., Jerak, A., and Küchenhoff, H. (2002). Characterization of cataclastic shear-zones of the KTB deep drill hole by regression analysis of drill cuttings data. *Geophysical Journal International*, 150(1):1–9.

Wu, T. Z., Yu, K., and Yu, Y. (2010). Single-index quantile regression. *Journal of Multivariate Analysis*, 101(7):1607–1621.

Wu, Y. and Liu, Y. (2009a). Stepwise multiple quantile regression estimation using non-crossing constraints. *Statistics and its Interface*, 2(3):299–310.

Wu, Y. and Liu, Y. (2009b). Variable selection in quantile regression. *Statistica Sinica*, 19(2):801–817.

Xue, L. (2009). Consistent variable selection in additive models. *Statistica Sinica*, 19(3):1281–1296.

Xue, L. and Qu, A. (2012). Variable Selection in High-dimensional Varying-coefficient Models with Global Optimality. *Journal of Machine Learning Research*, 13(1):1973–1998.

Yang, H., Guo, C., and Lv, J. (2016). Variable selection for generalized varying coefficient models with longitudinal data. *Statistical Papers*, 57(1):115–132.

Ying, Z., Jung, S.-H., and Wei, L.-J. (1995). Survival Analysis with Median Regression Models. *Journal of the American Statistical Association*, 90(429):178–184.

Zhang, H.-G. and Mei, C.-L. (2012). SiZer Inference for varying coefficient models. *Communications in Statistics-Simulation and Computation*, 41(10):1944–1959.

Zhang, H.-G., Mei, C.-L., and Wang, H.-L. (2013). Robust SiZer Approach for Varying Coefficient Models. *Mathematical Problems in Engineering*, 2013(Article ID 547874):DOI: 10.1155/2013/547874. Available from: `http://dx.doi.org/10.1155/2013/547874`.

Zhang, L., Wang, H. J., and Zhu, Z. (2014). Testing for change points due to a covariate threshold in quantile regression. *Statistica Sinica*, 24:1859–1877.

Zhao, W., Zhang, R., Lv, Y., and Liu, J. (2013). Variable selection of the quantile varying coefficient regression models. *Journal of the Korean Statistical Society*, 42(3):343–358.

Zhao, Z. and Xiao, Z. (2014). Efficient regressions via optimally combining quantile information. *Econometric theory*, 30(6):1272–1314.

Zheng, Q., Gallagher, C., and Kulasekera, K. (2013). Adaptive penalized quantile regression for high dimensional data. *Journal of Statistical Planning and Inference*, 143(6):1029–1038.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

Zou, H. and Li, R. (2008). One-Step Sparse Estimates in Nonconcave Penalized Likelihood Models. *The Annals of Statistics*, 36(4):1509–1533.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36(3):1108–1126.

Zou, H. and Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *The Annals of statistics*, 37(4):1733–1751.

Appendix A

R-code used to analyze the data examples

```
############################################################
### The data example in Section 2.4.1 ###
############################################################
library(QRegVCM)
data(CD4)
subj = CD4$subj ## subject indicator (a man)
dim = length(subj) ## number of rows in the data = 1817
y = CD4$CD4 ## the CD4 percentage
X0 = rep(1,dim) ## the intercept
X1 = CD4$Smooking ## the smoking status
X2 = CD4$Age ## age at HIV infection
X3 = CD4$PreCD4 ## the pre-infection CD4 percentage
times = CD4$Time ## the time in years
X = cbind(X0, X1, X2, X3) ## the covariate matrix
px=ncol(X)

lambdas = 10^seq(-2, 1, 0.1)
kn = rep(10,px) ## The number of knot intervals for each covariate.
degree = rep(3,px) ## the degree of splines
d = rep(1,px) ## the differencing order in the penalty term
## for each covariate
```

```
gam=0.25  ## the power used in estimating the smoothing parameter
## for each covariate (the \eta_P)
nr.bootstrap.samples=200 ## number of bootstrap samples
seed=110 ## the seed for the random generator in the bootstrap
## resampling
taus = seq(0.1,0.9,0.2) ## the quantiles of interest

test2=test_variability(times=times, subj=subj, X=X, y=y, d=d,
kn=kn, degree=degree, lambda=lambdas, gam=gam,
nr.bootstrap.samples=nr.bootstrap.samples,seed,
tau=taus,test="Y",model=4)

### test results
test2$comp ## the comparisons made
test2$P  ## p-values
test2$GR ## test statistics
View(test2$Gb) ## bootstrap test statistics

### estimation results
## The variability models
qhat5_s2_m4=test2$qhat5_s2_m4 # Model 4
qhat5_s2_m5=test2$qhat5_s2_m5 # Model 5
qhat5_s2_m0=test2$qhat5_s2_m0*rep(1,dim) # Model 0
## The estimated coefficients for Model 4
gamma0=test2$hat_btV_4[1:dim]
gamma1=test2$hat_btV_4[(dim+1):(dim*2)]
gamma2=test2$hat_btV_4[(dim*2+1):(dim*3)]
gamma3=test2$hat_btV_4[(dim*3+1):(dim*4)]



##########################################################
### The data example in Section 3.3 ###
##########################################################
library(QRegVCM)
data(wages)
y = wages$resp ## the hourly wage
times = wages$exper ## the duration of work experience in years
```

```
subj = wages$id ## subject indicator (individual)
dim=length(y) ## number of rows in the data = 6402
x0 = rep(1,dim) ## for intercept
### the covariates
x1 = wages$r1 # the dummy variable for black individuals
x2 = wages$r2 # the dummy variable for hispanic individuals
x3 = wages$hgc ## the highest grade completed by the indiviadual
X = cbind(x0, x1, x2, x3) ## the covariate matrix
px=ncol(X)

##########################
### Input parameters ####
##########################
lambda = 10^seq(-2, 1, 0.1) # the smoothing parameter
kn = rep(5,px) # The number of knot intervals for each covariate.
degree = rep(2,px) # the degree of splines
d = rep(1,px)
gam=0.25
nr.bootstrap.samples=200
seed=110
########################################
## Testing for the constancy of \beta_0
test1=simul_shapetest(times=times, subj=subj, X=X, y=y, d=d, kn=kn,
degree=degree, lambda=lambda, gam=gam, v=1,
nr.bootstrap.samples=nr.bootstrap.samples,seed=seed,
test=c("c",NA),omega=10^3)
#### Testing results
test1$result  #the testing procedures
test1$P  ## p-values
test1$R ## test statistics


#########################################################
### The data example in Section 4.3.1 ###
#########################################################
# First run the functions in the R-code available at
# https://ibiostat.be/online-resources/longitudinal
```

```
library(splines)
library(SparseM)
library(quantreg)
library(mvtnorm)
library(AER)
data(Grunfeld)

subj = Grunfeld$firm ## The firm indicator
dim = length(subj)
Y = Grunfeld$invest ## gross investment
x0 = rep(1,dim)
x1 = Grunfeld$value ## the market value of the firm
x2 = Grunfeld$capital ## the capital of the firm
times = Grunfeld$year ## the time in year
X = cbind(x0, x1, x2)

tolerance=1e-06; degree=3;
kn=4 # The number of internal knot for each covariate.
lambdasp = c(0.01,10^seq(-1, 2, 0.3))
lambdas = (1:23)*0.01 # regularization parameter

test1=variable_selection(times, subj, X, Y, d=1, kn, degree,
lambda=lambdas, lambdap=lambdasp, eta_p=1, eta_l=2,
tolerance=1e-06,variability="Y", technique="galassoP")
test1$result # The indicator of which covariates are selected
# for the median.
test1$resultV # The indicator of which covariates are selected
# for the variability.


##########################################################
### The data example in Section 5.5 ###
##########################################################
# First run the functions in the R-code available at
# https://ibiostat.be/online-resources/longitudinal
library(splines)
library(SparseM)
```

```
library(quantreg)
library(mvtnorm)
library(plm)
data(EmplUK)
subj = EmplUK$firm ## the firm indicator
times = EmplUK$year ## the time in year
dim = length(subj)
Y = log(EmplUK$emp) ## logarith of number employees
x0 = rep(1,dim)
x1 = EmplUK$wage ## average annual wage per employee
x2 = EmplUK$capital ## the caital of the company
x3 = EmplUK$output ## an index of value-added output

#### p-3=47 redundant covariates
nt=9
sigma = matrix(1, nrow=nt, ncol=nt)
for(i in 1:nt)
{
for(j in 1:nt)
{
sigma[i,j] = 4*exp(-abs(i-j))
}
}

ip=47
xr = NULL
N = rep(0,length(unique(subj)))
for(i in 1:length(unique(subj))){
N[i] = length(subj[which(subj==unique(subj)[i])])
xr = rbind(xr,t(rmvnorm(ip, mean=rep(0,N[i]),
sigma=sigma[1:N[i],1:N[i]])))
}
X = cbind(x0, x1, x2, x3,xr)

tolerance=1e-06; degree=3;
kn=3 # The number of internal knot for each covariate.
lambdasp = c(0.01,10^seq(0, 1.8, 0.6))
```

```
lambdas = (seq(1,30,by=3))*0.01 # regularization parameter

test1=variable_selection(times, subj, X, Y, d=1, kn, degree,
lambda=lambdas,lambdap=lambdasp, eta_p=1, eta_l=2, tolerance=1e-06,
variability="Y",technique="galassoB_lasso")

sum(test1$result[-(1:4),2]) ## number irrelevant variables in median
sum(test1$resultV[-(1:4),2]) ## number irrelevant variables in
# variability
```