

Quantile regression in varying-coefficient models: non-crossing quantile curves and heteroscedasticity

Peer-reviewed author version

ANDRIYANA, Yudhie; Gijbels, Irène & VERHASSELT, Anneleen (2018) Quantile regression in varying-coefficient models: non-crossing quantile curves and heteroscedasticity. In: STATISTICAL PAPERS, 59 (4), p. 1589-1621.

DOI: 10.1007/s00362-016-0847-7

Handle: <http://hdl.handle.net/1942/23408>

Quantile regression in varying-coefficient models: non-crossing quantile curves and heteroscedasticity

Y. Andriyana^{1,3}, I. Gijbels¹ and A. Verhasselt².

1. KU Leuven, Department of Mathematics and Leuven Statistics Research Center (LStat), Leuven, Belgium.
2. Universiteit Hasselt, Interuniversity Institute for Biostatistics and statistical Bioinformatics, CenStat, Hasselt, Belgium.
3. Statistics Department, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung - Indonesia.

October 15, 2016

Abstract

Quantile regression is an important tool for describing the characteristics of conditional distributions. Population conditional quantile functions cannot cross for different quantile orders. Unfortunately estimated regression quantile curves often violate this and cross each other, which can be very annoying for interpretations and further analysis. In this paper we are concerned with flexible varying-coefficient modelling, and develop methods for quantile regression that ensure that the estimated quantile curves do not cross. A second aim of the paper is to allow for some heteroscedasticity in the error modelling, and to also estimate the associated variability function. We investigate the finite-sample performances of the discussed methods via simulation studies. Some applications to real data illustrate the use of the methods in practical settings.

Keywords and phrases: B-splines, crossing quantile curves, longitudinal data, P-splines, quantile regression, quantile sheet, variability, varying-coefficient models.

MSC classification: 62Gxx; 62Jxx; 62Hxx.

1 Introduction

A (unconditional) quantile function is an increasing function in its argument, say τ (with $0 \leq \tau \leq 1$). In real applications, the impact of explanatory variables on a variable of interest, leads to the study of conditional quantile functions or (for short) regression quantiles. For a given value of τ , a conditional quantile function is thus a function of a covariate (or several covariates). In practice, the conditional quantile functions are estimated, from data, for various fixed values τ , i.e. we estimate for example the conditional median

function, but also the conditional first and third quartile functions. These conditional quantile functions are helpful to further describe the impact of covariates on the response variable. See Koenker (2005) for a general reference.

Conditional quantile functions are by definition, for any given fixed value(s) of the covariate(s), an increasing function in the argument τ . When estimating separately several regression quantiles based on a sample at hand, there is however no guarantee that the estimated regression quantiles are not crossing. This is more likely to happen in small to moderate samples, and is annoying for further analysis and interpretations. There is thus an interest to prevent this crossing to happen in finite-samples. In the literature one can find some methods to deal with this issue.

A first approach was given by Koenker (1984) in a linear regression model by considering parallel quantile planes. Cole (1988) and Cole and Green (1992) proposed a more general algorithm. They combined the Box-Cox power transformation (L), the mean or median function (M), and the coefficient of variation (S), leading all together to the LMS algorithm. However, this implicitly assumes that the L-step leads to normally distributed random variables. Moreover the choice of transformation is sensitive to outliers in the data. To remedy for these drawbacks, He (1997) proposed an algorithm such that the monotonicity in τ at all covariates is guaranteed, without needing to impose a normality assumption on the error structure. He's approach actually allows for a heteroscedastic modelling, but is also applicable to a homoscedastic regression model (being a special case). The use of He's algorithm in homoscedastic and heteroscedastic models was also investigated in the simulation studies of Wu and Liu (2009). They propose a stepwise procedure to ensure non-crossing estimated curves for both linear and non-linear quantile regression. They show that their algorithm improves the estimation accuracy although their estimation scheme costs more time to achieve the non-crossing of the estimated curves. All the former approaches are based on non-simultaneous estimation, meaning that they use single quantile objective functions. Schnabel and Eilers (2013b) considered a location-scale model for obtaining non-crossing expectile curves.

An alternative way is to use simultaneous objective functions to impose the non-crossing restriction, as was discussed by Bondell *et al.* (2010) and Liu and Wu (2011). The latter paper also deals with variable selection. Using the objective functions in a simultaneous way might share strengths among the regression quantiles. Consequently, it can improve the estimation accuracy compared to estimation based on individual objective functions. See Zou and Yuan (2008), Jiang *et al.* (2016), and Yang and Liu (2016), among others, for recent contributions in composite quantile regression methods.

Yet another, indirect approach is based on estimating conditional cumulative distribution functions first, and then pass to estimation of quantile curves via an inverting operation. Non-crossing of curves is enforced on the level of the estimation of the conditional cumulative distribution function. See for example Dette and Volgushev (2008) and

Chernozhukov *et al.* (2009).

To describe accurately the stochastic behaviour behind complex data, one often considers regression models that are on the one hand flexible enough to capture this complexity, but on the other hand still allow for estimation methods with good practical performance. In this paper we focus on varying-coefficient type of modelling, which naturally extends linear regression models by allowing the regression coefficients to change with another covariate. Since their introduction by Hastie and Tibshirani (1993), these models have been studied intensively in the literature, in particular with respect to estimating the mean regression function. They are very useful tools in statistical data analysis nowadays.

Quantile estimation in varying-coefficient models has been studied in Honda (2004), Kim (2007), Wang *et al.* (2009), among others. Andriyana *et al.* (2014) used the flexible P-splines estimation technique for quantile regression in varying-coefficient models. They established the asymptotic distributional behaviour of the estimators, discussed linear (or more generally convex) programming algorithms for solving the optimization problems, and provided data-driven choices of regularization parameters. The aim of the current paper is twofold: (i) to develop methods to prevent for crossing estimated quantile curves in a setting of varying-coefficient modelling; (ii) to allow for heteroscedasticity and to discuss estimation of the scaling/variability function, together with the quantile curves.

The paper is organized as follows. In Section 2 we introduce the modelling framework, and recall briefly the P-splines estimation technique for individual quantile estimation. We also detail the issue of non-crossing curves and discuss the distinction between homoscedastic and heteroscedastic modelling aspects. In Section 3 several methods for preventing for crossing estimated quantile curves in the varying-coefficient model setting are discussed. The optimization problems are solved using linear programming methods. Details of these can be obtained from the authors. Section 4 focuses on the estimation of the variability function in a heteroscedastic quantile regression varying-coefficient model. The finite-sample performances of the methods for the estimation of the quantile curves, preventing curves to cross, and for estimation of the variability function are investigated in simulation studies in Sections 5 and 6. The use of the developed methods is illustrated on a real data example in Section 7. The paper is concluded with Section 8. Additional theoretical considerations, some practical implementation issues, simulation results and an additional real data application are provided in the Supplementary material accompanying this paper.

2 P-splines quantile regression in varying-coefficient models

2.1 Conditional quantiles

Suppose that a variable of interest Y is influenced by p covariates $X^{(1)}, \dots, X^{(p)}$, placed in the column vector $\mathbf{X}^* = (X^{(1)}, \dots, X^{(p)})'$ of dimension p (the superscript $'$ refers to the transposed of a vector or matrix). Denoting the conditional distribution function of Y given \mathbf{X}^* by $F_{Y|\mathbf{X}^*}$, the conditional quantile of Y given \mathbf{X}^* of order τ (with $0 \leq \tau \leq 1$) is given by

$$q_\tau(Y | \mathbf{X}^*) = \inf \{u : F_{Y|\mathbf{X}^*}(u | \mathbf{X}^*) \geq \tau\} .$$

By definition it follows that, for all values of \mathbf{X}^* ,

$$\text{for } 0 \leq \tau_1 < \tau_2 \leq 1 : \quad q_{\tau_1}(Y | \mathbf{X}^*) \leq q_{\tau_2}(Y | \mathbf{X}^*) . \quad (2.1)$$

The conditional quantile curves, looked upon as functions of \mathbf{X}^* , do not cross each other, or $q_\tau(Y | \mathbf{X}^*)$ is monotone increasing in the argument τ .

In this paper we are particularly interested in a longitudinal data setting in which we consider having repeated observations on $(Y(T), (\mathbf{X}^*(T))', T)$ where T denotes a variable (with domain \mathcal{T}), $Y(T)$ is the measurement of the variable of interest at ‘time’ T , and $\mathbf{X}^*(T)$ is the vector of covariates at time T . In this setting the conditional τ -th order quantile of $Y(T)$ conditionally upon $(\mathbf{X}^*(T), T)$ is denoted by $q_\tau(Y(T) | \mathbf{X}^*(T), T)$, for which (2.1) now holds for all values of $(\mathbf{X}^*(T), T)$. See further Section 2.2.

Direct measures of spread can be obtained from the conditional quantiles by considering, for example, the median absolute deviation, defined as the conditional median $\text{median}(|Y(T) - q_{0.5}(Y(T) | \mathbf{X}^*(T), T)| | \mathbf{X}^*(T), T)$, or the conditional interquartile range $q_{0.75}(Y(T) | \mathbf{X}^*(T), T) - q_{0.25}(Y(T) | \mathbf{X}^*(T), T)$.

An approach towards estimation in general of conditional quantiles, such as $q_\tau(Y | \mathbf{X}^*)$ or $q_\tau(Y(T) | \mathbf{X}^*(T), T)$, is to first estimate the conditional distribution function $F_{Y|\mathbf{X}^*}$ (or $F_{Y(T)|\mathbf{X}^*(T), T}$) and then to invert this function. Note that direct (in particular nonparametric) estimation of conditional distributions and quantiles (and also conditional measures of spread) can be very cumbersome, due to the dimensionality of the influencing factors $\mathbf{X}^*(T)$. A fully parametric approach remains feasible, but allowing for more flexibility (and less distributional assumptions) here requires to putting some mild structures.

A flexible approach is to consider regression models, and in particular location-scale models, as will be done from Section 2.2 onwards. In this regression setting we consider a linear structure $q_\tau(Y(T) | \mathbf{X}^*(T), T) = (\mathbf{X}^*(T))' \boldsymbol{\beta}^\tau(T)$, with $\boldsymbol{\beta}^\tau(T)$ a vector of p unknown univariate functions, which restricts the task to estimation of these p unknown functions. This additional structure has the advantage to allow for a more detailed study of the

influence of the covariates on the response, not only in terms of the quantile function, but also in terms of the variability in the conditional distribution.

2.2 Quantile regression

In multiple linear regression the influence of the p covariates $X^{(1)}, \dots, X^{(p)}$ on the response variable Y is, for the mean influence, modelled via

$$Y = \beta_0 + \beta_1 X^{(1)} + \dots + \beta_p X^{(p)} + \tilde{\varepsilon} = \mathbf{X}'\boldsymbol{\beta} + \tilde{\varepsilon}, \quad (2.2)$$

where $\mathbf{X} = (1, X^{(1)}, \dots, X^{(p)})'$, and where the error term $\tilde{\varepsilon}$ satisfies $E(\tilde{\varepsilon} | \mathbf{X}^*) = 0$ and $\text{Var}(\tilde{\varepsilon} | \mathbf{X}^*) = \sigma^2(\mathbf{X}^*)$, an unknown p -variate function. In a homoscedastic error model, it is assumed that $\sigma^2(\mathbf{X}^*) \equiv \sigma^2$, a constant. In many applications though heteroscedasticity is an issue, and the interest is also to estimate the p -variate function $\sigma^2(\mathbf{X}^*)$.

A varying-coefficient model is more flexible and essentially allows the regression coefficients in (2.2) to change with (one of) the covariates or another variable involved. In the longitudinal data setting described in Section 2.1 one often uses the following basic varying-coefficient model

$$Y(t) = \beta_0(t)X^{(0)}(t) + \beta_1(t)X^{(1)}(t) + \dots + \beta_p(t)X^{(p)}(t) + \tilde{\varepsilon}(t) = \mathbf{X}'(t)\boldsymbol{\beta}(t) + \tilde{\varepsilon}(t), \quad (2.3)$$

with $t \in \mathcal{T}$, $\mathbf{X}(t) = (X^{(0)}(t), X^{(1)}(t), \dots, X^{(p)}(t))' = (X^{(0)}(t), (\mathbf{X}^*(t))')'$, putting $X^{(0)}(t) \equiv 1$ for all $t \in \mathcal{T}$, and denoting $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \dots, \beta_p(t))'$. The function $\beta_0(t)$ is called the baseline function (or intercept function). For simplicity of presentation, we do not write the possible dependence of $\tilde{\varepsilon}(t)$ on the covariates $\mathbf{X}^*(t)$. In a general *mean regression* model the basic assumptions on the error term are

$$E(\tilde{\varepsilon}(T) | \mathbf{X}^*(T), T) = 0 \quad \text{and} \quad \text{Var}(\tilde{\varepsilon}(T) | \mathbf{X}^*(T), T) = \sigma^2(\mathbf{X}^*(T), T).$$

Special settings would be:

$$\sigma^2(\mathbf{X}^*(T), T) = \sigma^2(T) \quad \text{or even} \quad \sigma^2(X^{(1)}(T), \dots, X^{(p)}(T), T) = \sigma^2, \text{ a constant.}$$

We now turn to *quantile regression*. Let $0 \leq \tau \leq 1$, and denote the τ -th order conditional quantile of the error term $\tilde{\varepsilon}(T)$, given $T = t$, in model (2.3) by $\tilde{a}^\tau(t)$, i.e.

$$\tilde{a}^\tau(t) = \inf \{u : P\{\tilde{\varepsilon}(T) \leq u | T = t\} \geq \tau\}.$$

For all $0 \leq \tau_1 < \tau_2 \leq 1$:

$$\tilde{a}^{\tau_1}(t) \leq \tilde{a}^{\tau_2}(t). \quad (2.4)$$

Under model (2.3) the τ -th order conditional quantile of the response variable $Y(T)$ given $\{\mathbf{X}(T), T = t\}$ is

$$q_\tau(Y(T) | \mathbf{X}(T), T = t) = \mathbf{X}'(t)\boldsymbol{\beta}(t) + \tilde{a}^\tau(t) = \mathbf{X}'(t)\boldsymbol{\beta}^\tau(t), \quad (2.5)$$

where $\boldsymbol{\beta}^\tau(t) = (\boldsymbol{\beta}_0^\tau(t), \boldsymbol{\beta}_1(t), \dots, \boldsymbol{\beta}_p(t))'$, with

$$\boldsymbol{\beta}_0^\tau(t) = \boldsymbol{\beta}_0(t) + \tilde{\boldsymbol{a}}^\tau(t). \quad (2.6)$$

In the sequel we will use the shorthand notation $q_\tau(Y(t)|\mathbf{X}(t), t)$ for this conditional quantile function. Recalling that $X^{(0)}(t) = 1$, it is easily seen from (2.4) and (2.5) that

$$\text{for all } 0 \leq \tau_1 < \tau_2 \leq 1: \quad q_{\tau_1}(Y(t) | \mathbf{X}(t), t) \leq q_{\tau_2}(Y(t) | \mathbf{X}(t), t), \quad (2.7)$$

for any fixed value of $(\mathbf{X}(t), t)$. So the conditional quantile curves do not cross.

Remark 2.1. *Note that one has to deal with an identification problem if one aims at estimating $\boldsymbol{\beta}_0(t)$ using quantile regression. Indeed, if both quantities $\boldsymbol{\beta}_0(t)$ and $\tilde{\boldsymbol{a}}^\tau(t)$ are unknown then one can only estimate $\boldsymbol{\beta}_0^\tau(t)$, since adding a term to $\boldsymbol{\beta}_0(t)$ and subtracting the same term from $\tilde{\boldsymbol{a}}^\tau(t)$ in (2.6) leads to a same result. The quantity $\boldsymbol{\beta}_0(t)$ can only be estimated under the additional assumption that $\tilde{\boldsymbol{a}}^\tau(t)$ is fully known, for example*

(A1): *the τ -th order conditional quantile of the error term $\tilde{\boldsymbol{\varepsilon}}(T)$, given $T = t$, equals zero, i.e. $\tilde{\boldsymbol{a}}^\tau(t) = 0$.*

When Assumption **(A1)** holds, the τ th-order conditional quantile of $Y(T)$ given $T = t$ is

$$q_\tau(Y(t) | \mathbf{X}(t), t) = \mathbf{X}'(t)\boldsymbol{\beta}(t).$$

Andriyana *et al.* (2014) studied a P-splines estimation technique to estimate the unknown vector of regression coefficient functions $\boldsymbol{\beta}(t)$, in the longitudinal data setting. In this setting there are repeated measurements on n subjects/individuals. For subject i , the repeated measurements occur at time points t_{i1}, \dots, t_{iN_i} . At time point t_{ij} one observes, for subject i , the response variable $Y(t_{ij})$ and the vector of covariate values $(X^{(1)}(t_{ij}), \dots, X^{(p)}(t_{ij}))$, which we denote shorthanded as Y_{ij} and $(X_{ij}^{(1)}, \dots, X_{ij}^{(p)})$ respectively. The longitudinal observations of $(Y(T), \mathbf{X}(T), T)$ thus consist of $(Y_{ij}, \mathbf{X}_{ij}, t_{ij})$, with $i = 1, \dots, n$ and $j = 1, \dots, N_i$, where t_{ij} is the j -th measurement time for the i th subject, N_i is the number of repeated measurements for the i th subject, and $\mathbf{X}_{ij} = (X_{ij}^{(0)}, \dots, X_{ij}^{(p)})'$, with $X_{ij}^{(0)} \equiv 1$. Typically one assumes that the measurements are independent for different subjects, but measurements at different time points for a same individual can be correlated. For more details see Andriyana *et al.* (2014), as well as the simulation models in Sections 5 and 6.

We now briefly recall the P-splines estimation method for the conditional quantile function $q_\tau(Y(T) | \mathbf{X}(T), T = t)$ in model (2.3). For notational simplicity of presentation we drop the dependence on τ of the unknown coefficient functions, but this dependence should be kept in mind. Each of the unknown regression coefficient functions, $\boldsymbol{\beta}_k(t)$

for $k = 0, \dots, p$, is approximated by a set of normalized B-splines of a given degree. More precisely, the unknown regression coefficient function $\beta_k(t)$ is approximated by m_k normalized B-splines of degree ν_k , denoted by $B_{k1}(t; \nu_k), \dots, B_{km_k}(t; \nu_k)$:

$$\beta_k(t_{ij}) \approx \alpha_{k1}B_{k1}(t_{ij}; \nu_k) + \dots + \alpha_{km_k}B_{km_k}(t_{ij}; \nu_k) = \sum_{\ell=1}^{m_k} \alpha_{k\ell}B_{k\ell}(t_{ij}; \nu_k), \quad (2.8)$$

where $(\alpha_{k1}, \dots, \alpha_{km_k})$ denotes the associated coefficient vector that is to be estimated. The functions $B_{k\ell}(t_{ij}; \nu_k)$, $\ell = 1, \dots, m_k$, with $m_k = u_k + \nu_k$, constitute a basis of normalized B-splines of degree ν_k with $u_k + 1$ equidistant knots. The objective function for individual quantile estimation is then given by

$$S_{1,\tau}(\boldsymbol{\alpha}) = \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_\tau \left(Y_{ij} - \sum_{k=0}^p \sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij}; \nu_k) X_{ij}^{(k)} \right) + \sum_{k=0}^p \sum_{\ell=d_k+1}^{m_k} \lambda_k |\Delta^{d_k} \alpha_{k\ell}|^\gamma, \quad (2.9)$$

and is a function of the global vector of all unknown coefficients denoted by $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_0, \dots, \boldsymbol{\alpha}'_p)'$, with $\boldsymbol{\alpha}_k = (\alpha_{k1}, \dots, \alpha_{km_k})'$.

The first term in the objective function (2.9) is the goodness-of-fit term for quantile regression involving the function $\rho_\tau(\cdot)$, the so-called ‘‘check-function’’, defined as

$$\rho_\tau(z) = \begin{cases} \tau z & \text{if } z > 0 \\ -(1 - \tau)z & \text{otherwise} \end{cases} = \tau z^+ + (1 - \tau)z^-, \quad (2.10)$$

where we used the notations: $z^+ = \max(z, 0)$ and $z^- = \max(-z, 0)$.

The second term in (2.9) is the penalty term. As in Eilers and Marx (1996) this term is needed when taking m_k large for having a good approximation in (2.8). However a large number of knot points can lead to overfitting, and preventing from this to happen leads to the inclusion of the penalty term. As in Eilers and Marx (1996) we restrict ourselves to a penalty function that penalizes for too large differences between the coefficients of adjacent B-splines, i.e. using Δ^{d_k} the d_k -th order differencing operator: $\Delta^{d_k} \alpha_{k\ell} = \sum_{t=0}^{d_k} (-1)^t \binom{d_k}{t} \alpha_{k(\ell-t)}$, with $d_k \in \mathbb{N}$. The parameters $\lambda_k > 0$, $k = 0, \dots, p$, in (2.9) are the smoothing/regularization parameters that control the trade-off between the goodness-of-fit term and the penalty term. The power $\gamma > 0$ leaves the possibility of using a variety of penalty terms. We refer to Andriyana *et al.* (2014) for a detailed discussion on the P-splines estimation method in its generality. For simplicity of presentation we will take $\gamma = 1$ in the sequel of this paper. The advantages of the choice $\gamma = 1$ are that one can rely on linear programming (as opposed to convex programming) and a Frisch-Newton interior point algorithm can be used (implying faster computing times). See Portnoy and Koenker (1997). For a general discussion on different types of penalties, see for example Antoniadis *et al.* (2011).

Important is to note that due to the B-splines approximation in (2.8) we get to a very similar formulation as in a usual multiple linear regression model. Indeed consider

the matrix \mathbf{B} , with a block structure containing all the B-splines basis functions for approximating all component functions $\beta_k(\cdot)$:

$$\mathbf{B}(t) = \begin{pmatrix} B_{01}(t, \nu_0) & \dots & B_{0m_0}(t, \nu_0) & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \ddots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & B_{p1}(t, \nu_p) & \dots & B_{pm_p}(t, \nu_p) \end{pmatrix}. \quad (2.11)$$

Recalling (2.5) and approximation (2.8) we then can write

$$q_\tau(Y(t) \mid \mathbf{X}(t), t) \approx \mathbf{X}'(t)\mathbf{B}(t)\boldsymbol{\alpha}, \quad (2.12)$$

and hence $q_\tau(Y_{ij} \mid \mathbf{X}_{ij}, t_{ij}) = \mathbf{X}'_{ij}\mathbf{B}(t_{ij})\boldsymbol{\alpha}$. The objective function $S_{1,\tau}(\boldsymbol{\alpha})$ can be written as

$$\begin{aligned} S_{1,\tau}(\boldsymbol{\alpha}) &= \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_\tau \left(Y_{ij} - \mathbf{X}'_{ij}\mathbf{B}(t_{ij})\boldsymbol{\alpha} \right) + \sum_{k=0}^p \sum_{\ell=d_k+1}^{m_k} \lambda_k |\Delta^{d_k} \alpha_{k\ell}| \\ &= \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_\tau \left(Y_{ij} - \mathbf{U}'_{ij}\boldsymbol{\alpha} \right) + \sum_{k=0}^p \sum_{\ell=d_k+1}^{m_k} \lambda_k |\Delta^{d_k} \alpha_{k\ell}|, \end{aligned} \quad (2.13)$$

where $\mathbf{U}'_{ij} = \mathbf{X}'_{ij}\mathbf{B}(t_{ij})$. Note the appearance of a linear term in $\boldsymbol{\alpha}$ in the goodness-of-fit term.

Minimization of $S_{1,\tau}(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$ leads to the estimated global vector of coefficients $\widehat{\boldsymbol{\alpha}} = (\widehat{\boldsymbol{\alpha}}'_0, \dots, \widehat{\boldsymbol{\alpha}}'_p)'$, with $\widehat{\boldsymbol{\alpha}}_k = (\widehat{\alpha}_{k1}, \dots, \widehat{\alpha}_{km_k})'$. The P-splines estimator of the unknown regression coefficient function $\beta_k(\cdot)$ is

$$\widehat{\beta}_k(t) = \sum_{\ell=1}^{m_k} \widehat{\alpha}_{k\ell} B_{k\ell}(t_{ij}; \nu_k),$$

and subsequently the P-splines estimator for the conditional quantile function:

$$\widehat{q}_\tau(Y(t) \mid \mathbf{X}(t), t) = \widehat{\beta}_0(t)X^{(0)}(t) + \widehat{\beta}_1(t)X^{(1)}(t) + \dots + \widehat{\beta}_p(t)X^{(p)}(t) = \mathbf{X}'(t)\mathbf{B}(t)\widehat{\boldsymbol{\alpha}}^\tau,$$

where we now added the dependence on τ again.

Estimation of various conditional quantile curves is obtained by minimizing the individual objective function $S_{1,\tau}(\boldsymbol{\alpha})$ for various values of τ . There is however no guarantee that the estimated quantile curves satisfy the finite-sample analogue of (2.7), namely

$$\text{for all } 0 \leq \tau_1 < \tau_2 \leq 1: \quad \widehat{q}_{\tau_1}(Y(t) \mid \mathbf{X}(t), t) \leq \widehat{q}_{\tau_2}(Y(t) \mid \mathbf{X}(t), t),$$

for any fixed values of $(\mathbf{X}(t), t)$. This is further investigated (theoretically) in Section A of the Supplementary material and in Section 5 (in the simulation study).

2.3 Homoscedasticity versus heteroscedasticity

When in a multiple linear regression model (or more generally in a mean regression model) the attention is drawn also on the variance of the error term and the aspect of homoscedasticity against heteroscedasticity, the error term $\tilde{\varepsilon}$ is often written in the form

$$\tilde{\varepsilon} = \sigma(X^*) \varepsilon \quad \text{with} \quad E(\varepsilon | X^*) = 0 \quad \text{and} \quad \text{Var}(\varepsilon | X^*) = 1 .$$

The last assumption is in fact needed for identifiability of the variance function $\sigma^2(\cdot)$.

The second aim of this paper is to deal with estimation of a similar kind of variability function in the setting of quantile regression in varying-coefficient models. Similarly as in the mean regression model, we therefore re-express the error term in model (2.3) as

$$\tilde{\varepsilon}(t) = V(\mathbf{X}^*(t), t) \varepsilon(t) ,$$

with $V(\cdot) \geq 0$, and denote the τ -th order conditional quantile of the error term $\varepsilon(T)$, given $T = t$, in model (2.3) by $a^\tau(t)$: $a^\tau(t) = \inf \{u : P\{\varepsilon(T) \leq u | T = t\} \geq \tau\}$. Some special cases are to be noted:

$$V(\mathbf{X}^*(T), T) = V(T) \quad \text{or even} \quad V(\mathbf{X}^*(T), T) = V, \text{ a nonnegative constant.}$$

As in the case of mean regression we will need to impose some additional assumption on the random error term $\varepsilon(t)$ to ensure identifiability of the function $V(\cdot)$. We refer to this function $V(\cdot)$ as the *variability function*.

In this paper we restrict to the special case that $V(\mathbf{X}^*(T), T) = V(T)$ and the error term $\varepsilon(T)$ is independent of $\mathbf{X}^*(T)$ for given T . We refer to this model as the (simple) heteroscedastic model. We speak about a homoscedastic model if, in addition, $V(T) = V$ (a constant).

In case of a simple heteroscedastic model

$$Y(t) = \mathbf{X}'(t)\boldsymbol{\beta}(t) + V(t) \varepsilon(t), \quad t \in \mathcal{T} , \quad (2.14)$$

with $V(\cdot) \geq 0$. In such a *location-scale model* we have that

$$\tilde{a}^\tau(t) = V(t)a^\tau(t) .$$

Under model (2.14) the τ -th order conditional quantile of the response variable $Y(T)$ given $\{\mathbf{X}(T), T = t\}$ is then

$$q_\tau(Y(t) | \mathbf{X}(t), t) = \mathbf{X}'(t)\boldsymbol{\beta}(t) + V(t)a^\tau(t) = \mathbf{X}'(t)\boldsymbol{\beta}^\tau(t) , \quad (2.15)$$

where $\boldsymbol{\beta}^\tau(t) = (\beta_0^\tau(t), \beta_1(t), \dots, \beta_p(t))^T$, with now

$$\beta_0^\tau(t) = \beta_0(t) + V(t)a^\tau(t) . \quad (2.16)$$

The interests are now in: (i) estimating the quantile curves $q_\tau(Y(t) | \mathbf{X}(t), t)$, through estimation of the functions $\beta_0^\tau(t), \beta_1(t), \dots, \beta_p(t)$; (ii) estimation of the variability function $V(\cdot)$.

Some examples of error structures for $\tilde{\varepsilon}(t)$ and associated conditional quantiles are: (i) for $\tilde{\varepsilon}(t) \sim N(\mu(t); \sigma^2(t))$, the function $\tilde{a}^\tau(t) = \mu(t) + \sigma(t)\Phi^{-1}(\tau)$; (ii) for $\tilde{\varepsilon}(t) \sim \text{Lognormal}(\mu(t); \sigma(t))$, the function $\tilde{a}^\tau(t) = \exp\{\mu(t)\} [\exp\{\sigma(t)\}]^{\Phi^{-1}(\tau)}$. Here Φ denotes the cumulative distribution function of a $N(0; 1)$ random variable.

Remark 2.2. From (2.16) it is clear what will be needed in order for $V(t)$ to be identifiable. Sufficient conditions for this to happen are that, for example, $a^\tau(\cdot)$ is known for at least two given values of τ , say τ_1 and τ_2 . Indeed, from (2.16) we have

$$\beta_0^{\tau_1}(t) - \beta_0^{\tau_2}(t) = V(t) (a^{\tau_1}(t) - a^{\tau_2}(t)) .$$

An estimator for $V(t)$ is then

$$\widehat{V}(t) = \frac{\widehat{\beta}_0^{\tau_1}(t) - \widehat{\beta}_0^{\tau_2}(t)}{a^{\tau_1}(t) - a^{\tau_2}(t)} .$$

So a sufficient condition for estimating the variability function $V(\cdot)$ in model (2.14) is the knowledge of two quantile functions of the error term:

(B): the τ_1 -th order and the τ_2 -th conditional quantile of the error term ε , given $T = t$, with $0 \leq \tau_1, \tau_2 \leq 1$ and $\tau_1 \neq \tau_2$, are fully known (i.e. $a^{\tau_1}(t)$ and $a^{\tau_2}(t)$ are fully known).

It is interesting to look when the identifiability is ensured for the examples of error distributions mentioned above: (i) for $\tilde{\varepsilon}(t) \sim N(\mu(t); \sigma^2(t))$, identifiability is ensured for example when $\mu(t) = 0$, leading to $V(t) = \sigma(t)$ and $a^\tau(t) = \Phi^{-1}(\tau)$; (ii) for $\tilde{\varepsilon}(t) \sim \text{Lognormal}(\mu(t); \sigma(t))$, identifiability is guaranteed for example in case $\sigma(t) = 1$, leading to $V(t) = \exp\{\mu(t)\}$, and $a^\tau(t) = [\exp\{1\}]^{\Phi^{-1}(\tau)}$.

From now on we entirely focus on the location-scale model (2.14) and our two-fold estimation task:

Task 1. estimation of the quantile regression curves in (2.15), via estimation of $\beta_0^\tau(t), \beta_1(t), \dots, \beta_p(t)$, and guaranteeing that the estimated curves do not cross;

Task 2. estimation of the variability function $V(\cdot)$.

Estimation **Task 1** is dealt with mainly in Section 3 and Section A of the Supplementary material, whereas estimation **Task 2** is discussed in Section 3 (partly) and Section 4.

3 Methods to prevent for crossing estimated quantile curves

The aim of this section is to develop methods that deal with the issue of crossing estimated quantile curves, based on the P-splines estimation technique. We present several methods, and evaluate and compare their performances via a finite-sample simulation study in Section 5 (see also the Supplementary material).

The proposed methods are inspired by three approaches available in the literature: (a) considering simultaneous objective functions (for various values of τ); (b) the approach followed by He (1997); (c) the stepwise approach of Wu and Liu (2009). In the first approach an important issue is of how to combine the individual objective functions into an aimed simultaneous objective function, such that the non-crossing is enforced. This discussion leads to the methods described in Sections 3.1 and 3.2. Approach (c) leads to the method discussed in Section 3.3, and the adaptation of approach (b) in the current setting of varying-coefficient models can be found in Section 3.4.

All methods proposed involve minimizing objective functions, using P-splines approximation. All optimization problems are translated into a linear programming problem. Due to space limitations we do not provide these implementation details here.

3.1 Method of weighted simultaneous quantile objective functions

Recall the matrix notation of the objective function in (2.13). Consider now H different values of τ denoted by $0 \leq \tau_1 < \dots < \tau_H \leq 1$, with $H \in \mathbb{N}_0$. Theoretically the H quantile curves $q_{\tau_h}(Y(t) | \mathbf{X}(t), t)$ cannot cross, for any given value of the covariate vector and the variable T . We want to prevent this from happening with the P-splines estimator $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$.

Since we need to look at several values for τ simultaneously, we combine the different individual objective functions. An obvious way to do so is to consider

$$\sum_{h=1}^H \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_h} \left(Y_{ij} - \mathbf{U}'_{ij} \boldsymbol{\alpha}_h \right) + \sum_{h=1}^H \sum_{k=0}^p \sum_{\ell=d_k+1}^{m_k} \lambda_{\tau_h, k} \left| \Delta^{d_k} \alpha_{hk\ell} \right| ,$$

where we needed to introduce additional necessary indices. If we want to ensure that the P-splines estimators $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$ do not cross at each of the observations, i.e.

$$\hat{q}_{\tau_h}(Y_{ij} | \mathbf{X}_{ij}, t_{ij}) \geq \hat{q}_{\tau_{h-1}}(Y_{ij} | \mathbf{X}_{ij}, t_{ij}) \quad \text{for all } h = 2, \dots, H ,$$

we need to include additional constraints into the simultaneous objective function. This would involve $Nm(H-1)$ constraints, with $N = \sum_{i=1}^n N_i$ the total number of observations,

and $m = \sum_{k=0}^p m_k$ the number of columns in the matrix \mathbf{B} in (2.11). Because of the large number of constraints the computational costs would be very substantial. Therefore, we need to find a way to reduce the number of additional constraints. The function values of normalized B-splines are always nonnegative, and if the covariate values $X^{(k)}(t_{ij})$ are nonnegative, for each possible t_{ij} , then the constraints for non-crossing curves can be simply imposed on the coefficient vectors $\boldsymbol{\alpha}_h$:

$$\boldsymbol{\alpha}_h \geq \boldsymbol{\alpha}_{h-1} \quad \text{for all } h = 2, \dots, H, \quad (3.1)$$

where the inequality sign means this should hold for all components of the vectors. This reduces the number of constraints to be $m(H - 1)$.

In practice we thus first transform the covariates $X^{(k)}(t)$ to be nonnegative on the range of the observations. One way to do so is by subtracting from $X^{(k)}(t)$ the smallest observed value of $X^{(k)}(t_{ij})$, i.e. by subtracting $\min_{1 \leq i \leq n, 1 \leq j \leq N_i} \{X^{(k)}(t_{ij})\}$.

Imposing the constraints (3.1) leads to adding terms in the optimization problem, together with a Lagrange multiplier parameter, denoted by $\lambda^{(\text{NC})}$ (where the superscript (NC) refers to the word NON-CROSSING). To make clear distinction with the regularization parameters $\lambda_{\tau_h, k}$ for the penalty terms, we add the superscript (PS) to the latter parameters. This results into the following simultaneous objective function

$$\begin{aligned} S_2(\boldsymbol{\alpha}) = & \sum_{h=1}^H \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_h} \left(Y_{ij} - \mathbf{U}'_{ij} \boldsymbol{\alpha}_h \right) + \lambda^{(\text{NC})} \sum_{h=2}^H w(\tau_h) \sum_{k=0}^p \sum_{\ell=1}^{m_k} |\Delta^{(\text{NC})} \alpha_{hk\ell}| \\ & + \sum_{h=1}^H \sum_{k=0}^p \sum_{\ell=d_k+1}^{m_k} \lambda_{\tau_h, k}^{(\text{PS})} |\Delta^{d_k} \alpha_{hk\ell}|, \end{aligned} \quad (3.2)$$

where we denoted $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1, \dots, \boldsymbol{\alpha}'_H)'$, $\Delta^{(\text{NC})} \alpha_{hk\ell} = \alpha_{hk\ell} - \alpha_{(h-1)k\ell}$ for $h = 2, \dots, H$, $k = 0, \dots, p$ and $\ell = 1, \dots, m_k$, and where the weight function defines the constraints on the coefficients of all $\boldsymbol{\alpha}_h$'s.

$$w(\tau_h) = \begin{cases} 1 & \text{if } \Delta^{(\text{NC})} \alpha_{hk\ell} < 0 \text{ for some } k \text{ and } \ell \\ 0 & \text{otherwise.} \end{cases}$$

The weight function reduces the number of constraints in case the inequalities are fulfilled in some or even all estimates.

For given tuning parameters, $\lambda^{(\text{NC})}$, and $\lambda_{\tau_h, k}^{(\text{PS})}$, for $h = 1, \dots, H$ and $k = 0, \dots, p$, we need to minimize $S_2(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$. To do so we transform the optimization problem into a (primal) linear programming problem and use a *Frisch-Newton* interior point algorithm to solve the linear programming problem. See also Koenker and Ng (2005).

The total number of tuning parameters to be chosen in (3.2) is $H(p+1)+1$. Combining all $H(p+1)+1$ possible grids of λ 's is of course not possible due to the computational

demand. A practical data-driven choice for these tuning parameters, that is used for all methods in this paper, can be found in Section B of the Supplementary material.

The objective function in (3.2) is obtained by combining the individual objective functions and adding constraints for imposing the non-crossing property. In the goodness-of-fit term each individual quantile curve gets the same weight. Because the quality and/or difficulty of estimation can be different for different τ 's it might make sense to include a weight factor W_h into the simultaneous objective function, i.e. to consider a goodness-of-fit term of the form

$$\sum_{h=1}^H W_h \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_h} \left(Y_{ij} - \mathbf{U}'_{ij} \boldsymbol{\alpha}_h \right)$$

instead of the first term in (3.2).

This issue of appropriate weighting in case of simultaneous estimation of quantile curves has received some attention in the literature of quantile regression. Zou and Yuan (2008), Bondell *et al.* (2010), Guo *et al.* (2012), Schnabel and Eilers (2013a), among others, take $W_h = 1$, for all τ_h -values, which corresponds with the approach in (3.2). Alexander *et al.* (2011) considers also constant weights $W_h = 1/H$, whereas a fixed weighting scheme assigning the highest weight W_h for median estimation is opted for in Koenker (2004) and Lamarche (2010), among others. Zhao and Xiao (2014) and Jiang *et al.* (2014) obtain W_h 's by minimizing the asymptotic variance of the associated quantile coefficient in the simultaneous objective function. Such an approach is only applicable when an expression for the asymptotic variance of the quantile estimator is available, and moreover can be estimated from the data. In complex (and flexible) models, as the varying-coefficient model, this is not feasible. Our approach below is inspired by Liu and Wu (2011). They take $W_h = 1/\text{E}[\rho_{\tau_h}(\epsilon - q_{\tau_h}(\epsilon))]$, assuming that the regression error term is $\epsilon \sim N(0, 1)$ and hence $\text{E}[\rho_{\tau_h}(\epsilon - q_{\tau_h}(\epsilon))] = \phi(\Phi^{-1}(\tau_h))$, where ϕ and Φ are respectively the density and cumulative distribution function of a standard normal random variable.

We now briefly discuss our alternative weighting scheme. Note that for the simple heteroscedastic model (2.14) and the expression for the quantile regression curve it follows that

$$Y(t) = q_{\tau}(Y(t) \mid \mathbf{X}(t), t) + V(t) [\varepsilon(t) - a^{\tau}(t)] ,$$

and this holds for all values of τ . Therefore

$$Y(t) - q_{\tau_h}(Y(t) \mid \mathbf{X}(t), t) = V(t) [\varepsilon(t) - a^{\tau_h}(t)] , \quad (3.3)$$

and since $\rho_{\tau}(z) = z(\tau - I\{z < 0\})$ and $V(t) \geq 0$, we find that

$$\begin{aligned} \rho_{\tau_h}(V(t) [\varepsilon(t) - a^{\tau_h}(t)]) &= V(t) [\varepsilon(t) - a^{\tau_h}(t)] (\tau_h - I\{V(t) [\varepsilon(t) - a^{\tau_h}(t)] < 0\}) \\ &= V(t) [\varepsilon(t) - a^{\tau_h}(t)] (\tau_h - I\{\varepsilon(t) - a^{\tau_h}(t) < 0\}) \\ &= V(t) \rho_{\tau_h}(\varepsilon(t) - a^{\tau_h}(t)) . \end{aligned}$$

Taking expectation (conditionally upon a given value for $T = t$), we get

$$E \{ \rho_{\tau_h} (V(t) [\varepsilon(t) - a^{\tau_h}(t)]) \} = V(t) E \{ \rho_{\tau_h} (\varepsilon(t) - a^{\tau_h}(t)) \} ,$$

which shows that the influence of the expectation by τ is captured by $E \{ \rho_{\tau_h} (\varepsilon(t) - a^{\tau_h}(t)) \}$. It thus seems reasonable to assign a larger weight to a smaller term in the global goodness-of-fit term, i.e. to consider

$$W_h = \frac{1}{E \{ \rho_{\tau_h} (\varepsilon(t) - a^{\tau_h}(t)) \}} .$$

Since the conditional distribution of $\varepsilon(t)$ is not known, the expectation in this weight is approximated by

$$\frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_h} (\varepsilon(t_{ij}) - a^{\tau_h}(t_{ij}))} .$$

Using (3.3), the unknown expression in the denominator is then further approximated by

$$\widehat{W}_h = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_h} \left(\frac{Y_{ij} - \widehat{q}_{\tau_h}(Y_{ij} | \mathbf{X}_{ij}, t_{ij})}{\widehat{V}(t_{ij})} \right)} , \quad (3.4)$$

where $\widehat{V}(\cdot)$ is a consistent estimator of the variability function $V(\cdot)$, that needs to be obtained in a first step. In Section 4 we discuss methods for estimating $V(\cdot)$ using the unweighted simultaneous objective function in (3.2).

To summarize, in a first step the unweighted simultaneous objective procedure in (3.2) is used, leading also to an estimator for $V(\cdot)$. In a second step one then uses the estimated weight in the simultaneous weighted objective function

$$\begin{aligned} S_{2,w}(\boldsymbol{\alpha}) &= \sum_{h=1}^H \frac{1}{\widehat{W}_h} \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_h} \left(Y_{ij} - \mathbf{U}'_{ij} \boldsymbol{\alpha}_h \right) + \lambda^{(\text{NC})} \sum_{h=2}^H w(\tau_h) \sum_{k=0}^p \sum_{\ell=1}^{m_k} |\Delta^{(\text{NC})} \alpha_{hk\ell}| \\ &+ \sum_{h=1}^H \sum_{k=0}^p \sum_{\ell=d_k+1}^{m_k} \lambda_{\tau_h,k}^{(\text{PS})} |\Delta^{d_k} \alpha_{hk\ell}| , \end{aligned}$$

which is then minimized with respect to $\boldsymbol{\alpha}$.

3.2 Quantile sheet method

In this section, we propose an alternative to overcome the crossing curves, using also a simultaneous objective function approach. The idea is based on the paper of Schnabel and Eilers (2013a), on univariate nonparametric quantile regression. The basic idea of their method is to consider a bivariate surface representing the behaviour of the quantile curve

in *both* the covariate and the quantile order τ . They use a tensor product of B-splines to describe the behaviour in *each* of the two variables (the covariate and τ). The non-crossing property of the quantile curves is then imposed by requiring that the bivariate function is monotonic increasing in the argument τ . The latter is done by adding a specific penalty. We adapt this method to the context of varying-coefficient models, and refer to it as quantile sheet (following the terminology of Schnabel and Eilers (2013a)).

Here we have the following simultaneous objective function,

$$\begin{aligned} & \sum_{h=1}^H \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_h} \left(Y_{ij} - \sum_{\tilde{h}=1}^{\tilde{m}} \sum_{k=0}^p \sum_{\ell=1}^{\tilde{m}_k} \alpha_{\tilde{h}k\ell} \left[\tilde{B}_{\tilde{h}}(\tau_h; \tilde{\nu}) \times \left(\check{B}_{k\ell}(t_{ij}; \check{\nu}_k) X_{ij}^{(k)} \right) \right] \right) \\ & + \lambda^{(\text{NC})} \sum_{\tilde{h}=2}^{\tilde{m}} w(\tau_h) \sum_{k=0}^p \sum_{\ell=1}^{\tilde{m}_k} |\Delta^{(\text{NC})} \alpha_{\tilde{h}k\ell}| + \sum_{\tilde{h}=1}^{\tilde{m}} \sum_{k=0}^p \sum_{\ell=d_k+1}^{\tilde{m}_k} \lambda_k^{(\text{PS})} |\Delta^{d_k} \alpha_{\tilde{h}k\ell}|, \end{aligned} \quad (3.5)$$

where $\tilde{B}_{\tilde{h}}(\tau_h; \tilde{\nu})$, $\tilde{h} = 1, \dots, \tilde{m}$ denotes a B-spline basis on the domain $\tau \in [0, 1]$ of degree $\tilde{\nu}$ with $\tilde{u} + 1$ equidistant knots with $\tilde{m} = \tilde{u} + \tilde{\nu}$. The functions $\check{B}_{k\ell}(t_{ij}; \check{\nu}_k)$, $\ell = 1, \dots, \tilde{m}_k$, form a B-spline basis, on the domain \mathcal{T} , of degree $\check{\nu}_k$ with $\check{u}_k + 1$ equidistant knots for the k -th component of the covariates with $\tilde{m}_k = \check{u}_k + \check{\nu}_k$. $\Delta^{(\text{NC})} \alpha_{\tilde{h}k\ell} = \alpha_{\tilde{h}k\ell} - \alpha_{(\tilde{h}-1)k\ell}$, for all $\tilde{h} = 2, \dots, \tilde{m}$. Further $\boldsymbol{\alpha}_{\tilde{h}} = \left(\alpha_{\tilde{h}01}, \dots, \alpha_{\tilde{h}0\tilde{m}_0}, \dots, \alpha_{\tilde{h}p1}, \dots, \alpha_{\tilde{h}p\tilde{m}_p} \right)'$ is the associated coefficient vector of dimension $\tilde{m} = \sum_{k=0}^p \tilde{m}_k$.

Expression (3.5) can be written more compactly by introducing some additional matrix notation. Similarly as in (2.11) in Section 3.1 we denote the matrix of B-spline basis functions, formed by the functions $\check{B}_{k\ell}(t, \check{\nu}_k)$, by $\check{\mathbf{B}}(t)$. Next, we put $\check{\mathbf{U}}'_{ij} = \mathbf{X}'_{ij} \check{\mathbf{B}}(t_{ij})$, a row vector of dimension \tilde{m} . By stacking the vectors $\check{\mathbf{U}}'_{ij}$, for $i = 1, \dots, n$ and $j = 1, \dots, N_i$, as rows in a matrix, we build the matrix $\check{\mathbf{U}}$ of dimension $N \times \tilde{m}$. In the domain of the quantile variable τ , we build in a similar fashion the matrix $\tilde{\mathbf{B}}$ of dimension $H \times \tilde{m}$ in which the h -th row consists of $(\tilde{B}_1(\tau_h, \tilde{\nu}), \dots, \tilde{B}_{\tilde{m}}(\tau_h, \tilde{\nu}))$. We then define $\mathbf{U} = \tilde{\mathbf{B}} \otimes \check{\mathbf{U}}$, the Kronecker product of the two matrices. The matrix \mathbf{U} is a block matrix of dimension $H N \times \tilde{m} \tilde{m}$, where the row of this matrix corresponding to observations on subject i at time t_{ij} , with respect to a fixed value τ_h , is referred to as \mathbf{U}_{hij} . Each row is of dimension $\tilde{m} \times \tilde{m}$. With this matrix notation, we can rewrite the objective function in (3.5) as

$$\begin{aligned} S_3(\boldsymbol{\alpha}) &= \sum_{h=1}^H \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_h} \left(Y_{ij} - \mathbf{U}'_{hij} \boldsymbol{\alpha}_h \right) \\ &+ \lambda^{(\text{NC})} \sum_{\tilde{h}=2}^{\tilde{m}} w(\tau_h) \sum_{k=0}^p \sum_{\ell=1}^{\tilde{m}_k} |\Delta^{(\text{NC})} \alpha_{\tilde{h}k\ell}| + \sum_{\tilde{h}=1}^{\tilde{m}} \sum_{k=0}^p \sum_{\ell=d_k+1}^{\tilde{m}_k} \lambda_k^{(\text{PS})} |\Delta^{d_k} \alpha_{\tilde{h}k\ell}|, \end{aligned} \quad (3.6)$$

with $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1, \dots, \boldsymbol{\alpha}'_{\tilde{m}})'$ where $\boldsymbol{\alpha}_h = \left(\alpha_{h01}, \dots, \alpha_{h0\tilde{m}_0}, \dots, \alpha_{hp1}, \dots, \alpha_{hp\tilde{m}_p} \right)'$, for $h = 1, \dots, \tilde{m}$.

It should be mentioned that Schnabel and Eilers (2013a) solve their optimization problem using a classic iteratively re-weighted least squares approach combined with the fast array algorithm GLAM (Currie *et al.* (2006); Eilers *et al.* (2006)) for multidimensional P-spline fitting. In contrast, in our context, we transform the optimization problem into a linear programming problem exploiting the sparsity in some matrices. Also here we can use the *Frisch-Newton* interior point algorithm. See for example Koenker and Ng (2005). As before, we first need to transform our covariates $\mathbf{X}(t)$ to be positive.

3.3 Stepwise individual quantile regression estimation

Another approach to ensure that estimated quantile curves do not cross, was introduced by Wu and Liu (2009). They proposed a stepwise procedure that starts by estimating a particular quantile function, then in the next step moves on to estimating the next (higher or lower order) quantile in the given set of quantile orders ($0 \leq \tau_1 < \dots < \tau_H \leq 1$). To ensure the non-crossing property of the quantiles involved in the first and second step, they include the necessary constraints. The procedure continues by moving to the next step (next quantile order). A specific choice to be made when applying this method is: which quantile is estimated at the starting step? Wu and Liu (2009) advice to use median estimation as a starting point, and we follow this recommendation when adapting this approach to our setting. In the upward step (moving-up), we add a constraint so that the estimated larger order quantile curve exceeds the preceding estimated quantile curve. Otherwise, in the downward step, we put a constraint such that the estimated smaller order quantile curve does not exceed the preceding estimated quantile curve. To reduce the number of constraints (see also the description in Section 3.1), we again transform the covariates to be positive.

The main steps of the Upward and Downward stepwise procedures read as follows.

STEP 1: Estimating the median regression curve

The P-splines median quantile estimator is given by

$$\hat{\alpha}^{0.5} = \operatorname{argmin}_{\alpha} \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5} \left(Y_{ij} - \mathbf{U}'_{ij} \alpha \right) + \sum_{k=0}^p \sum_{\ell=d_k+1}^{m_k} \lambda_{0.5,k}^{(\text{PS})} \left| \Delta^{d_k} \alpha_{k\ell} \right| .$$

STEP 2: Complete Up (CU)

Starting from $\tau_h = 0.5$, the next larger order ($\tau_{h+1} > \tau_h$) is obtained from the following constrained minimization problem:

minimize

$$\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_{h+1}} \left(Y_{ij} - \mathbf{U}'_{ij} \alpha \right) + \sum_{k=0}^p \sum_{\ell=d_k+1}^{m_k} \lambda_{\tau_{h+1},k}^{(\text{PS})} \left| \Delta^{d_k} \alpha_{k\ell} \right| \quad (3.7)$$

with respect to $\boldsymbol{\alpha}$, subject to

$$\boldsymbol{\alpha} \geq \widehat{\boldsymbol{\alpha}}^{\tau_h} . \quad (3.8)$$

The Complete Upward (CU) stepwise procedure consists of minimizing (3.7) subject to (3.8) for subsequently larger values τ_{h+1} (for h taking values in a given grid).

STEP 2: Complete Down (CD)

This is similar to the Complete Upward procedure, but moving downwards from Step 1 (median estimation) on, replacing (3.7) and its constraint (3.8) by:

minimize

$$\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_{h-1}} \left(Y_{ij} - \mathbf{U}'_{ij} \boldsymbol{\alpha} \right) + \sum_{k=0}^p \sum_{\ell=d_k+1}^{m_k} \lambda_{\tau_{h-1},k}^{(PS)} \left| \Delta^{d_k} \alpha_{k\ell} \right|$$

with respect to $\boldsymbol{\alpha}$, subject to

$$\boldsymbol{\alpha} \leq \widehat{\boldsymbol{\alpha}}^{\tau_h} ,$$

and carrying this out sequentially.

In Section 5 we investigate the performance of this method to ensuring non-crossing of the estimated quantile curves.

3.4 AHe approach: an adaptation of He's (1997) approach

Yet another method can be proposed by adapting the approach of He (1997) to our context of varying-coefficient models. Crucial here are the following assumptions on the error structure.

(H1): the (conditional) median quantile of the error term $\varepsilon(t)$ equals zero: $q_{0.5}(\varepsilon(t)) = 0$ (i.e. $a^{0.5}(t) = 0$).

(H2): the (conditional) median quantile of the absolute value of the error term $\varepsilon(t)$ equals one: $q_{0.5}(|\varepsilon(t)|) = 1$.

These assumptions are sufficient to ensure identifiability when it comes to estimating the quantile curves, as well as the variability function. See also Remark 2.2.

The procedure consists of three steps: (i) in a first step the median quantile function is estimated; (ii) due to the assumptions and the first step this allows to estimate the function $V(t)$ in a second step; (iii) in a final step the estimation results from the two previous steps are used to obtain the general quantile regression estimates. More precisely the various steps in this estimation procedure are as follows.

Step 1. Under Assumption **(H1)**, the median quantile function under model (2.14) is given by

$$q_{0.5}(Y(t) \mid \mathbf{X}(t), t) = \beta_0(t)X^{(0)}(t) + \beta_1(t)X^{(1)}(t) + \cdots + \beta_p(t)X^{(p)}(t) = \mathbf{X}'(t)\boldsymbol{\beta}(t) ,$$

with $\boldsymbol{\beta}(t) = (\beta_0(t), \dots, \beta_p(t))'$.

Using the P-splines estimation method of Andriyana *et al.* (2014) (with $\tau = 0.5$), as described in Section 2, we obtain the estimated regression coefficient functions $(\widehat{\beta}_0(t), \dots, \widehat{\beta}_p(t))$, and subsequently the estimated median regression curve:

$$\widehat{q}_{0.5}(Y(t) \mid \mathbf{X}(t), t) = \widehat{\beta}_0(t)X^{(0)}(t) + \widehat{\beta}_1(t)X^{(1)}(t) + \cdots + \widehat{\beta}_p(t)X^{(p)}(t) = \mathbf{X}'(t)\widehat{\boldsymbol{\beta}}(t) ,$$

Step 2. From Model (2.14) and since $V(t) \geq 0$, we have

$$|Y(t) - \mathbf{X}'(t)\boldsymbol{\beta}(t)| = V(t)|\varepsilon(t)| ,$$

and using Assumption **(H2)** we find that the conditional median of $|Y(t) - \mathbf{X}'(t)\boldsymbol{\beta}(t)|$ equals $V(t)$.

$$q_{0.5} \left(|Y(t) - \mathbf{X}'(t)\boldsymbol{\beta}(t)| \right) = V(t) .$$

The unknown variability function $V(\cdot)$ can then be estimated using a P-splines estimation technique, based on the pseudo-observations of the quantity $|Y(t) - \mathbf{X}'(t)\boldsymbol{\beta}(t)|$ resulting from **Step 1**, i.e. using $\left| Y(t_{ij}) - \mathbf{X}'(t_{ij})\widehat{\boldsymbol{\beta}}(t_{ij}) \right|$, for $i = 1, \dots, n$ and $j = 1, \dots, N_i$. We therefore approximate the unknown variability function $V(\cdot)$ by a B-spline basis, in a similar fashion as the approximations in (2.8). We denote the B-splines basis of degree ν^v , using $u^v + 1$ equidistant knot points, by $B_\ell^v(\cdot; \nu^v)$, $\ell = 1, \dots, m^v$, with $m^v = u^v + \nu^v$. The superscript v is to draw the attention that this is the B-splines basis for estimation of $V(\cdot)$. In other words, we consider

$$V(t_{ij}) \approx \sum_{\ell=1}^{m^v} \alpha_\ell^v B_\ell^v(t_{ij}; \nu^v) ,$$

where $(\alpha_1^v, \dots, \alpha_{m^v}^v)$ denotes the associated coefficient vector that is to be estimated. The estimator $\widehat{V}(t)$ is then given by

$$\widehat{V}(t) = \sum_{\ell=1}^{m^v} \widehat{\alpha}_\ell^v B_\ell^v(t; \nu^v) , \quad (3.9)$$

where $(\widehat{\alpha}_1^v, \dots, \widehat{\alpha}_{m^v}^v)$ is obtained by minimizing

$$\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5} \left(\left| Y(t_{ij}) - \mathbf{X}'(t_{ij})\widehat{\boldsymbol{\beta}}(t_{ij}) \right| - \sum_{\ell=1}^{m^v} \alpha_\ell^v B_\ell^v(t_{ij}; \nu^v) \right) + \sum_{\ell=d_v+1}^{m^v} \lambda_v |\Delta^{d_v} \alpha_\ell^v| , \quad (3.10)$$

with respect to $\boldsymbol{\alpha}^v = (\alpha_1^v, \dots, \alpha_{m^v}^v)$, where $\lambda_v > 0$ is the penalization parameter, and where d_v is the order of the differencing operator in the penalty term.

Step 3. Equipped with the estimated vector $(\widehat{\beta}_0(t), \dots, \widehat{\beta}_p(t))$ from **Step 1** and the estimator $\widehat{V}(t)$ from **Step 2**, and recalling that

$$Y(t) - \mathbf{X}'(t)\boldsymbol{\beta}(t) = V(t)\varepsilon(t) ,$$

we can now herein substitute $\boldsymbol{\beta}(t)$ by $\widehat{\boldsymbol{\beta}}(t)$ and $V(t)$ by $\widehat{V}(t)$, and approximate the unknown (conditional) quantile $a^{\tau_h}(t)$ of the error term $\varepsilon(t)$ by m_h B-spline basis functions, say of degree ν_h :

$$a^{\tau_h}(t_{ij}) \approx \sum_{\ell=1}^{m_h} \alpha_{h,\ell}^q B_\ell^q(t_{ij}; \nu_h) .$$

The coefficients $(\alpha_{h,1}^q, \dots, \alpha_{h,m_h}^q)$ in the expansion are obtained again by P-splines approximation. More precisely minimizing the objective function

$$\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_h} \left(Y(t_{ij}) - \mathbf{X}'(t_{ij})\widehat{\boldsymbol{\beta}}(t_{ij}) - \widehat{V}(t_{ij}) \sum_{\ell=1}^{m_h} \alpha_{h,\ell}^q B_\ell^q(t_{ij}; \nu_h) \right) + \sum_{\ell=d_h^q+1}^{m_h} \lambda_h^q \left| \Delta^{d_h^q} \alpha_{h,\ell}^q \right| ,$$

with respect to $\boldsymbol{\alpha}_h^q = (\alpha_{h,1}^q, \dots, \alpha_{h,m_h}^q)$, where $\lambda_h^q > 0$ is the penalization parameter, and where d_h^q is the order of the differencing operator in the penalty term for this quantile estimation. Denote the resulting estimator of $\boldsymbol{\alpha}_h^q$ by $\widehat{\boldsymbol{\alpha}}_h^q$. The estimator of $a^{\tau_h}(t)$ is then

$$\widehat{a}^{\tau_h}(t) = \sum_{\ell=1}^{m_h} \widehat{\alpha}_{h,\ell}^q B_\ell^q(t; \nu_h) ,$$

and the estimated quantile function is given by

$$\widehat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t) = \mathbf{X}'(t)\widehat{\boldsymbol{\beta}}(t) + \widehat{V}(t)\widehat{a}^{\tau_h}(t) .$$

Remark 3.1. *Some remarks can be made about the above estimation procedure. A first remark is that the estimator $\widehat{V}(t)$ is not necessarily positive. We do not elaborate on this point, but just mention that this issue can be addressed in several ways. A first possibility would be to approximate $\log(V(t))$ by B-splines, instead of $V(t)$. An alternative would be to include some nonnegativity constraints on the coefficients of the B-spline approximation of $V(\cdot)$. In our implementations of the methods we apply the first option, and approximate $\log(V(t))$ with B-splines. A second remark is that in Step 3 it is not granted that the estimated quantile curves $\widehat{a}^{\tau_h}(t)$, looked upon as functions of τ_h are non-crossing. But this is easily ensured by applying the simultaneous estimation technique exposed in Section 3.1, if needed. A final remark is that an advantage of this method is that the estimation of the variability function comes as part of it, in Step 2.*

4 Estimation of the error variability function

In Section 3 we focused on estimation of the quantile curves $q_\tau(Y(t) \mid \mathbf{X}(t), t)$, and discussed several methods to ensure non-crossing of the estimated curves. In this section we pay attention to estimation **Task 2**: estimating the variability function $V(t)$.

We investigate several methods for estimating $V(t)$. A first method is the adaptation of He's (1997) approach presented in Section 3.4, with the estimator defined in (3.9). Recall that this method relies on Assumptions **(H1)** and **(H2)**.

A second method was already discussed in Remark 2.2, and assumes Assumption **(B)**.

A third alternative method is as follows. Suppose now that the intercept coefficient term is zero, i.e. $\beta_0(t) = 0$, for all t . We refer to this as Assumption **(NI)**. In that case, Model (2.14), implies that

$$Y(t) - \sum_{k=1}^p \beta_k(t) X^{(k)}(t) = V(t)\varepsilon(t) ,$$

and (see also (2.16))

$$\beta_0^\tau(t) = V(t)a^\tau(t) . \tag{4.1}$$

Assuming **(H2)**, we then obtain that

$$q_{0.5} \left(\left| Y(t) - \sum_{k=1}^p \beta_k(t) X^{(k)}(t) \right| \right) = V(t) .$$

Thus under Assumptions **(NI)** and **(H2)**, the function $V(t)$ is identifiable, and its estimation can be achieved via the following two-steps procedure.

Step 1. Obtain the estimated coefficient functions $\widehat{\beta}_k(t)$, for $k = 1, \dots, p$, using the (simultaneous) estimation method of Section 3.1.

Step 2. Use the pseudo-observations from **Step 1**, denoted by $\left| Y(t_{ij}) - \sum_{k=1}^p \widehat{\beta}_k(t_{ij}) X^{(k)}(t_{ij}) \right|$; approximate the unknown variability function $V(\cdot)$ by B-splines, and use a P-splines estimation method with check function $\rho_{0.5}(\cdot)$ to estimate the function $V(\cdot)$, as in (3.10).

Yet, a fourth method is obtained under the Assumption **(NI)** and the assumption

(A2): the τ -th order conditional quantile of the error term $\varepsilon(T)$, given $T = t$, denoted by $a^\tau(t)$ is fully known.

Indeed, from (4.1) it follows that, under Assumptions **(NI)** and **(A2)**, the variability function $V(\cdot)$ is identifiable, and can be estimated by

$$\widehat{V}(t) = \frac{\widehat{\beta}_0^\tau(t)}{a^\tau(t)}.$$

Remark 4.1. *From the discussions in this and previous sections, several sets of sufficient conditions for identifiability of the variability function V were provided:*

- Assumptions **(H1)** and **(H2)** hold;
- Assumption **(B)** holds;
- Assumptions **(NI)** and **(H2)** hold;
- Assumptions **(NI)** and **(A2)** hold.

These are just some examples, and other sets of sufficient conditions are possible.

In the simulation study in Section 6 all simulation models satisfy Assumption **(NI)** (i.e. no intercept function).

5 Simulation study I: Non-crossing property of quantile curves

Computer codes for the methods described in Sections 3 and 4 have been developed and are collected in an available R Package QRegVCM that has been developed by the first author. The package is available at the CRAN website <https://cran.r-project.org/web/packages/> and the codes are also available at <http://wis.kuleuven.be/stat/codes.html>.

In this first simulation study we aim to investigate the performances of the methods ensuring non-crossing estimated quantile curves discussed in Section 3. We therefore consider simulation models in which $V(t) = 1$ (or a constant), and the focus is only on estimation of the quantile functions. We take the general model

$$Y(t_{ij}) = \beta_0(t_{ij}) + \sum_{k=1}^3 \beta_k(t_{ij})X^{(k)}(t_{ij}) + \varepsilon(t_{ij}),$$

and consider different settings for the unknown coefficient functions $\beta_k(\cdot)$, for $k = 0, 1, 2, 3$ and for the error term structure. See Table 1 for all details.

In all simulation models, we take the sample size $n = 100$, and for each subject i , (with $i = 1, \dots, n$) measurements can happen only at the time points $\{0, 1, 2, \dots, 49\}$. From this fixed set of possible time points, each time point (except for the starting time point 0) has a probability of 40% to be skipped. This creates different numbers of repeated measurements N_i for each subject i ($i = 1, \dots, n$). The actual measurement times are generated by adding a $U[0, 0.5]$ random variable to the non-skipped scheduled times.

Table 1: *Homoscedastic models.*

Model	Coefficient functions	Error design
1	$\beta_0(t) = \frac{t^2}{215}$ $\beta_1(t) = \cos\left(\frac{(t-25)\pi}{20}\right)$ $\beta_2(t) = \frac{\sin(\pi t)}{30}$ $\beta_3(t) = -4 + \frac{(20-t)^2}{100}$	$\varepsilon(t) = \frac{\eta(t) - q_{0.5}(\eta(t))}{q_{0.5}(\eta(t) - q_{0.5}(\eta(t)))}$ $\Rightarrow q_{0.5}(\varepsilon(t)) = 0$ $\Rightarrow q_{0.5}(\varepsilon(t)) = 1$
2	$\beta_0(t) = \frac{(t-5)^2}{1500}$ $\beta_1(t) = \frac{1}{4} \cos\left(\frac{t-5}{500}\right)$ $\beta_2(t) = \frac{\sin(\pi t)}{1000}$ $\beta_3(t) = \frac{(15-t)^2}{1000}$	$\varepsilon(t) = \frac{\eta(t) - q_{0.5}(\eta(t)) + \max_{t_{ij}}(\eta(t))}{q_{0.5}(\eta(t) - q_{0.5}(\eta(t)) + \max_{t_{ij}}(\eta(t)))}$ $\Rightarrow q_{0.5}(\varepsilon(t)) = 1$ $\Rightarrow q_{0.5}(\varepsilon(t)) = 1$
3	$\beta_0(t) = \frac{ t-25 ^3}{330}$ $\beta_1(t) = 2 - 3 \cos\left(\frac{(t-25)\pi}{15}\right)$ $\beta_2(t) = 10 + 2 \sin\left(\frac{\pi t}{30}\right)$ $\beta_3(t) = -4 + \frac{(20-t)^2}{100}$	$\varepsilon(t) = \eta(t)$ $\Rightarrow q_{0.5}(\varepsilon(t)) \neq 0$ $\Rightarrow q_{0.5}(\varepsilon(t)) \neq 1$

The covariates $X^{(1)}(t_{ij})$, $X^{(2)}(t_{ij})$ and $X^{(3)}(t_{ij})$ are independent where $X^{(1)}(t_{ij})$ has a standard exponential distribution, $X^{(2)}(t_{ij})$ a standard normal distribution, and $X^{(3)}(t_{ij})$ is uniformly $U[-1, 1]$ distributed.

The error terms are, for each individual, generated from a multivariate normal distribution, followed by a transformation applied to it, as to fulfill certain assumptions (regarding its median value or median of its absolute value). More precisely, we start from generating $\eta(\cdot)$ from a multivariate normal distribution with covariance structure $\text{Cov}(\eta(t_{ij}), \eta(t_{ij'})) = 30 \exp(-|j - j'|)$. We then transform this error, as indicated in Table 1, so that the transformed error $\varepsilon(\cdot)$ has the indicated median properties.

For the P-splines approximation technique we always used 10 equidistant knot points and worked with B-splines of degree 3.

We investigate the finite-sample performances of the methods (see Section 3):

- INDIVIDUAL: the individual quantile regression estimator using P-splines (see Section 2; and Andriyana *et al.* (2014));
- SIMULTANEOUS (UNWEIGHTED): the simultaneous objective function method of Section 3.1, with $W_h = 1$;
- SIMULTANEOUS (WEIGHTED): the simultaneous objective function method, with data-driven weights as in (3.4); see Section 3.1;
- QUANTILE SHEET: the adaptation of the quantile sheet method of Schnabel and Eilers (2013a); see Section 3.2;
- STEPWISE: the complete Up and Down procedure as described in Section 3.3;
- AHE APPROACH: the adaptation of He's approach; see Section 3.4.

For each simulation model, we simulate 200 samples, and summarize results based on these. Simulation results illustrating the problem of crossing estimated quantile curves are provided in Section C.1 in the Supplementary material.

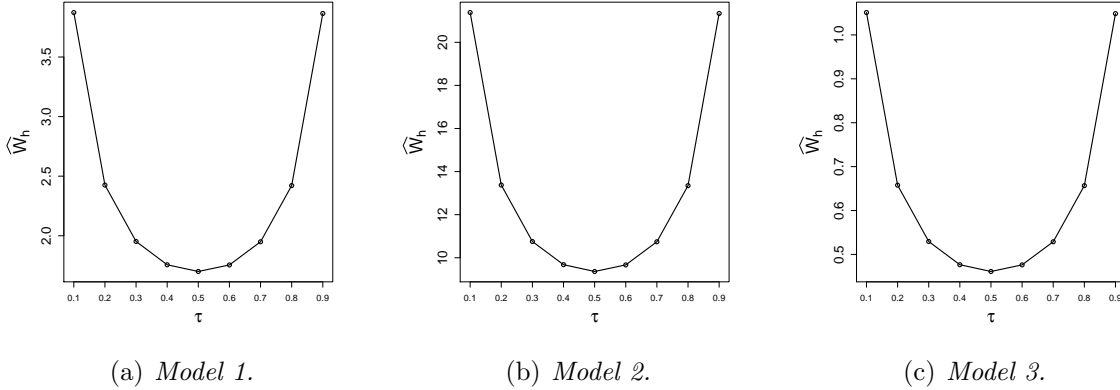


Figure 5.1: *Homoscedastic models: the average of the weights \widehat{W}_h over 200 simulations.*

We first report on the performances of the six methods, of which five are designed for dealing with the issue of crossing curves. For each sample s ($s = 1, \dots, N_{\text{sim}}$), with $N_{\text{sim}}=200$ the number of simulations, consisting of observations $(Y_{ij}, \mathbf{X}_{ij}, t_{ij})$ we calculate the empirical root mean integrated squared errors defined as

$$\text{RMISE}(\widehat{q}_{\tau_h}^{(s)}(\cdot)) = \left(\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} (\widehat{q}_{\tau_h}^{(s)}(Y_{ij} | \mathbf{X}_{ij}, t_{ij}) - q_{\tau_h}(Y_{ij} | \mathbf{X}_{ij}, t_{ij}))^2 \right)^{1/2}, \quad h = 1, \dots, H. \quad (5.1)$$

Note that we do not write the dependence on the sample s of the second term within brackets, for not making the notation too heavy. But it is good to realize that due to the evaluation in the observations, this term also changes with the sample.

Before summarizing (a selection of) the results for each of the models, we look a bit closer to the simultaneous method with the data-driven weights. In Figure 5.1 we depict the mean of the data-driven weights \widehat{W}_h (over the 200 simulations) for each of the models. Note that the largest weights are assigned to the most extreme-order quantiles, and that the median gets in general a smaller weight. All estimated weights have a similar shape when looked upon as a function of τ .

In Figure 5.2 boxplots report on the values of $\text{RMISE}(\widehat{q}_{\tau_h}^{(s)}(\cdot))$, for the 200 simulations, for each of the six methods, for Model 1. Results for Models 2 and 3 are given in Figure C.1 in Section C.2 in the Supplementary material. From Figure 5.2 we can see that the unweighted and weighted simultaneous methods perform quite comparable in this (homoscedastic) model setting. In addition, their performances are quite close to that of the quantile sheet method, although the latter method has some difficulty for more extreme-order quantiles (e.g. $\tau = 0.1$ and $\tau = 0.9$). In all three models the stepwise method and the AHe approach perform less well (with exception of higher order quantiles

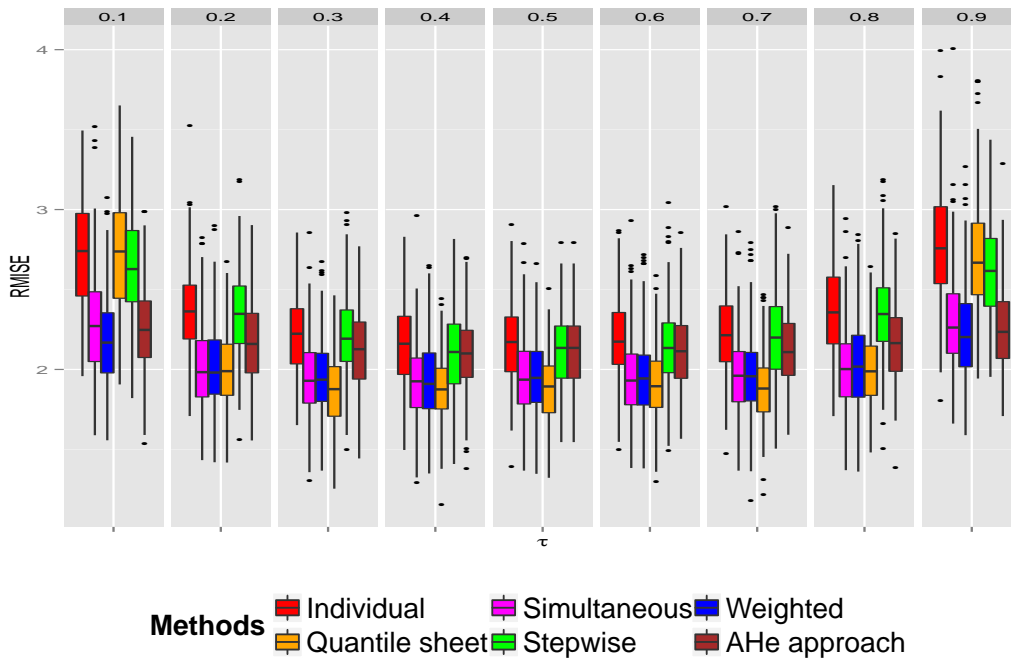


Figure 5.2: *Homoscedastic model 1: boxplots of $RMISE(\hat{q}_{\tau_h})$ over 200 simulations.*

for the AHe approach), with performances that are in fact quite close to that of the individual quantile estimation method.

For a more detailed evaluation of the quality of the estimated quantile curves, we look at some representative estimates (among the 200 estimated quantile curves). We present results for the sample that gave a median performance for the individual estimation method (according to the empirical root mean integrated squared error criterion in (5.1), but after averaging over all considered quantile orders τ). Figure 5.3 (d)—(i) depicts these representative estimates for each of the six methods, for Model 1. For that specific sample we also plot, in Figures 5.3 (a), the values $Y_{ij} = Y(t_{ij})$, when fixing the value of the covariate vector $\mathbf{X}(\cdot)$ to be the maximum over all observed values (i.e. $\max_{i,j} \mathbf{X}(t_{ij})$). Moreover, the true quantile curves are plotted in Figure 5.3 (b). Similar pictures for the simulation results for Models 2 and 3 are provided in respectively Figure C.2 and Figure C.3 in the Supplementary material.

Note that all methods designed for preventing crossing curves are indeed taking care of the crossings that are noticeable in the panels (d) of Figure 5.3, Figure C.2 and Figure C.3. The performances of the unweighted and the weighted simultaneous methods as well as of the quantile sheet method are quite comparable (except for may be the most extreme quantile orders). In Section C.3 of the Supplementary material we report on average computing times for each of the discussed six methods. Since the computational cost is quite high for the quantile sheet method, we do not include this method in the second part of the simulation study.

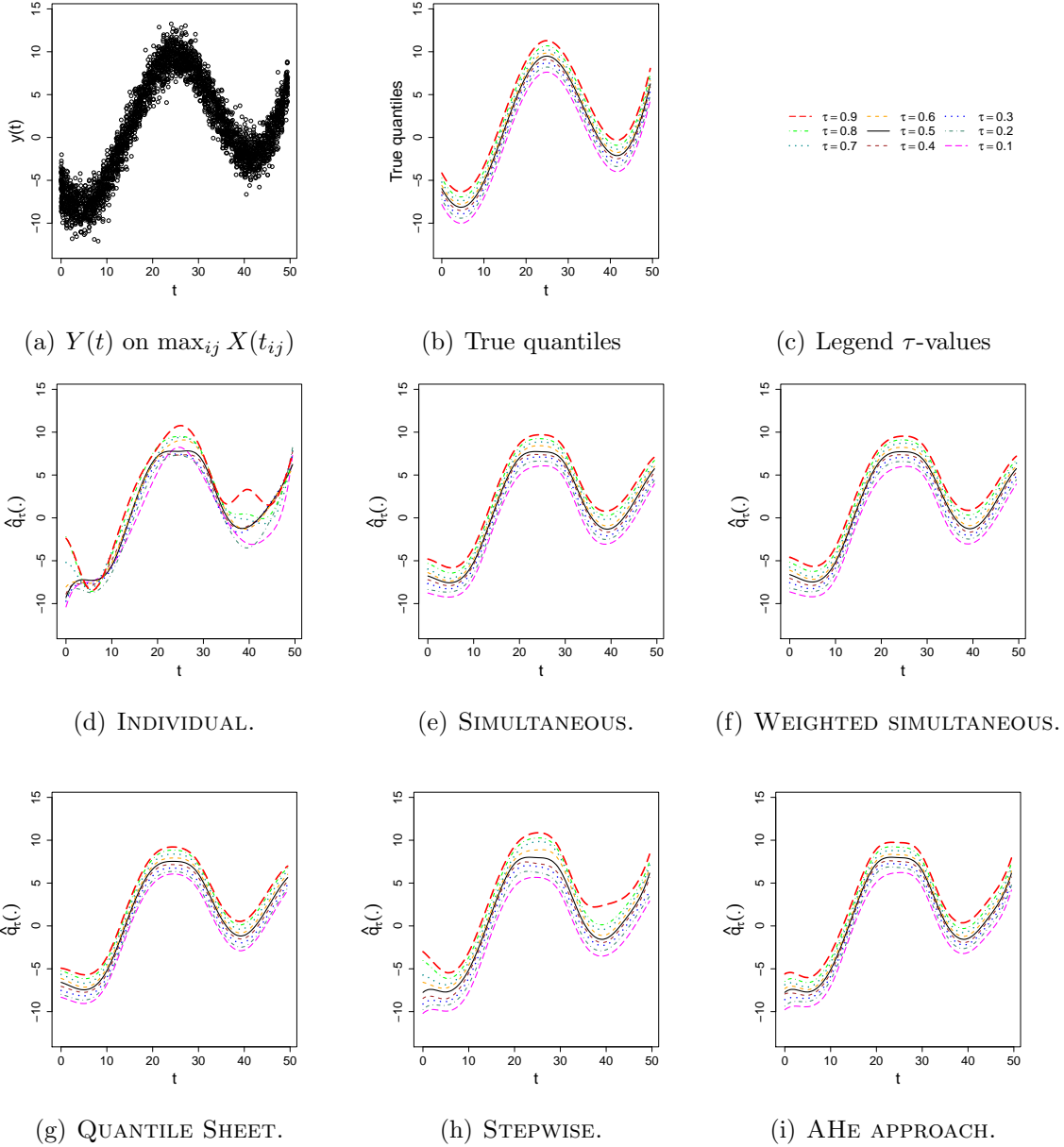


Figure 5.3: *Model 1. Top panel: Scatter plots in case of median performance of $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$ for the individual method. Other panels: performances of $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$ for all six methods, applied to the maximum values of the covariates.*

6 Simulation study II: heteroscedasticity

Here we consider a simple heteroscedastic model

$$Y(t_{ij}) = \sum_{k=1}^3 \beta_k(t_{ij})X^{(k)}(t_{ij}) + V(t_{ij})\varepsilon(t_{ij}),$$

with $V(\cdot) \geq 0$, and as indicated in Table 2. For identifiability reasons regarding the function $V(\cdot)$, we do not include an intercept coefficient function (i.e. no $\beta_0(\cdot)$). Further,

the covariates and the error terms are generated as in Section 5. We draw 200 times samples of sizes $n = 100$ from the various models.

Table 2: *Simple heteroscedastic models.*

Model	Coefficients	Error design	$V(t)$
1	$\beta_1(t) = \cos\left(\frac{(t-25)\pi}{1000}\right)$ $\beta_2(t) = \sin\left(\frac{\pi t}{30}\right) + 3$ $\beta_3(t) = -4 + \frac{(20-t)^2}{1000}$	$\varepsilon(t) = \frac{\eta(t) - q_{0.5}(\eta(t))}{q_{0.5}(\eta(t) - q_{0.5}(\eta(t)))}$ $\Rightarrow q_{0.5}(\varepsilon(t)) = 0$ $\Rightarrow q_{0.5}(\varepsilon(t)) = 1$	$V(t) = \frac{1}{5}\sqrt{t}$
2	$\beta_1(t) = \cos\left(\frac{(t-25)\pi}{100}\right)$ $\beta_2(t) = \sin\left(\frac{\pi t}{100}\right)$ $\beta_3(t) = \frac{(20-t)^2}{100}$	$\varepsilon(t) = \frac{\eta(t) - q_{0.5}(\eta(t)) + \max_{t_{ij}}(\eta(t))}{q_{0.5}(\eta(t) - q_{0.5}(\eta(t)) + \max_{t_{ij}}(\eta(t)))}$ $\Rightarrow q_{0.5}(\varepsilon(t)) = 1$ $\Rightarrow q_{0.5}(\varepsilon(t)) = 1$	$V(t) = \frac{(50-t)^{2/3}}{5}$
3	$\beta_1(t) = \cos\left(\frac{(t-25)\pi}{100}\right)$ $\beta_2(t) = \sin\left(\frac{\pi t}{30}\right) + 13$ $\beta_3(t) = -4 + \frac{(20-t)^2}{100}$	$\varepsilon(t) = \frac{\eta(t) - q_{0.5}(\eta(t))}{q_{0.5}(\eta(t) - q_{0.5}(\eta(t)))}$ $\Rightarrow q_{0.5}(\varepsilon(t)) = 0$ $\Rightarrow q_{0.5}(\varepsilon(t)) = 1$	$V(t) = \frac{(25-t)^2 + 3t}{100}$

In this part we focus on the estimation of the variability function $V(\cdot)$, and investigate the performances of two methods for estimating this quantity:

- AHE-BASED METHOD: the estimator defined in (3.9);
- TWO-STEPS METHOD: the two-steps alternative estimation method described in Section 4 (utilizing (4.1)).

For a given simulated sample s ($s = 1, \dots, N_{\text{sim}}$, with $N_{\text{sim}} = 200$) the performance of a method is evaluated via the empirical root mean integrated squared error, defined as

$$\text{RMISE}(\widehat{V}^{(s)}(\cdot)) = \left(\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\widehat{V}^{(s)}(t_{ij}) - V(t_{ij}) \right)^2 \right)^{1/2}.$$

Table 3: *Estimation of $V(\cdot)$: the averages (and standard deviations) of $\text{RMISE}(\widehat{V}(t))$ over 200 simulations, for Methods 1 and 2.*

Method	Model 1	Model 2	Model 3
1. AHe-based method	0.69 (0.17)	14.78 (0.28)	2.79 (0.80)
2. Two-steps method	0.69 (0.17)	0.78 (0.19)	2.75 (0.81)

Table 3 summarizes the simulation results for Models 1–3. Presented in the table are the mean and the standard deviations of the empirical root mean integrated squared errors for the AHe-based method (Method 1) and the two-steps method (Method 2).

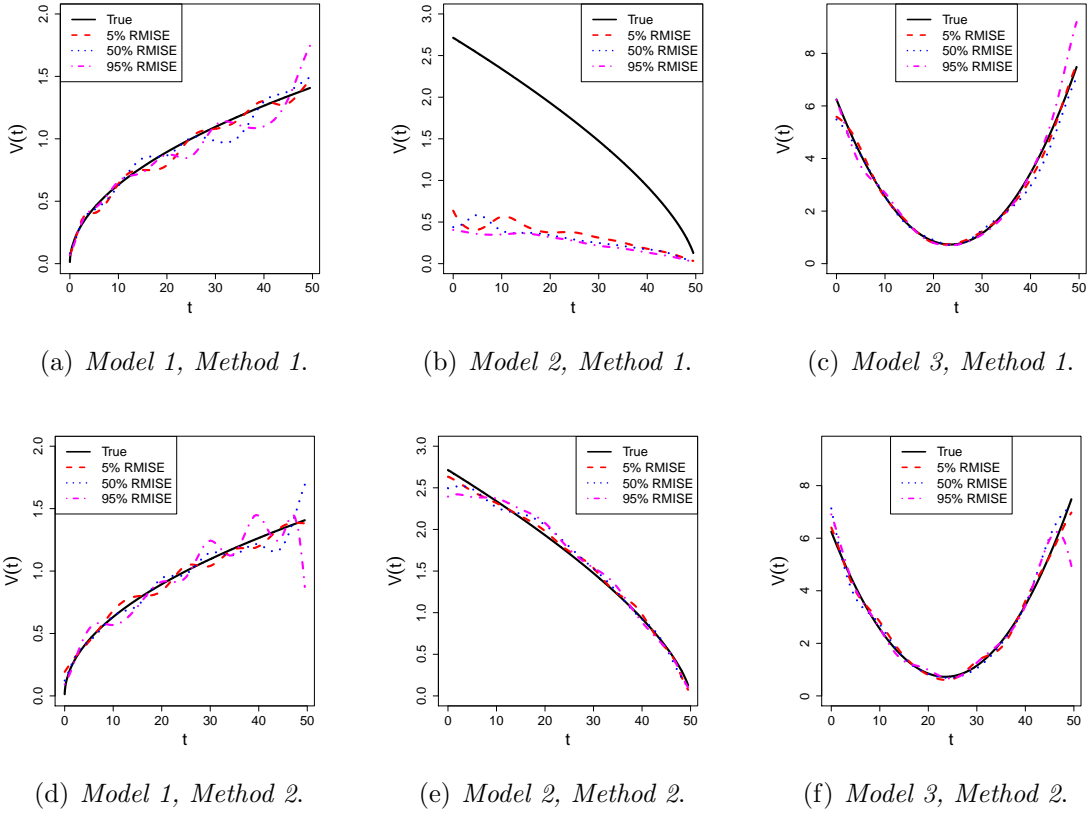
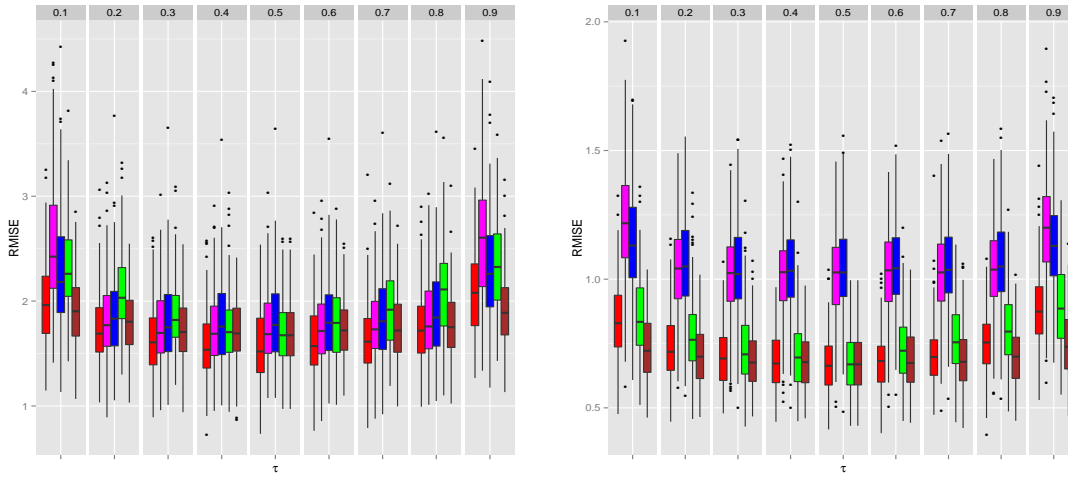


Figure 6.1: *Simple heteroscedastic models: three representative estimates for $\widehat{V}(\cdot)$ (solid curve), according to the 5-th, 50-th, and 95-th percentiles of $\text{RMISE}(\widehat{V}(t))$; results for Method 1 and Method 2. Top panels: results for Method 1; Bottom panels: results for Method 2.*

Depicted in Figure 6.1 are some representative estimated variability curves for each of the methods. These representative estimates correspond to the 5-th, 50-th, and 95-th percentiles of the evaluation criterion RMISE over the 200 simulated samples (and this for each method and each model).

Note, from Table 3 and Figure 6.1, that the performances of the AHe-based method and the two-steps method are quite comparable for Models 1 and 3, but that the AHe-based method seems to fail for Model 2. Recall that for the latter model the median of the error term does not equal zero, and hence Assumption **(H1)** is not satisfied here. The two-steps method does not rely on this assumption, and hence it performs well.

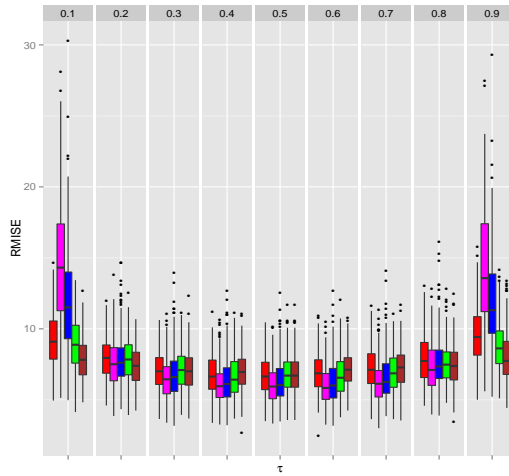
In what follows we use the two-steps method to estimate the variability function $V(\cdot)$ when needed. This is for example the case when applying the weighted simultaneous estimation method. To save space we do not present averages of the data-driven weights \widehat{W}_h , since they look very similar to the plots given in Figure 5.1. This is to be expected, given that the error terms $\varepsilon(t)$ have the same distribution, and given the good quality of the estimation of $V(\cdot)$.



(a) Heteroscedastic model 1

(b) Heteroscedastic model 2

Methods ■ Individual ■ Simultaneous ■ Weighted
■ Stepwise ■ AHe approach



(c) Heteroscedastic model 3

Figure 6.2: *Simple heteroscedastic models: boxplots of $RMISE(\hat{q}_{\tau_h})$ over 200 simulations for all models.*

To finish we present some pictures that illustrate the performances in estimation of the quantile curves in case of simple heteroscedastic models. For each of the five methods, and all three models, we present the boxplots of the RMISE evaluation criterion (see (5.1)). See Figure 6.2. Note that in case of these simple heteroscedastic models, the weighting can lead to improvements in the simultaneous method, in particular for the higher-order quantiles. In such cases the weighted version is also less variable (despite the extra estimation of the variability function). The stepwise method and the AHe approach have a more or less comparable performance. In terms of criterion (5.1) there is no

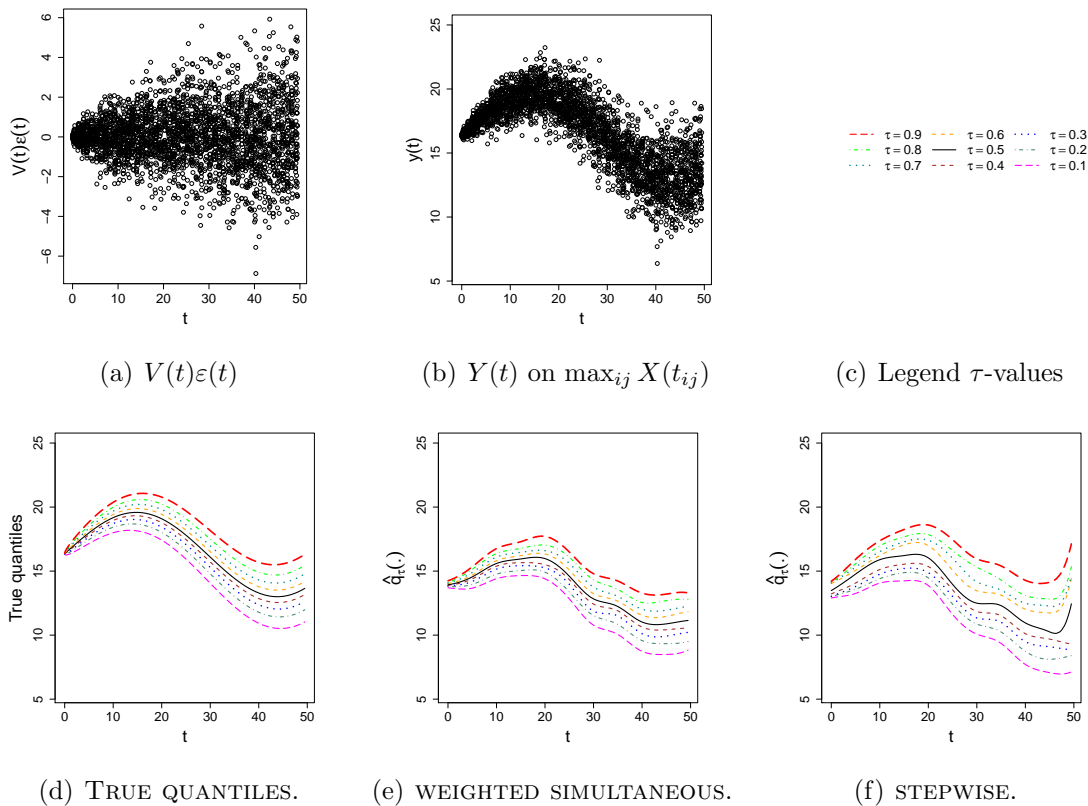


Figure 6.3: *Simple heteroscedastic model 1. Top panel: Scatter plots on the median performance of $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$. Bottom panel: estimated quantile curves $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$ applied to the maximum values of the covariates, with median RMISE-performances. Results are for the weighted simultaneous method and the stepwise method.*

real gain when comparing the individual quantile estimation method with the methods designed to prevent for crossing curves. The stepwise method and the AHe approach method give, in case of Model 2, better performances than the simultaneous methods (see Figure 6.2(b)). Recall that for the simultaneous methods one either takes $W_h = 1$ for all h (in the unweighted simultaneous method) or uses the data-driven weights (see (3.4); in the weighted simultaneous method), in which obviously the quality of estimation of the variability function $V(\cdot)$ will play a role. For Model 2, a possible poor estimation of the variability function influences the performance of the weighted simultaneous quantile estimation method. These choices of weights explain why both simultaneous methods perform, in this model, considerably worse than the other three methods, which do not rely on choices of weights and/or on an estimator for the variability function. Figure C.1(a) in Section C.2 of the Supplementary material presents the boxplots for the conditional quantile estimators for the homoscedastic Model 2. Comparing Figures 6.2(b) and C.1(a) we see in the latter a good performance of both simultaneous methods.

Finally we depict for each of the models: a typical scatter plot of the terms $V(t)\varepsilon(t)$; and a typical scatter plot of the observed response, as a function of time. These can

be found in the top panels of Figures 6.3 and 6.4, and in Figure C.4 in Section C.2 of the Supplementary material. For brevity we only present representative (median performance) estimates of the quantile curves for two (out of the five) methods: the weighted simultaneous method (which involves the estimated variability function) and the stepwise method (see Section 3.3). See the bottom panels of Figures 6.3, 6.4 and Figure C.4.

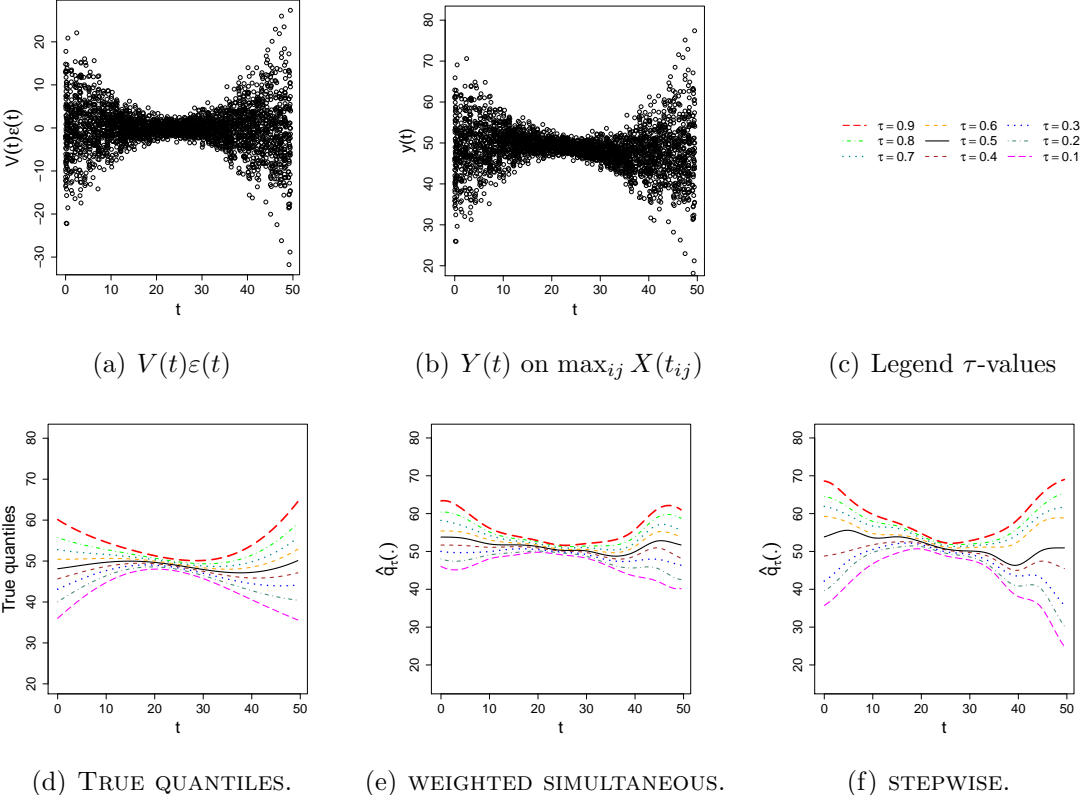


Figure 6.4: *Simple heteroscedastic model 3. Top panel: Scatter plots on the median performance of $\hat{q}_{\tau_n}(Y(t) | \mathbf{X}(t), t)$. Bottom panel: estimated quantile curves $\hat{q}_{\tau_n}(Y(t) | \mathbf{X}(t), t)$ applied to the maximum values of the covariates, with median RMISE-performances. Results are for the weighted simultaneous method and the stepwise method.*

In Section C.2.3 of the Supplementary material we present simulation results for two additional models: (i) Model 1 but where the error is lognormal distributed; and (ii) a simulation model that is not of a location-scale type. From these additional simulation results we can conclude that also for such very difficult settings the discussed methods continue to work for the quantile estimation task. Obviously the AHe approach inherently assumes a location-scale model and is inapplicable for estimating a variability function outside this modelling structure.

7 Real data example: CD4 data

The data that we consider here are a subset of the data collected in the Multicenter AIDS Cohort Study. This study concerns the investigation of the evolvement of an AIDS infection over time, and involves 283 homosexual men who became HIV-positive in a period of 6 years (1984 to 1991). For each of the subjects in the study, one has repeated measurements on physical examination characteristics, on laboratory results, on CD4 cell counts and on CD4 percentages. The number of repeated measurements ranges, over the different individuals, from 1 to 14, with a median of 6 and a mean of 6.57. The number of distinct time points, across the individuals, is 59. More details on the data can be found in Kaslow *et al.* (1987). In this application we investigate the effect of the covariates cigarette smoking status, pre-HIV infection CD4 cell percentage and age at HIV infection, on the quantiles of the CD4 percentage after infection (the response variable). Denote by t_{ij} the time in years (after HIV infection) of the j -th measurement on the i -th individual and by Y_{ij} the i -th individual's CD4 percentage at time point t_{ij} . The covariates are $X_i^{(1)}$ the smoking status of the i -th individual (1 or 0 if the individual ever or never smoked cigarettes), $X_i^{(2)}$ the centered (with the median) age at HIV infection for the i -th individual and $X_i^{(3)}$ the centered (with the median) pre-infection CD4 percentage.

In Andriyana *et al.* (2014) the median quantile curve for the post infection CD4 cell percentage was estimated. See that paper for, among others, the estimated baseline function, as well as the estimated coefficient functions for the median regression curve. With the methods developed here, we can ensure that various estimated quantile curves do not cross, and we can also further investigate the variability in the data, using the simple heteroscedastic modelling framework. As in Andriyana *et al.* (2014), we use B-splines of degree 3 with 10 equidistant knots on the time interval and take differencing order 1 everywhere.

We first report on estimating several quantile curves (for $\tau = \{0.1, 0.2, \dots, 0.9\}$). We present, for brevity, the results for two of the developed methods: the weighted simultaneous method and the AHe approach. The estimated quantile curves are depicted in Figures 7.1 (b) and (c). As we can see in Figure 7.1 (a), there is a clear issue of crossing estimated curves when one uses the individual objective function method. This problem is solved when using any of the other methods, as is seen from Figures 7.1 (b) and (c).

Next we can wonder about the variability aspects. If we want to make the assumption that the error term has median zero (assumption **(H1)**) then we should include an intercept term, since from Andriyana *et al.* (2014), we have learned that the estimated baseline function differs from zero. We apply the AHe-based method for estimating the variability function. The estimated curve is plotted in Figure 7.2 (a). In Figure 7.2 (b) we plot the standardized residuals, together with estimated regression quantile curves (for quantile orders 0.05, 0.25, 0.50, 0.75 and 0.95). As can be seen the estimated median

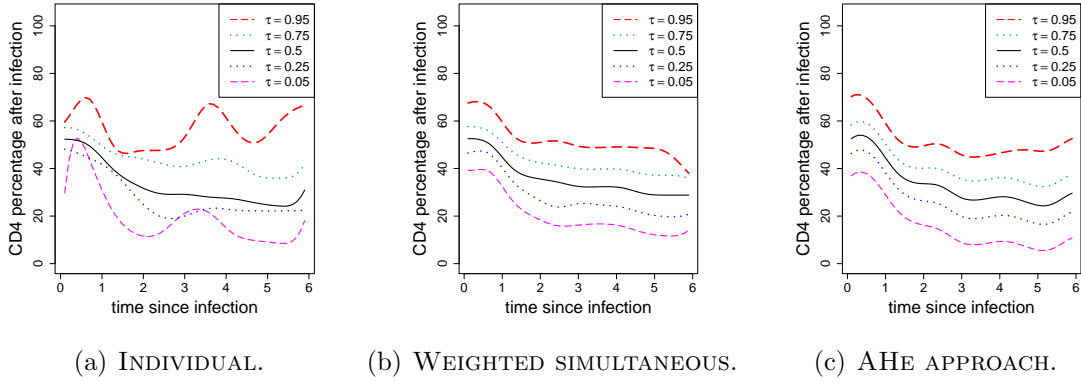


Figure 7.1: CD_4 data. Estimated CD_4 percentage quantile curves for $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ at the maximum values of all covariates using the methods: (a) Individual quantile estimation, (b) (weighted) Simultaneous quantile estimation, and (c) AHe approach.

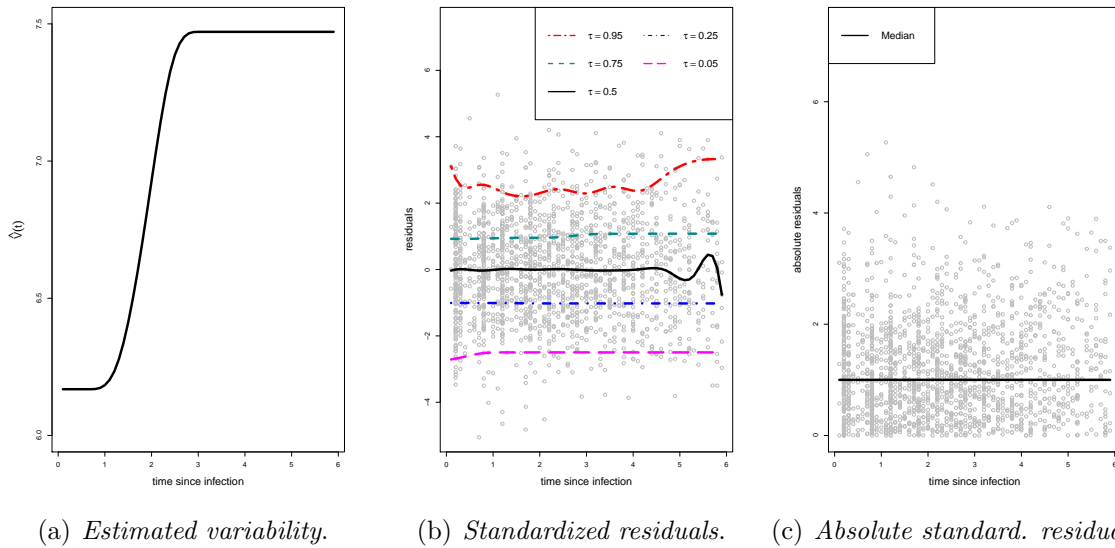


Figure 7.2: CD_4 data. (a). The estimated variability curve using the AHe-approach; (b) & (c). Median regression applied to the standardized residuals (using the estimated variability function) and to the absolute value of the standardized residuals.

regression curve is close to the constant 0. Furthermore the quantile curves are close to being horizontal, showing no obvious dependence anymore on the covariates. Figure 7.2 (c) depicts the absolute value of the standardized residuals, with a median regression estimator applied to it, revealing indeed a constant median value of 1.

In Section D in the Supplementary material the methods developed in this paper are used to analyze air pollution data.

8 Further discussion

A first aim of this paper was to develop quantile estimation methods for varying-coefficient models, based on P-splines approximations, that ensure that the estimated quantile curves do not cross. An alternative approach could be to apply rearrangement techniques such as these used by Chernozhukov *et al.* (2009).

A second aim of the paper was to deal with quantile regression allowing for heteroscedasticity. The focus in this paper was on a location-scale type of modelling. Distinguishing between location models, location-scale models and more general heteroscedastic models is an important issue when it comes to building and choosing statistical models. Tests for testing homoscedasticity versus heteroscedasticity, specification tests, testing for errors being independent of covariates, among others, have been developed in a nonparametric regression context without time varying-coefficients. See, for example, Einmahl and Van Keilegom (2008), Neumeyer (2009) and Birke *et al.* (2013). Developing such tests in the context of varying-coefficient models would be of interest, and is a possible direction for future research.

Concerning the P-splines approximation we did not discuss the choice of the order of the differencing operators, and took them always one or two in the simulation studies and real data applications. Of course, this choice together with the choice of other procedure parameters, such as the number of knot points and the degree of the B-spline, do matter in practice. For a detailed study on the complex interplay between these procedure parameters, see for example Gijbels and Verhasselt (2010), and references therein.

Acknowledgements.

This research is supported by the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy), and project GOA/12/014 of the Research Fund of the KU Leuven. The authors thank the Editor, an Associate Editor and the anonymous reviewers for their very valuable comments which led to a considerable improvement of the paper.

References

- Aldrin, M. and Hobaek Haff, I. (2005). Generalised additive modelling of air pollution, traffic volume and meteorology. *Atmospheric Environment*, **39**, 2145-2155.
- Alexander, M., Harding, M. and Carlos Lamarche, C. (2011). Quantile regression for time-series-cross-section data. *International Journal of Statistics and Management System*, **6**, 47-72.
- Andriyana, Y., Gijbels, I. and Anneleen Verhasselt. A. (2014). P-splines quantile regression estimation in varying coefficient models. *Test*, **23**, 153-194.
- Antoniadis, A., Gijbels, I. and Nikolova, M. (2011). Penalized likelihood regression for generalized linear models with nonquadratic penalties. *The Annals of the Institute of Statistical Mathematics*, **63**, 585-615.

- Birke, M., Neumeier, N. and Volgushev, S. (2013). The independence process in conditional quantile location-scale models and an application to testing for monotonicity. *Technical Report 2013-03*, University of Hamburg. <http://preprint.math.uni-hamburg.de/public/ims.html>
- Bondell, H.D., Reich, B.J. and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, **97**, 825–838.
- Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, **96**, 559–575.
- Cole, T.J. (1988). Fitting smoothed centile curves to reference data. *Journal of The Royal Statistical Society, Series A*, **151**, 385–418.
- Cole, T.J. and Green, P.J. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine*, **11**, 1305–1319.
- Currie, I.D., Durban, M. and Eilers, P.H.C. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of The Royal Statistical Society, Series B*, **68**, 259–280.
- Dette, H. and Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society, Series B*, **70**, 609–627.
- Eilers, P.H.C., Currie, I.D. and Durban, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis*, **50**, 61–76.
- Eilers, P.H.C. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–102.
- Einmahl, J.H.J. and Van Keilgeom, I. (2008). Specification tests in nonparametric regression. *Journal of Econometrics*, **143**, 88–102.
- Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. John Wiley and Sons, New York.
- Gijbels, I. and Verhasselt, A. (2010). P-splines regression smoothing and difference type of penalty. *Statistics and Computing*, **20**, 499–511.
- Guo, J., Tian, M. and Zhu, K. (2012). New efficient and robust estimation in varying-coefficient models with heteroscedasticity. *Statistica Sinica*, **22**, 1075–1101.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of The Royal Statistical Society, Series B*, **55**, 757–796.
- He, X. (1997). Quantile curves without crossing. *The American Statistician*, **51**, 186–192.
- Honda, T. (2004). Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inference*, **121**, 113–125.
- Jiang, J., Jiang, X. and Song, X. (2014). Weighted composite quantile Regression estimation of DTARCH models. *Econometric Journal*, **17**, 1–23.
- Jiang, R., Qian, W.-M. and Zhou, Z.-G. (2016). Single-index composite quantile regression with heteroscedasticity and general error distributions. *Statistical Papers*, **57**, 185–203.
- Karlsson, M. and Lindmark, A. (2014). truncSP: An R Package for Estimation of Semi-Parametric Truncated Linear Regression Models. *Journal of Statistical Software*, **57**, Issue 14.

- 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**, 310–318.
- Kim, M.-O. (2007). Quantile regression with varying coefficients. *The Annals of Statistics*, **35**, 92–108.
- Koenker, R. (1984). A note on L-estimates for linear models. *Statistics & Probability Letters*, **2**, 323–325.
- Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**, 74–89.
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.
- Koenker, R. and Ng, P. (2005). A Frisch-Newton algorithm for sparse quantile regression. *Acta Mathematicae Applicatae Sinica (English Series)*, **21**, 225–236.
- Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometric*, **157**, 396–408.
- Liu, Y. and Wu, Y. (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constrains. *Journal of Nonparametric Statistics*, **23**, 415–437.
- Neumeier, N. (2009). Testing independence in nonparametric regression. *Journal of Multivariate Analysis*, **100**, 1551–1566.
- Oftedal, B., Walker, S.-E., Gram, F., McInnes, H. and Nafstad, P. (2009). Modelling long-term averages of local ambient air pollution in Oslo, Norway: evaluation of nitrogen dioxide, PM10 and PM2.5. *International Journal of Environment and Pollution*, **36**, 110–126.
- Portnoy, S., and Koenker, R. (1997). The gaussian hare and the laplacian tortoise: computability of squared -error versus absolute-error estimators. *Statistical Science*, **12**, 279-300.
- Schnabel, S.K. and Eilers, P.H.C. (2013a). Simultaneous estimation of quantile curves using quantile sheets. *AStA Advances in Statistical Analysis*, **97**, 77–87.
- Schnabel, S.K. and Eilers, P.H.C. (2013b). A location-scale model for non-crossing expectile curves. *Stat*, **2**, 171–183.
- Şentürk, D. and Müller, H.-G. (2010). Functional varying coefficient models for longitudinal data. *Journal of American Statistical Association*, **105**, 1256–1264.
- Wang, H.J., Zhu, Z. and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *The Annals of Statistics*, **37**, 3841–3866.
- Wu, Y. and and Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, **19**, 801–817.
- Yang, H. and Liu, H. (2016). Penalized weighted composite quantile estimators with missing covariates. *Statistical Papers*, **57**, 69-88.
- Zhao, Z. and Xiao, Z. (2014). Efficient regressions via optimally combining information. *Econometric Theory*, **30**, 1272–1314.
- Zou, H. and Yuan, M. (2008). Regularized simultaneous model selection in multiple quantiles regression. *Computational Statistics & Data Analysis*, **52**, 5296–5304.

Supplementary material to the paper

Quantile regression in varying-coefficient models: non-crossing quantile curves and heteroscedasticity

by

Andriyana, Y., Gijbels, I. and Verhasselt, A.

This supplementary material consists of the following additional material.

1. In Section A we establish that non-crossing of the estimates quantile curves is guaranteed for a specific covariates and time setting;
2. Section B presents the data-driven choices for the tuning parameters ($\lambda^{(NC)}$, and $\lambda_{\tau_h, k}^{(PS)}$, for $h = 1, \dots, H$ and $k = 0, \dots, p$) that is used for all methods;
3. In Section C we provide more details on the results of the extensive simulation study, and we also report on results for two additional simulation models.
4. The methods developed in the paper are applied to a second data example, the Air Pollution data, in Section D.

A Guaranteed non-crossing at a specific covariates and time setting

In this section we establish that for the P-splines estimation method non-crossing of the estimated quantile curves is guaranteed for a specific fixed mean value of the covariates and the time (under some conditions).

Recall that the optimization problem to solve is

minimize

$$\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau} \left(Y_{ij} - \mathbf{X}'_{ij} \boldsymbol{\beta}(t_{ij}) \right) + \sum_{k=0}^p \sum_{\ell=d_k+1}^{m_k} \lambda_k |\Delta^{d_k} \alpha_{k\ell}|$$

with respect to $\boldsymbol{\alpha}$; or equivalently (see (2.8)) minimize

$$\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau} \left(Y_{ij} - \mathbf{X}'_{ij} \boldsymbol{\beta}(t_{ij}) \right) + \sum_{k=0}^p \lambda_k J_k(\boldsymbol{\beta}_k), \quad (\text{A.1})$$

with respect to the set of functions $\boldsymbol{\beta}$; where we denoted the penalty term applied to each function with a generic notation J_k , where the dependence on k comes from the fact that the difference order d_k may be different for the various component functions $\boldsymbol{\beta}_k$.

Important to note is that $J_k(\beta_k) \geq 0$. The solution of the minimization problem depends of course on the specific value of τ considered. The above equivalence only formally holds if the approximation in (2.8) is an equality, meaning that the function $\beta_k(\cdot)$ belongs to the space spanned by the B-splines functions $B_{k1}(\cdot), \dots, B_{km_k}(\cdot)$, for $k = 0, \dots, p$. In that case, the penalization is also superficial.

Theorem A.1 states that the non-crossing of the conditional quantile estimate is guaranteed in a specific fixed value for $(\mathbf{X}(t), t)$ that corresponds to some mean value of the observed $(X(t_{ij}), t_{ij})$ where the mean is obtained via the B-splines basis functions.

Theorem A.1. *If $\beta_k(\cdot)$ in Model (2.14) belongs to the space spanned by the B-splines functions $B_{k1}(\cdot), \dots, B_{km_k}(\cdot)$, for $k = 0, \dots, p$, then it holds that*

$$\text{for all } 0 \leq \tau_1 < \tau_2 \leq 1: \quad \widehat{q}_{\tau_1}(Y(t^*) \mid \mathbf{X}(t^*), t^*) \leq \widehat{q}_{\tau_2}(Y(t^*) \mid \mathbf{X}(t^*), t^*), \quad (\text{A.2})$$

where $(\mathbf{X}(t^*), t^*)$ is such that $\mathbf{X}'(t^*)\mathbf{B}(t^*) = \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{X}'_{ij}\mathbf{B}(t_{ij})$.

Proof. The proof is along the same lines as the proof of Theorem 2.5 of Koenker (2005), but needs several adaptations for the varying-coefficient model setting.

Let $0 \leq \tau_1 < \tau_2 \leq 1$. We start by evaluating the first term in (A.1), and look at the differences of this first term for the two values of τ considered. Using the definition of ρ_τ in (2.10) we get, for a given set of $\boldsymbol{\beta}$ functions,

$$\begin{aligned} & \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \left[\rho_{\tau_2} \left(Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij}) \right) - \rho_{\tau_1} \left(Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij}) \right) \right] \\ &= \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \left[\tau_2 (Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij}))^+ + (1 - \tau_2) (Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij}))^- - \tau_1 (Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij}))^+ \right. \\ & \quad \left. - (1 - \tau_1) (Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij}))^- \right] \\ &= \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \left[(\tau_2 - \tau_1) (Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij}))^+ + [(1 - \tau_2) - (1 - \tau_1)] (Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij}))^- \right] \\ &= (\tau_2 - \tau_1) \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij})). \end{aligned}$$

In conclusion:

$$\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_1} \left(Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij}) \right) + (\tau_2 - \tau_1) \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{ij} - \mathbf{X}'_{ij}\boldsymbol{\beta}(t_{ij}))$$

$$= \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_2} \left(Y_{ij} - \mathbf{X}'_{ij} \boldsymbol{\beta}(t_{ij}) \right). \quad (\text{A.3})$$

Denote by $\widehat{\boldsymbol{\beta}}^\tau$ the solution of minimization problem (A.1). Then

$$\begin{aligned} & \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_1} \left(Y_{ij} - \mathbf{X}'_{ij} \widehat{\boldsymbol{\beta}}^{\tau_1}(t_{ij}) \right) + \sum_{k=0}^p \lambda_k J_k(\widehat{\boldsymbol{\beta}}_k^{\tau_1}) \\ & \leq \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_1} \left(Y_{ij} - \mathbf{X}'_{ij} \widehat{\boldsymbol{\beta}}^{\tau_2}(t_{ij}) \right) + \sum_{k=0}^p \lambda_k J_k(\widehat{\boldsymbol{\beta}}_k^{\tau_2}). \end{aligned}$$

Adding to both sides the term $(\tau_2 - \tau_1) \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{ij} - \mathbf{X}'_{ij} \boldsymbol{\beta}^{\tau_2}(t_{ij}))$ and applying twice equation (A.3), we obtain

$$\begin{aligned} & \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_1} \left(Y_{ij} - \mathbf{X}'_{ij} \widehat{\boldsymbol{\beta}}^{\tau_1}(t_{ij}) \right) + (\tau_2 - \tau_1) \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{ij} - \mathbf{X}'_{ij} \boldsymbol{\beta}^{\tau_2}(t_{ij})) + \sum_{k=0}^p \lambda_k J_k(\widehat{\boldsymbol{\beta}}_k^{\tau_1}) \\ & \leq \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_1} \left(Y_{ij} - \mathbf{X}'_{ij} \widehat{\boldsymbol{\beta}}^{\tau_2}(t_{ij}) \right) + (\tau_2 - \tau_1) \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{ij} - \mathbf{X}'_{ij} \boldsymbol{\beta}^{\tau_2}(t_{ij})) + \sum_{k=0}^p \lambda_k J_k(\widehat{\boldsymbol{\beta}}_k^{\tau_2}) \\ & = \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_2} \left(Y_{ij} - \mathbf{X}'_{ij} \widehat{\boldsymbol{\beta}}^{\tau_2}(t_{ij}) \right) + \sum_{k=0}^p \lambda_k J_k(\widehat{\boldsymbol{\beta}}_k^{\tau_2}) \\ & \leq \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_2} \left(Y_{ij} - \mathbf{X}'_{ij} \widehat{\boldsymbol{\beta}}^{\tau_1}(t_{ij}) \right) + \sum_{k=0}^p \lambda_k J_k(\widehat{\boldsymbol{\beta}}_k^{\tau_1}) \\ & = \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_1} \left(Y_{ij} - \mathbf{X}'_{ij} \widehat{\boldsymbol{\beta}}^{\tau_1}(t_{ij}) \right) + (\tau_2 - \tau_1) \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{ij} - \mathbf{X}'_{ij} \widehat{\boldsymbol{\beta}}^{\tau_1}(t_{ij})) + \sum_{k=0}^p \lambda_k J_k(\widehat{\boldsymbol{\beta}}_k^{\tau_1}). \end{aligned}$$

Subtracting the first and third term from both sides, we find

$$\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{X}'_{ij} \widehat{\boldsymbol{\beta}}^{\tau_2}(t_{ij}) \geq \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{X}'_{ij} \widehat{\boldsymbol{\beta}}^{\tau_1}(t_{ij}). \quad (\text{A.4})$$

Now recalling that

$$\widehat{\boldsymbol{\beta}}_k^\tau(t_{ij}) = \sum_{\ell=1}^{m_k} \widehat{\alpha}_{k\ell}^\tau B_{k\ell}(t_{ij}; \nu_k).$$

and using the matrix notations in (2.11), the inequality in (A.4) can be rewritten as

$$\left(\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{X}'_{ij} \mathbf{B}(t_{ij}) \right) \widehat{\boldsymbol{\alpha}}^{\tau_2} \geq \left(\sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{X}'_{ij} \mathbf{B}(t_{ij}) \right) \widehat{\boldsymbol{\alpha}}^{\tau_1}$$

which proves (A.2). \square

B Data-driven choice of the tuning parameters

The minimization problems, such as this involving $S_2(\boldsymbol{\alpha})$ in (3.2) require the choice of $H(p+1)+1$ tuning parameters, $\lambda^{(\text{NC})}$ and $\lambda_{\tau_h, k}^{(\text{PS})}$, for $h = 1, \dots, H$ and $k = 0, \dots, p$. A practical data-driven method that was experienced to give a good performance consists of first choosing a global parameter λ and then refining this further, taking into account guesses of the specific irregularities of the unknown univariate functions. More precisely, the two steps of this data-driven tuning parameter selection are

Step 1. First take $\lambda^{(\text{NC})} = \lambda_{\tau_h, k}^{(\text{PS})} = \lambda$ for all $h = 1, \dots, H$ and $k = 0, \dots, p$, and from a given grid of λ -values, choose the λ that minimizes the Schwarz Information Criterion

$$\text{SIC}(\lambda) = \log \left(\frac{1}{nH} \sum_{h=1}^H \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_h} (Y_{ij} - \hat{q}_{\tau_h}(Y_{ij}|\mathbf{X}_{ij}, t_{ij})) \right) + \frac{\log(N)}{2N} p_\lambda,$$

where p_λ is the size of the elbow set \mathcal{E}_λ , where

$$\mathcal{E}_\lambda = \{(i, j, h) : Y_{ij} - \hat{q}_{\tau_h}(Y_{ij}|\mathbf{X}_{ij}, t_{ij}) = 0\},$$

i.e. the set of all fits which led to a perfect fitted value $\hat{q}_{\tau_h}(Y_{ij}|\mathbf{X}_{ij}, t_{ij})$ for the observed response value Y_{ij} .

Denote the resulting choice of λ by $\hat{\lambda}$.

Step 2. Using $\hat{\lambda}$, obtain $\lambda_{\tau_h, k}^{(\text{PS})}$ for all $h = 1, \dots, H$ and $k = 0, \dots, p$, from

$$\hat{\lambda}_{\tau_h, k}^{(\text{PS})} = \hat{\lambda} \left(\mathcal{R}(\hat{\beta}_k^{\tau_h, B}(\cdot)) \right)^{-\kappa}$$

where $\hat{\beta}_k^{\tau_h, B}(t)$ is the quantile regression estimator of $\beta_k^{\tau_h}(t)$ using B-splines (so with individual objective function putting $\lambda_{\tau_h, k}^{(\text{PS})} = 0$), $\mathcal{R}(\hat{\beta}_k^{\tau_h, B}(\cdot))$ is the range of all values $\hat{\beta}_k^{\tau_h, B}((t_{ij}))$, and $\kappa > 0$ is a given number (in our simulations we take $\kappa = 0.5$).

The rationale behind the second step is that a larger value of $\lambda_{\tau_h, k}^{(\text{PS})}$ can be used for a function $\beta_k(\cdot)$ that is smoother (more regular); and the correction factor $\left(\mathcal{R}(\hat{\beta}_k^{\tau_h, B}(\cdot)) \right)^{-\kappa}$ to the data-driven global choice $\hat{\lambda}$ for λ from the first step, aims to achieve this. A very rough guess of the ‘irregularity’ of the unknown function $\beta_k(\cdot)$ is obtained by considering the range of all estimated realized function values $\hat{\beta}_k^{\tau_h, B}((t_{ij}))$. By taking a negative power of this rough guess, we hint for obtaining larger data-driven values for $\lambda_{\tau_h, k}^{(\text{PS})}$ when $\beta_k(\cdot)$ is less irregular.

C More details about the simulation study

C.1 Illustration of the problem of crossing estimated quantile curves

To illustrate the problem of possible crossings of individually estimated quantile curves, we calculate for the individual method the percentages of average number of crossings (in observation points) for each two estimated quantile curves, i.e.

$$\text{percentage of average number of crossings} = \left(\frac{1}{N_{\text{sim}}} \sum_{s=1}^{N_{\text{sim}}} \frac{\#\text{crossing-points}}{N} \right) \times 100\% ,$$

where $N_{\text{sim}} = 200$ denotes the number of simulations, and $N = \sum_{i=1}^n N_i$ the total number of observations. Tables C.1 and C.2 report on these percentages, for Models 2 and 3, for each set of two estimated quantile curves. Obviously the problem is more pronounced for estimated quantile curves with subsequent orders as compared to orders further apart (for example for the $\tau = 0.6$ and $\tau = 0.7$ curves, as compared to the $\tau = 0.1$ and $\tau = 0.7$ curves). The problem of crossing curves is more severe for Model 3 than for Model 2.

Table C.1: *Homoscedastic Model 2: Percentages of average number of crossings in the individually estimated quantile curves, based on 200 simulations.*

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	.	12.27	4.27	1.70	1.11	0.70	0.23	0.16	0.04
0.2		.	10.63	3.46	1.25	0.55	0.21	0.16	0.14
0.3			.	10.82	4.00	1.34	0.46	0.34	0.12
0.4				.	13.19	3.95	1.46	0.70	0.25
0.5					.	11.96	4.14	1.38	0.59
0.6						.	13.02	3.95	1.82
0.7							.	10.33	3.76
0.8								.	11.51
0.9									.

Table C.2: *Homoscedastic Model 3: Percentages of average number of crossings in the individually estimated quantile curves, based on 200 simulations.*

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	.	26.00	15.58	9.96	6.68	4.60	3.19	2.02	1.49
0.2		.	25.99	15.21	9.91	6.77	4.35	2.55	1.88
0.3			.	24.72	16.00	9.81	6.62	4.06	3.31
0.4				.	26.39	16.04	10.53	6.97	4.78
0.5					.	26.19	16.53	10.20	6.77
0.6						.	27.00	15.05	10.18
0.7							.	24.69	15.35
0.8								.	26.77
0.9									.

C.2 Additional simulation results

C.2.1 Simulation study I

Boxplots reporting on the values of $\text{RMISE}(\hat{q}_{\tau_h}^{(s)}(\cdot))$ for Models 2 and 3 are given in Figure C.1. A more detailed impression on the performances for quantile estimation for the six discussed methods can be obtained from Figure C.2 and Figure C.3.

C.2.2 Simulation study II

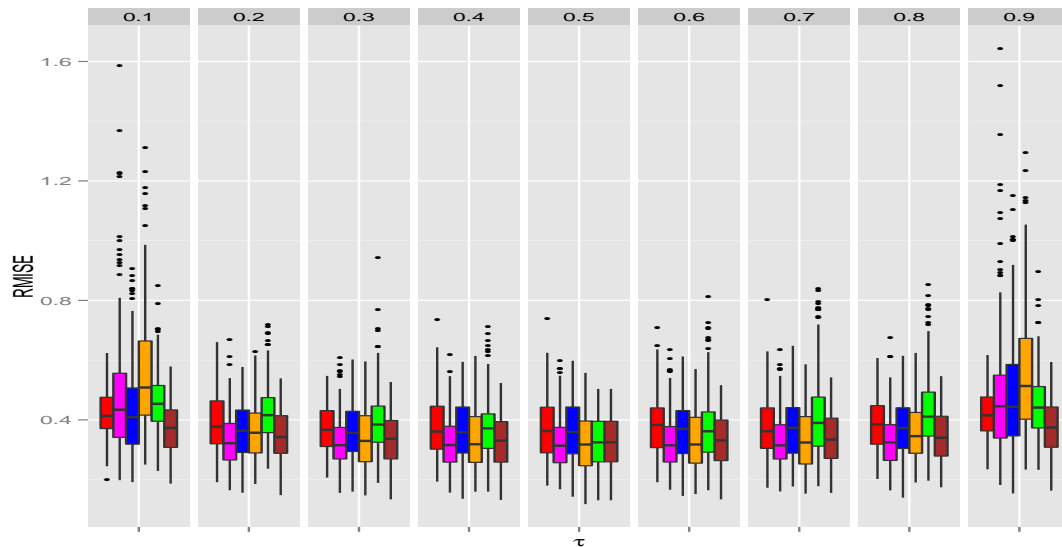
Pictures reporting on the performances of the quantile estimators for the simple heteroscedastic model 2 are to be found in Figure C.4.

C.2.3 Additional simulation study

Asymmetric errors

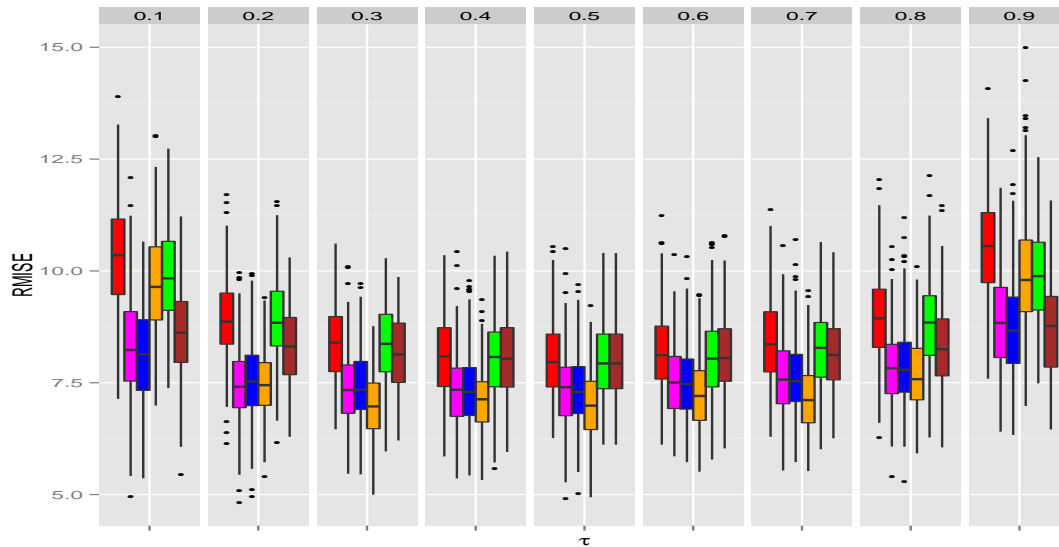
In the simulation studies presented so far the error structure was always coming from a multivariate normal distribution. In order to investigate the performance of the methods in case of an asymmetric error structure, we consider simulation model 1, but now with lognormal errors.

We consider the simple heteroscedastic Model 1 of Section 6, but with variability function $V(t) = 10^{-7}\sqrt{t}$. The error terms are now, for each individual, obtained by generating $\eta(\cdot)$ from a lognormal distribution with covariance structure $\text{Cov}(\eta(t_{ij}), \eta(t_{ij'})) = 30 \exp(-|j - j'|)$. We then transform this error, in correspondence to the model assumptions listed in the third column of Table 2.



(a) *Model 2.*

Methods Individual Simultaneous Weighted
 Quantile sheet Stepwise AHe approach



(b) *Model 3.*

Figure C.1: *Homoscedastic models 2 and 3: boxplots of $RMISE(\hat{q}_{\tau_h})$ over 200 simulations.*

This is a very challenging and difficult simulation setting. As can be seen from Figure C.5 (a) the error structure is a very difficult one. In Figure C.5 (b) we only plot the points $Y(t)$ on a restricted vertical range $([0, 22])$ to allow to see the structure. The maximum value of $Y(t)$ was close to 1000 in this setting. In Figure C.6 we present the boxplots of $RMISE(\hat{q}_{\tau_h})$, but on a vertical log-scale for a better visual impression. As can be seen all methods show a high estimation variance, which is even more pronounced in

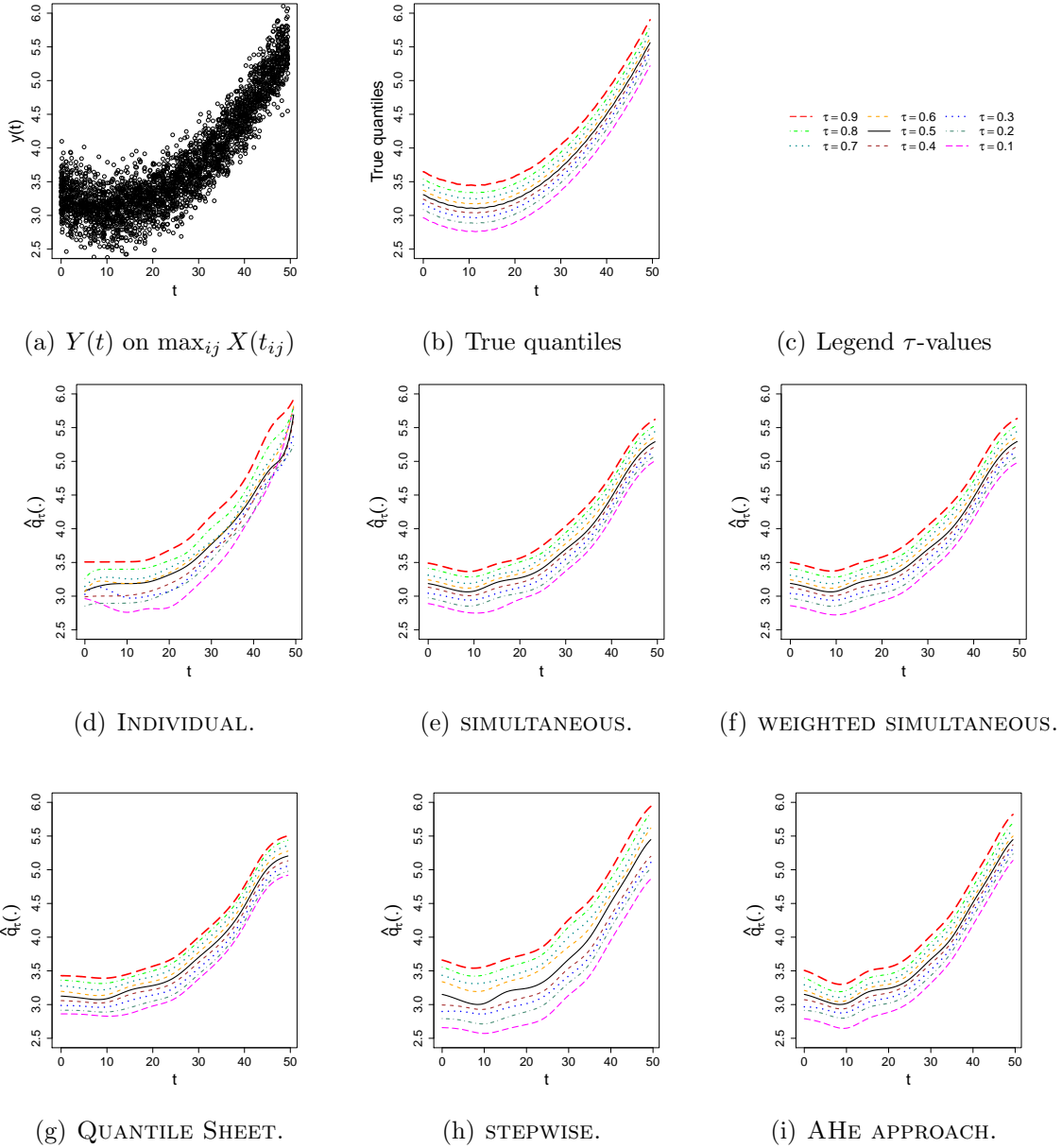


Figure C.2: *Model 2. Top panel: Scatter plots in case of median performance of $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$ for the individual method. Other panels: performances of $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$ for all six methods, applied to the maximum values of the covariates.*

the higher quantiles. Since the highest quantiles tend to get more weights in the weighted simultaneous method (see Section 5), this method suffers in particular from this.

A simulation model that is not a location-scale model

So far we always focussed on location-scale models. However, the methods developed in this paper are also applicable to models that are not of a location-scale type. At least as long as it concerns the estimation of conditional quantile curves.

We therefore consider a gamma model. Denote by \mathbf{G}_i a vector of N_i independent

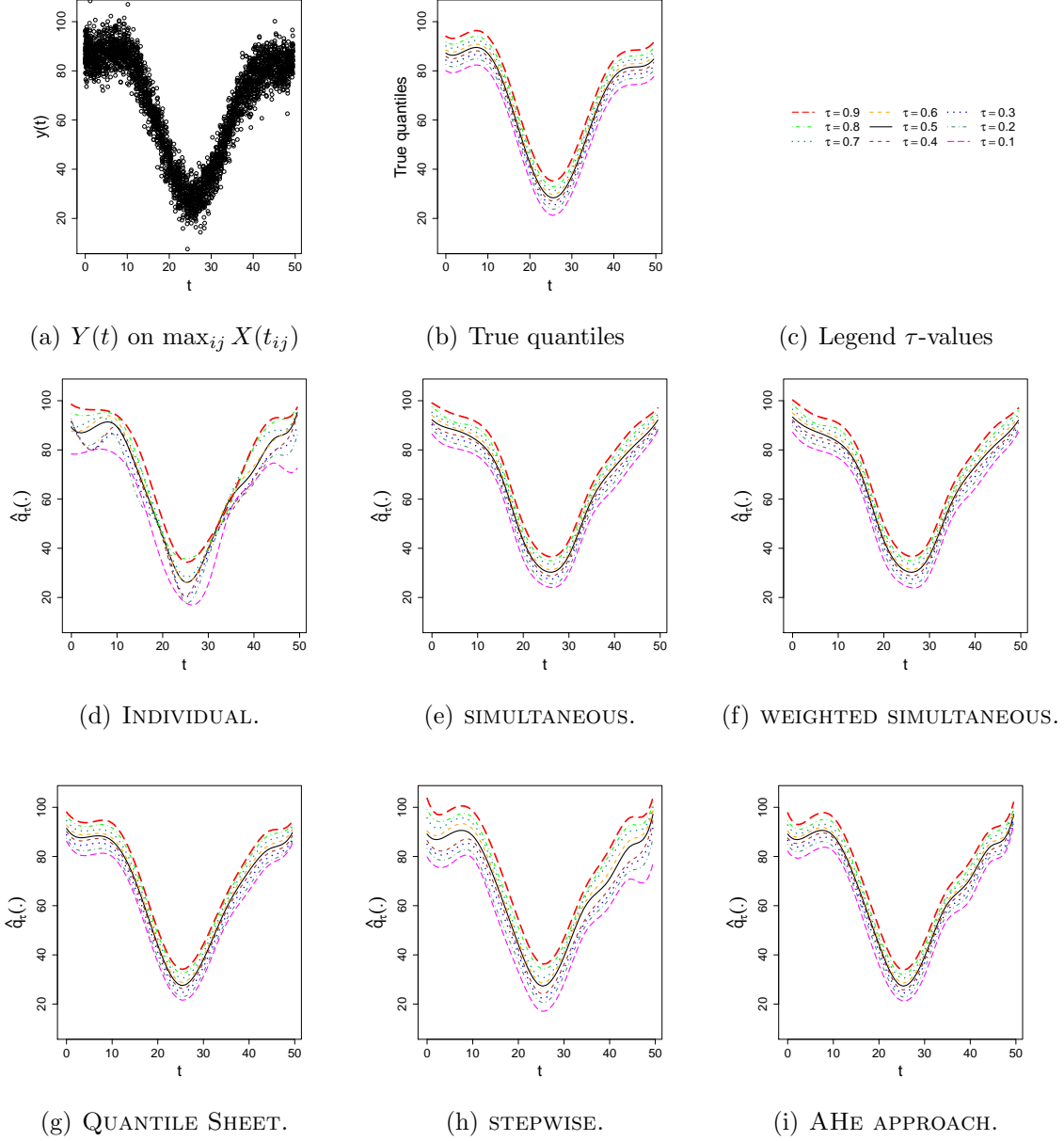


Figure C.3: *Model 3. Scatter plots for the sample with median performance of $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$ for the individual method. Other panels: performances of $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$ for all six methods, applied to the maximum values of the covariates.*

Gamma-distributed variables (corresponding to individual i) with specific shape and scale parameter. Then $\mathbf{Y}_i = \mathbf{R}_i \mathbf{G}_i$, with $\mathbf{R}_i \mathbf{R}_i'$ the variance-covariance matrix (the same as for all other models) with (j, j') element equal to $30 \exp(-|j - j'|)$. The shape parameter is taken to be $\exp\{(X^{(1)}(t_{ij})\beta_1(t_{ij}) + X^{(2)}(t_{ij})\beta_2(t_{ij}) + X^{(3)}(t_{ij})\beta_3(t_{ij}))/10\}$ and the scale parameter equals $\exp\{(X^{(1)}(t_{ij})a_1(t_{ij}) + X^{(2)}(t_{ij})a_2(t_{ij}) + X^{(3)}(t_{ij})a_3(t_{ij}))/10\}$ for a given data point $(\mathbf{X}^*(t_{ij}), t_{ij})$. Herein the functions $\beta_1(\cdot)$, $\beta_2(\cdot)$, and $\beta_3(\cdot)$ are the same as in the simple

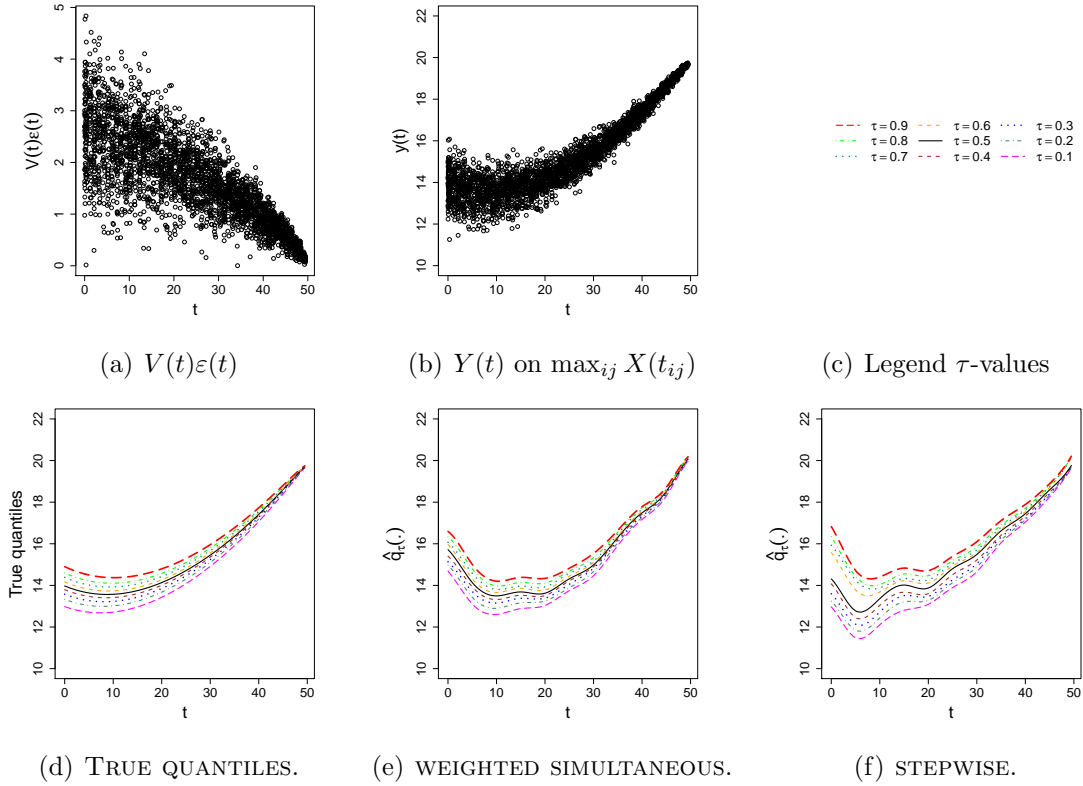


Figure C.4: *Simple heteroscedastic model 2. Top panel: Scatter plots on the median performance of $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$. Bottom panel: estimated quantile curves $\hat{q}_{\tau_h}(Y(t) | \mathbf{X}(t), t)$ applied to the maximum values of the covariates, with median RMISE-performances. Results are for the weighted simultaneous method and the stepwise method.*

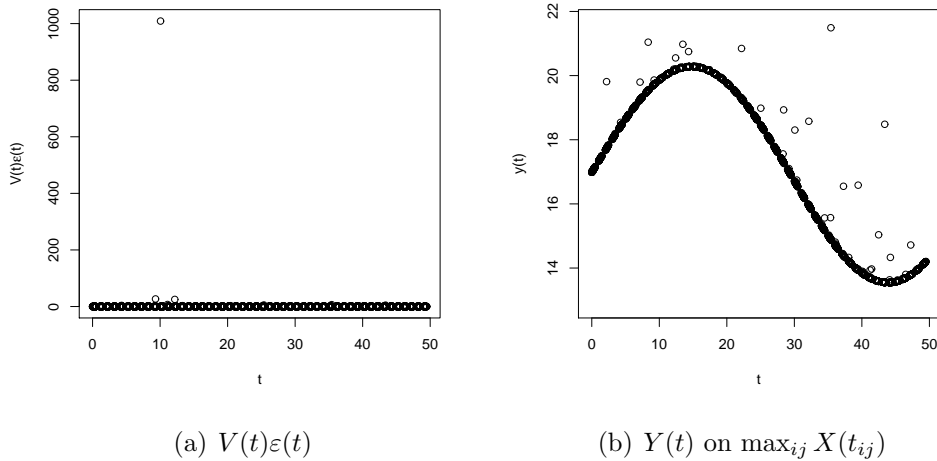


Figure C.5: *Simple heteroscedastic model 1 with lognormal error structure: typical scatter plots of the data.*

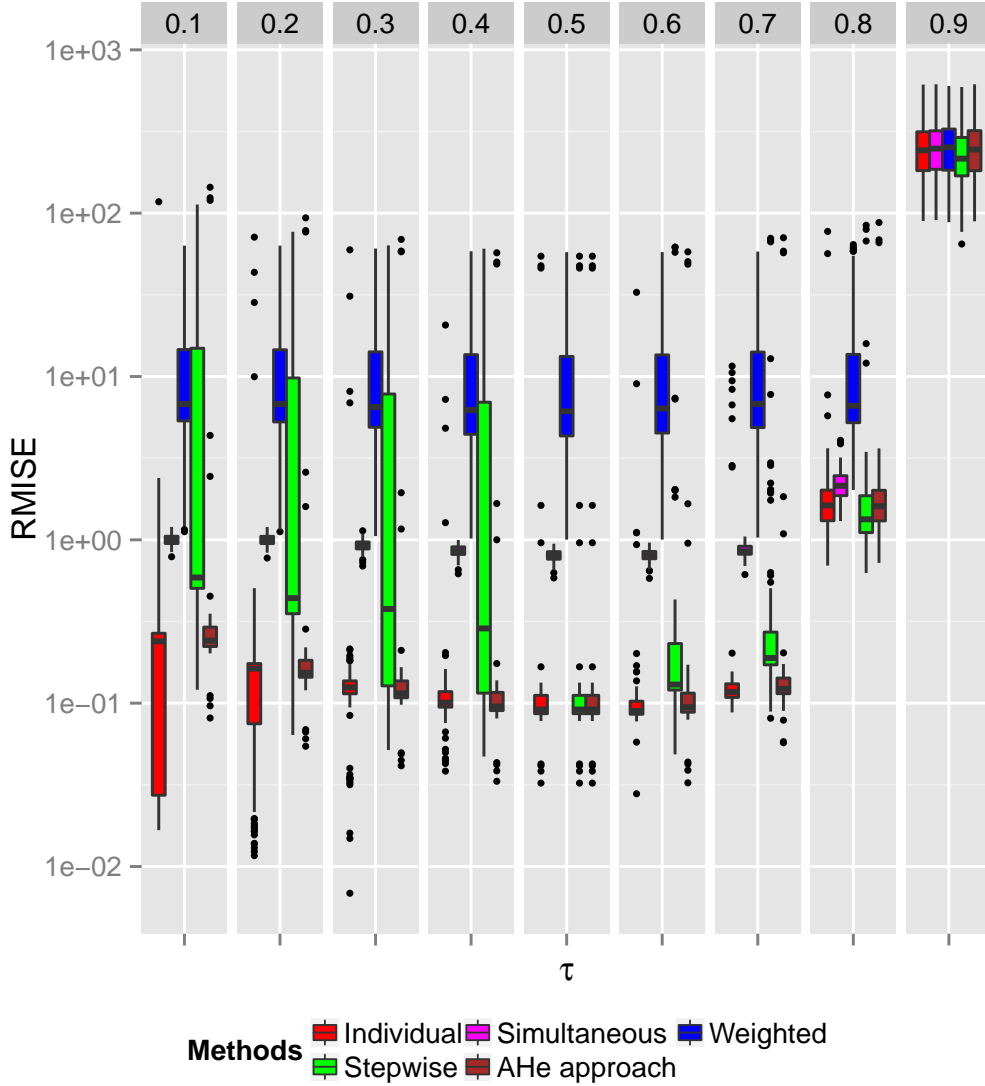


Figure C.6: *Simple heteroscedastic model 1 with lognormal error structure: boxplots of $RMISE(\hat{q}_{\tau_h})$ over 200 simulations. The vertical axis is on a log-scale.*

heteroscedastic model 1, and the functions $a_1(\cdot)$, $a_2(\cdot)$ and $a_3(\cdot)$ are given by

$$a_1(t) = \left(\frac{t}{100}\right)^3 + 1 \quad a_2(t) = 0.1t + 0.5 \quad a_3(t) = 0.5\sqrt{t}.$$

Also this simulation setting is very challenging as can be sensed from Figure C.7 which shows increasingly more widely spread data when moving to the right in the picture. From the boxplots in Figure C.8 (where again we used a log-scale for the vertical axis) it is seen that the quality of quantile estimation decreases with increasing quantile order.

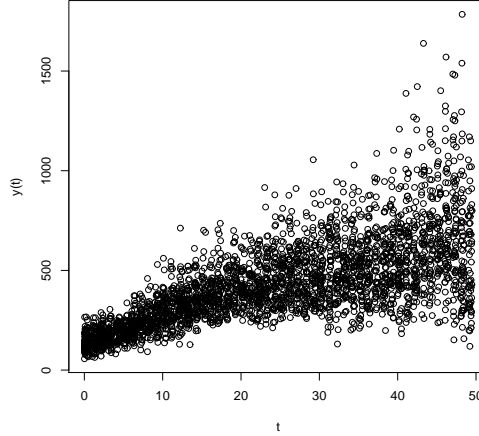


Figure C.7: *Gamma model: Plot of $Y(t)$ versus $\max_{ij} X(t_{ij})$.*

C.3 Computing times

Table C.3 presents average computing times (average over all 200 simulations) for all discussed methods. The computational cost is smallest for the AHe approach, and is largest for the quantile sheet method.

Table C.3: *Computing times (in minutes)*

Methods	Homoscedastic Models			Heteroscedastic Models		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Individual	1.25	1.22	1.38	1.20	1.27	1.37
Simultaneous	1.66	1.64	1.74	2.18	2.68	2.85
Weigthed Simultaneous	3.52	3.89	3.21	4.29	5.23	5.10
Quantile sheet	10.03	8.29	9.38	NOT INCLUDED		
AHe-approach	0.40	0.36	0.44	0.38	0.37	0.43
Stepwise	1.19	1.09	1.22	1.18	1.26	1.29

D Additional real data example: Air Pollution data

The data are a subsample of 500 observations from a data set that originates in a study where air pollution at a road is related to traffic volume and meteorological variables. The data were collected by the Norwegian Public Roads Administration, measured at Alnabru in Oslo, Norway, between October 2001 and August 2003. During each of the 273 days, measurements were performed at different time points (hours), with a range of 1 to 6 measurements per day, and a median of 2 measurements per day. Information about the data can be found at StatLib (<http://lib.stat.cmu.edu>) and also in the `truncSP` R-package (see Karlsson and Lindmark (2014)). Guo *et al.* (2012) analyzed these data considering

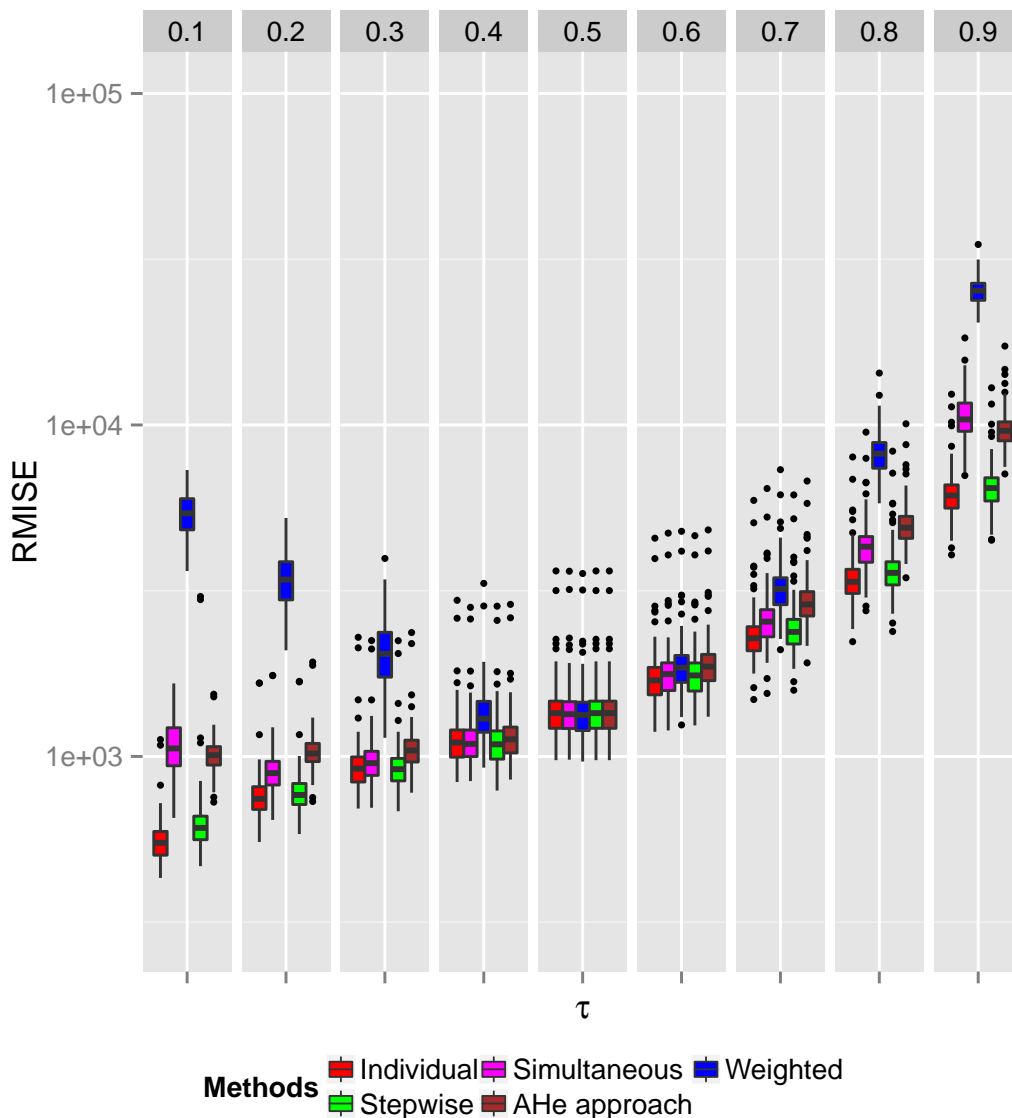


Figure C.8: *Gamma model: boxplots of $RMISE(\hat{q}_{\tau_h})$ over 200 simulations. The vertical axis is on a log-scale.*

also a varying-coefficient model, but in an i.i.d. setting (considering all observations as being independent realisations of a same population random vector). In contrast, in our analysis, we consider the day being the subject and the different measurements per day as the repeated measurements, allowing as such possible dependence between measurements on the same day. The response variable $Y(t)$, consists of hourly values of the logarithm of the concentration of PM10. PM10 is a particle pollution known as “particulate matter”, one of the main constituent air pollutants with negative effect on human health. It is a mixture of solid and liquid droplets with diameter less than 10 micrometers. See for example Aldrin and Hobaek Haff (2005) and Oftedal *et al.* (2009) for studies on air pollution involving PM10 and other particle pollutants. In their analysis Guo *et al.* (2012) considered two covariates, $X_1(t)$ the logarithm of the number of cars per hour and the

second covariate, $X_2(t)$, the wind speed (meters/second). For the illustration here we also consider only these two covariates.

To analyze this data set, we use B-splines of degree 3 with 10 equidistant knot points on the time interval with differencing order 1. The estimated quantile curves, using the individual objective function method, and two of the developed methods (with guaranteed non-crossing estimated quantile curves)— the weighted simultaneous objective function approach and the adaptation of AHe approach, are depicted in Figures D.1 (a)—(c).

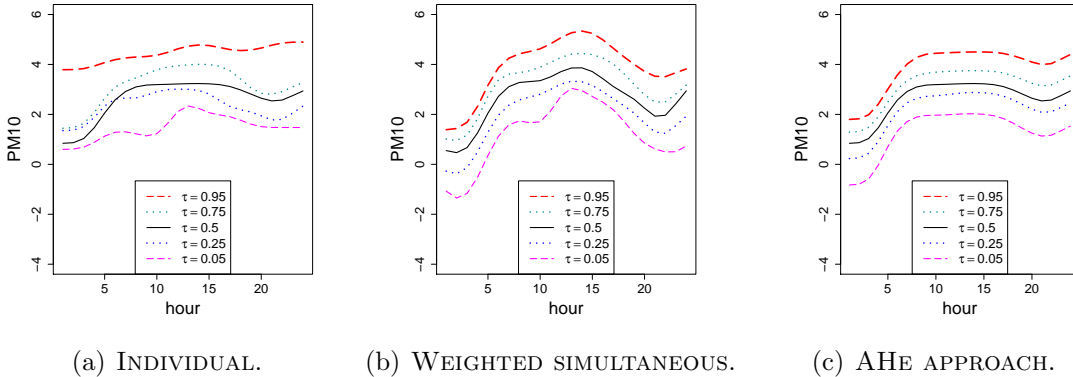


Figure D.1: *Air Pollution Data. Estimated log(concentration of PM10) quantile curves for $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$ at the maximum values of all covariates using the methods: (a) Individual quantile estimation, (b) (weighted) Simultaneous quantile estimation, and (c) AHe approach.*

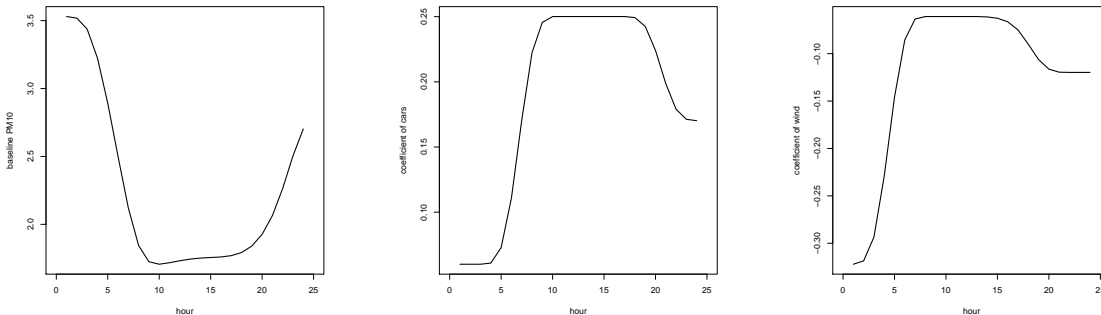


Figure D.2: *Air Pollution Data. The estimated coefficient functions, using the AHe method.*

The estimated coefficient functions $\beta_0(t)$, $\beta_1(t)$ and $\beta_2(t)$ are shown in Figures D.2 (a), (b) and (c), respectively. It should be noted that Guo *et al.* (2012) did not include an intercept term in their model (but standardized, in their i.i.d. setting the Y , X_1 and X_2 variables, not the T variable). This should be kept in mind when comparing the

estimated coefficient functions. The shapes of the estimated coefficient functions $\beta_1(t)$ and $\beta_2(t)$ are somewhat similar (with ours being more smooth), showing the largest effect of both covariates from early morning till late in the afternoon, and with a reduced effect on particle pollution for higher wind speeds (see the negative values of the estimated coefficient $\beta_2(t)$).

In Figure D.3 (a) we plot the estimated variability function, revealing that the estimated variability is lowest between 5 and 7 in the morning, but then increases steadily during the rest of the day.

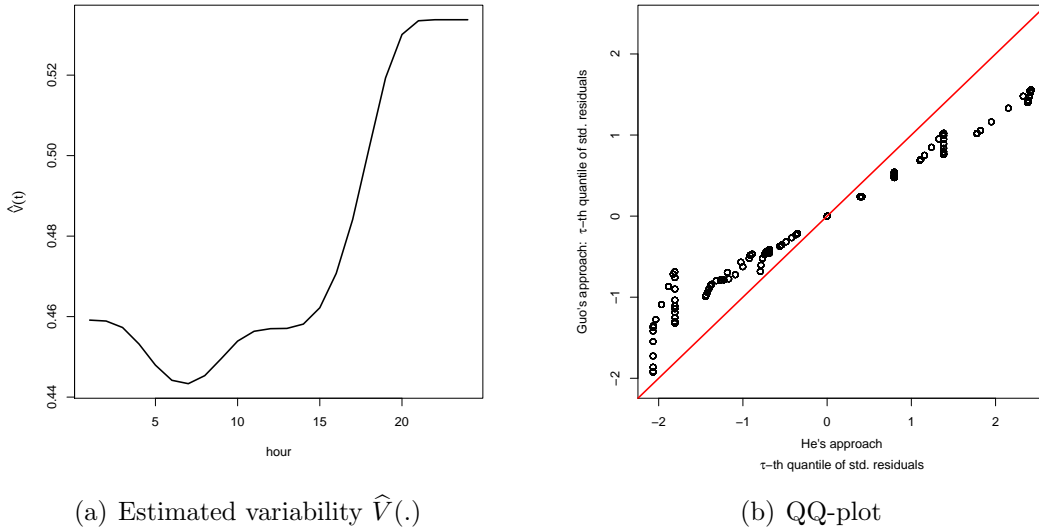


Figure D.3: *Air Pollution Data*. (a) *The estimated variability curve*; (b) *QQ-plots of the τ -th quantile of the standardized residuals using (D.1) versus estimation via the approach in Section 3.4.*

Analyzing these data in a general longitudinal data setting provides some additional insights, compared to previous analysis. Guo *et al.* (2012) also estimate a kind of variability function, but when doing so they assume that the (i.i.d.) errors are standard normally distributed. Translated to our setting this would mean

$$\beta_0^\tau(t) = \beta_0(t) + V(t)q_N^\tau,$$

where q_N^τ denotes the τ th quantile of a standard normal distribution. Based on all known quantiles, an estimator for the variability function is then

$$\hat{V}_{\text{GTZ}}(t) = \frac{1}{H-1} \sum_{h=2}^H \frac{\hat{\beta}_0^{\tau_h}(t) - \hat{\beta}_0^{\tau_{h-1}}(t)}{q_N^{\tau_h} - q_N^{\tau_{h-1}}}. \quad (\text{D.1})$$

In Figure D.3(b) we plot the quantiles of the standardized residuals, using an estimated variability function following Guo *et al.* (2012), relying on (D.1), versus the quantiles of

the standardized residuals using our approach. As can be seen here, the further from the median, the more differences are noticeable between the two quantiles, revealing that the normality assumption is questionable, which is also confirmed by tests for normality.