ORIGINAL PAPER

# Quantile regression in heteroscedastic varying coefficient models

**Y. Andriyana[1,2] · I. Gijbels[1]**

**Abstract** Varying coefficient models are flexible models to describe the dynamic structure in longitudinal data. Quantile regression, more than mean regression, gives partial information on the conditional distribution of the response given the covariates. In the literature, the focus has been so far mostly on homoscedastic quantile regression models, whereas there is an interest in looking into heteroscedastic modelling. This paper contributes to the area by modelling the heteroscedastic structure and estimating it from the data, together with estimating the quantile functions. The use of the proposed methods is illustrated on real-data applications. The finite-sample behaviour of the methods is investigated via a simulation study, which includes a comparison with an existing method.

**Keywords** B-splines · Heteroscedastic error · Longitudinal data · P-splines · Quantile regression · Varying coefficient models

## 1 Introduction

In many applications, the same characteristics for a subject are observed at multiple points in time. In this longitudinal data setting, the multiple observations on the same subject typically are not independent, whereas independence between observations from different subjects still can be considered.

✉ I. Gijbels
 irene.gijbels@wis.kuleuven.be

[1] Department of Mathematics, Leuven Statistics Research Center (LStat), KU Leuven, Leuven, Belgium

[2] Statistics Department, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung, Indonesia

Simple multiple linear regression models are often insufficient to adequately describe such data, and among the more flexible models are varying coefficient models, introduced by Hastie and Tibshirani (1993) and further studied for longitudinal data in Hoover et al. (1998), Huang et al. (2004), Qu and Li (2006), and Şentürk and Müller (2006, 2010), among others. Most papers in the literature focus on estimation of the dynamic influence of the covariates on the mean response. A more overall picture of the conditional distribution function of the response given the covariates can be obtained from investigating quantiles. Quantile regression in a longitudinal data setting has been dealt with in Honda (2004), Kim (2006, 2007), Wang et al. (2009) and Andriyana et al. (2014), among others. Andriyana et al. (2014) considered a homoscedastic quantile regression varying coefficient model, and approximated the unknown coefficient functions by B-splines. They established theoretical properties (including consistency and rate of convergence, and asymptotic distribution) of the estimated coefficient functions.

In regression modelling, the assumptions made on the error term are crucial. Among the important issues is whether the error term has a structure that changes with the covariates or not. In a simple (non-dynamic) covariate setting, this results in looking at heteroscedastic (versus homoscedastic) regression models. In mean regression, various methods have been developed for testing for a homoscedastic error structure (versus a heteroscedastic one). When the aim is a more general knowledge of the conditional distribution of the response, and one resorts to quantile regression the main focus has been so far on a homoscedastic error structure. Recently, Andriyana et al. (2016) allowed for some dynamic structure of the error term in a location-scale model, and developed an estimation method for its variability. This structure, however, does not allow the error to change with (depend on) the covariates, whereas such flexible modelling is sometimes needed as will be seen in Sect. 5 and the example below.

In the UK employment data example of Sect. 5.2, the aim is to investigate how covariates (*wage*, *capital*, *output*) influence the conditional quantile function for the response variable, the *employment*, which is the logarithm of the number of employees in a firm. This influence may change over time, and this is captured by a varying coefficient model. The top and bottom panels of Fig. 1 show the estimated conditional quantile curves as functions of time (for nine different orders of quantiles: 0.1, 0.2, up to 0.9) under two different fixed covariate settings: for the top panel for mean values of the covariates, and for the bottom panel for maximal values of the covariates. Figure 1a,d and b, e are obtained by the proposed methods allowing for heteroscedasticity in the error structure, whereas Fig. 1c, f utilizes a state-of-the-art method. This figure already reveals that some heteroscedasticity is not captured by the latter method (see Fig. 1f, as opposed to Fig. 1d, e). Indeed, some of the covariates turn out to have an influence on the error variability. In Sect. 5.2, we further investigate this heteroscedasticity in the data, as a function of the covariates and time. Figure 2 depicts the estimated influence from the covariate *capital* on a main quantile pattern (in Fig. 2a), as well as on the observed heteroscedasticity (in Fig. 2b). Important to note is that the influence of the covariate *capital* increases over time, but with a more or less constant influence the last 3 years of the studied period. See further Sect. 5.2.
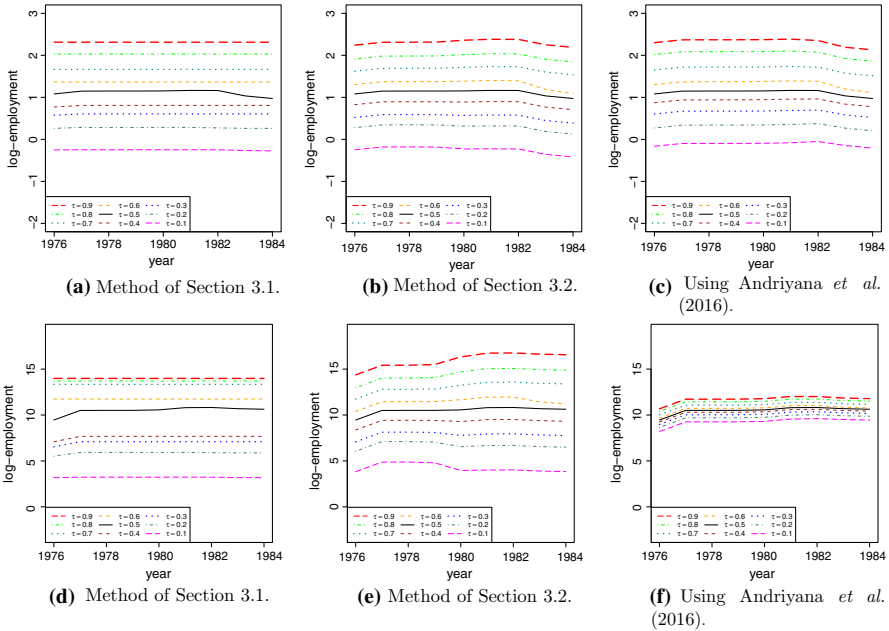
**Fig. 1** UK employment data. Estimated log-employment quantile curves for $\tau = 0.1, 0.2, \ldots, 0.9$ at the mean values (*top panels*) and at the maximum values (*bottom panels*) of all covariates using: **a**, **d** method of Sect. 3.1, **b**, **e** method of Sect. 3.2 , and **c**, **f** AHe $V(t)$ approach
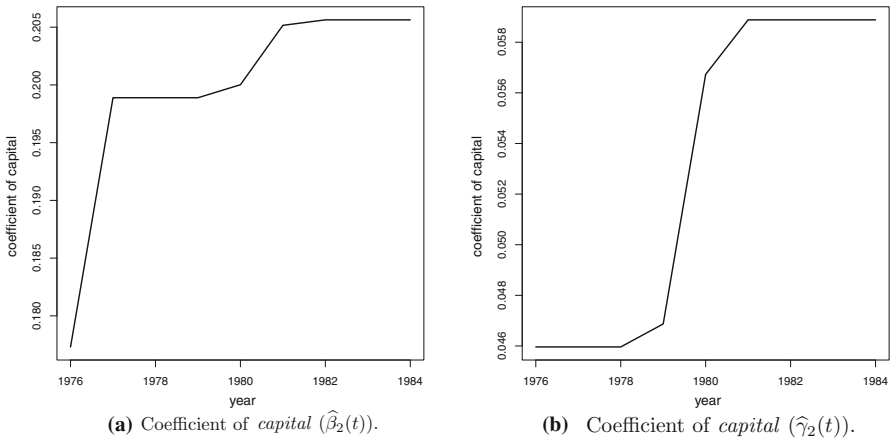


**Fig. 2** UK employment data. The estimated influence of the covariate *capital* on the main quantile function (**a**) and on the heteroscedasticity (**b**), using the estimation method of Sect. 3.2

In this paper, we thus allow for a heteroscedastic error structure, and estimate also the function that describes how the error structure changes with the covariates. This is in contrast to Andriyana et al. (2014) who consider homoscedastic errors and Andriyana et al. (2016) that only allow for a too simplistic heteroscedasticity,

as is illustrated in the example above. The main and innovative contribution of this paper is, thus, to (1) allow for a flexible enough heteroscedasticity modelling in the context of quantile regression for varying coefficient models; (2) to provide a method to estimate not only the conditional quantile function, but *also* the impact of the covariates on this heteroscedasticity structure; and moreover (3) to estimate the entire error variability structure. The paper is further organized as follows. In Sect. 2, the modelling framework is described, and Sect. 3 discusses estimation methods. The finite-sample behaviour of the estimators is examined in Sect. 4 via a simulation study. The practical use of the proposed method is then further illustrated via real-data applications in Sect. 5. In Sect. 6, some concluding remarks and some further discussions are given.

## 2 Varying coefficient models and heteroscedasticity

Suppose we have repeated observations on $(Y(T), (X^{(1)}(T), \ldots, X^{(p)}(T)), T)$, with $T$ being a 'time' variable taking values in the domain $\mathcal{T}$, with $Y(T)$ being the response variable at time $T$, and with $(X^{(1)}(T), \ldots, X^{(p)}(T))$ being the vector of covariates at time $T$.

In a longitudinal data setting, one has repeated measurements on $n$ subjects/individuals. For subject $i$, the repeated measurements occur at time points $t_{i1}, \ldots, t_{iN_i}$, with $N_i$ denoting the number of repeated measurements for the $i$th subject. At time point $t_{ij}$, one observes the response variable $Y(t_{ij})$ and the vector of covariate values $(X^{(1)}(t_{ij}), \ldots, X^{(p)}(t_{ij}))$, which are shortly denoted as $Y_{ij}$ and $(X_{ij}^{(1)}, \ldots, X_{ij}^{(p)})$, respectively. The longitudinal observations of $(Y(T), \mathbf{X}(T), T)$, with $\mathbf{X}(T) = (X^{(0)}(T), X^{(1)}(T), \ldots, X^{(p)}(T))$, where $X^{(0)}(T) = 1$, consist thus of $(Y_{ij}, \mathbf{X}_{ij}, t_{ij})$, $i = 1, \ldots, n$ and $j = 1, \ldots, N_i$, with $\mathbf{X}_{ij} = \left( X_{ij}^{(0)}, \ldots, X_{ij}^{(p)} \right)^{\mathrm{T}}$, and $X_{ij}^{(0)} = 1$. 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) and Sect. 4.

The interest in this paper is in the conditional $\tau$-th order quantile of $Y(T)$ given $\{\mathbf{X}(T), T\}$, with $0 < \tau < 1$, denoted by $q_\tau(Y(T)|\mathbf{X}(T), T)$. Based on these conditional quantiles, one could consider conditional measures of spread such as the conditional median absolute deviation, or the conditional interquartile range (i.e. $q_{0.75}(Y(T)|\mathbf{X}(T), T) - q_{0.25}(Y(T)|\mathbf{X}(T), T)$). Without any modelling assumption, one could attempt to estimate, based on the available longitudinal data, the conditional distribution function of $Y(T)$ given $\{\mathbf{X}(T), T\}$, followed by estimation of conditional quantile functions, and some conditional measures of spread. In a fully nonparametric setting, this would result in a very cumbersome task (e.g. due to a possible impact of several covariates, having an effect on the dimensionality of the conditioning argument). On the other hand, a fully parametric modelling setting is often too restrictive, and leads to important model misspecification errors. Therefore in this paper, we opt for adding some minor model structure, while still keeping enough flexibility. This is done by considering a location-scale type of model. Of main importance is also that such modelling allows us to quantify the possible impact of covariates on the error variability.

More specifically, we consider the following flexible varying coefficient model:

$$Y(T) = \beta_0(T)X^{(0)}(T) + \beta_1(T)X^{(1)}(T) + \cdots + \beta_p(T)X^{(p)}(T) + V(\mathbf{X}(T), T)\,\varepsilon(T),$$
(2.1)

where $\varepsilon(T)$ is independent of $(X^{(1)}(T), \ldots, X^{(p)}(T))$, and $\beta_k(\cdot)$, for $k = 0, \ldots, p$, are the dynamic regression coefficients (unknown univariate functions). We further assume that

$$V(\mathbf{X}(T), T) = \gamma_0(T)X^{(0)}(T) + \gamma_1(T)X^{(1)}(T) + \cdots + \gamma_p(T)X^{(p)}(T) \geq 0, \quad (2.2)$$

for all $T \in \mathcal{T}$, where $\gamma_k(\cdot)$, for $k = 0, \ldots, p$, are unknown univariate functions. Expression (2.2) in fact means that we assume for the error structure itself also a kind of varying coefficient model. We refer to $V(\mathbf{X}(t), t)$ as the *heteroscedasticy function*.

A special case of (2.2) would be when $\gamma_1(T) = \cdots = \gamma_p(T) = 0$, for all $T \in \mathcal{T}$, which means that the error structure simplifies to $\gamma_0(T)\,\varepsilon(T)$, and in fact $V(\mathbf{X}(t), t) \equiv V(t) = \gamma_0(t)$. Such a simple heteroscedastic model was studied in Andriyana et al. (2016), but it turns out to be too simple to describe the heteroscedasticity in some data, as is illustrated by the real-data examples in Sects. 1 and 5.

Model (2.1) is inspired also by previous considerations in the longitudinal data setting. For example, Davidian and Giltinan (1995) consider that the error variance is proportional to the square of the mean response, and in Fortin et al. (2007) the variance is expressed as an unknown power of the mean function. Assuming such parametric relationships between conditional mean and variance is, however, again quite restrictive, in contrast to (2.2) which allows for a very flexible heteroscedastic error structure.

From Model (2.1), the $\tau$-th conditional quantile (with $0 < \tau < 1$) of $Y(T)$ given $\{\mathbf{X}(T), T = t\}$ equals

$$q_\tau(Y(t)|\mathbf{X}(t), t) = \beta_0(t)X^{(0)}(t) + \beta_1(t)X^{(1)}(t) + \cdots + \beta_p(t)X^{(p)}(t) + V(\mathbf{X}(t), t)a^\tau(t),$$
(2.3)

with $a^\tau(t)$ being the $\tau$-th conditional quantile of $\varepsilon(T)$, given $T = t$, i.e. $a^\tau(t) = \inf\{u : P\{\varepsilon(T) \leq u|T = t\} \geq \tau\}$.

Note that the Model (2.1) together with the error structure (2.2) can be re-expressed as:

$$Y(t) = \mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\beta}(t) + \left(\mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\gamma}(t)\right)\varepsilon(t) \tag{2.4}$$

where $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \ldots, \beta_p(t))^{\mathrm{T}}$ and $\boldsymbol{\gamma}(t) = (\gamma_0(t), \gamma_1(t), \ldots, \gamma_p(t))^{\mathrm{T}}$, where $\mathbf{A}^{\mathrm{T}}$ denotes the transpose of a vector or matrix $\mathbf{A}$. Consequently, the $\tau$-th conditional quantile in (2.3) can be re-written as:

$$q_\tau(Y(t)|\mathbf{X}(t), t) = \mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\beta}(t) + \left(\mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\gamma}(t)\right)a^\tau(t) = \mathbf{X}^{\mathrm{T}}(t)\left[\boldsymbol{\beta}(t) + \boldsymbol{\gamma}(t)a^\tau(t)\right].$$
(2.5)

Note that, since $X^{(0)}(t) = 1$, the intercept term in (2.5) equals $\left[\beta_0(t) + \gamma_0(t)\right]a^\tau(t)$. Note, more generally, that the dependence on $\tau$ comes in via the second term between brackets in (2.5).

The quantities that we would like to estimate are:

- the univariate regression functions $(\beta_0(t), \beta_1(t), \ldots, \beta_p(t))^\mathrm{T}$, impacting the main conditional quantile (the systematic part $\mathbf{X}^\mathrm{T}(t)\boldsymbol{\beta}(t)$), and $\boldsymbol{\gamma}(t) = (\gamma_0(t), \gamma_1(t), \ldots, \gamma_p(t))^\mathrm{T}$, impacting the heteroscedasticity;
- subsequently, the heteroscedasticity function $V(\mathbf{X}(t), t)$ defined in (2.2);
- the $\tau$-th order conditional quantile of $\varepsilon(T)$, i.e. $a^\tau(t)$;
- finally, the $\tau$-th order conditional quantile of $Y(T)$, expressed in (2.5).

Given that $V(\mathbf{X}(t), t) \geq 0$, and using that $a^{\tau_1}(t) \leq a^{\tau_2}(t)$ for $0 < \tau_1 < \tau_2 < 1$, we obtain that the population quantile functions are non-crossing under the general heteroscedastic model (2.1),

$$q_{\tau_1}(Y(t)|\mathbf{X}(t), t) \leq q_{\tau_2}(Y(t)|\mathbf{X}(t), t) \quad \text{for } 0 < \tau_1 < \tau_2 < 1, \tag{2.6}$$

for all $(\mathbf{X}(t), t)$ given. Estimated quantile functions should preferably also satisfy this constraint.

It is clear from (2.5) (and the sentences following this displayed expression) that for the vector $\boldsymbol{\gamma}(t)$ to be identifiable in the general heteroscedastic model (2.1), we need to impose some conditions. We briefly discuss some sufficient conditions that ensure identifiability of all quantities to be estimated.

A first sufficient condition for identifiability is having knowledge about two quantiles of $\varepsilon(t)$, say $a^{\tau_1}(t)$ and $a^{\tau_2}(t)$. It is clear from (2.5) that

$$q_{\tau_1}(Y(t)|\mathbf{X}(t), t) - q_{\tau_2}(Y(t)|\mathbf{X}(t), t) = \mathbf{X}^\mathrm{T}(t)\boldsymbol{\gamma}(t)\left[a^{\tau_1}(t) - a^{\tau_2}(t)\right],$$

and, hence, an estimator of $V(\mathbf{X}(t), t)$ is given by

$$\widehat{V}(\mathbf{X}(t), t) = \frac{\widehat{q}_{\tau_1}(Y(t)|\mathbf{X}(t), t) - \widehat{q}_{\tau_2}(Y(t)|\mathbf{X}(t), t)}{a^{\tau_1}(t) - a^{\tau_2}(t)}, \tag{2.7}$$

with $\widehat{q}_\tau(Y(t)|\mathbf{X}(t), t)$ being an estimator for $q_\tau(Y(t)|\mathbf{X}(t), t)$. See also Sect. 3.1. A sufficient condition to ensure identifiability of $V(\mathbf{X}(t), t)$ in the general heteroscedastic model is:

**Assumption (A)**       The $\tau_1$-th and $\tau_2$-th order (with $0 < \tau_1, \tau_2 < 1$ and $\tau_1 \neq \tau_2$) conditional quantile of the error term $\varepsilon(T)$, given $T = t$, are fully known, i.e. $a^{\tau_1}(t)$ and $a^{\tau_2}(t)$ are known for all $t \in \mathcal{T}$.

An alternative set of conditions that ensure identifiability of all quantities involved is inspired by assumptions imposed in classical quantile regression (see, e.g. He 1997):

**Assumption (B)**       **(B1)** the conditional median of the error term $\varepsilon(t)$ equals zero: i.e. $a^{0.5}(t) = 0$ for all $t$.
**(B2)** the conditional median of the absolute value of the error term $\varepsilon(t)$ equals one: $q_{0.5}(|\varepsilon(t)|) = 1$.
Indeed, under Assumption (B1), it follows from (2.1) and (2.5) that

$$q_{0.5}(Y(t)|\mathbf{X}(t), t) = \mathbf{X}^\mathrm{T}(t)\boldsymbol{\beta}(t) \tag{2.8}$$

which allows to estimate $(\beta_0(t), \beta_1(t), \ldots, \beta_p(t))^{\mathrm{T}}$. Subsequently, under Assumption (B2) and since $V(\mathbf{X}(t), t) \geq 0$, it follows that

$$q_{0.5}\left(\left|Y(t) - \mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\beta}(t)\right|\right) = V(\mathbf{X}(t), t), \tag{2.9}$$

which then guides to estimation of the heteroscedasticity function. Finally, expression (2.5) is the key to estimate $a^\tau(t)$ followed by estimation of $q_\tau(Y(t)|\mathbf{X}(t), t)$. See further Sect. 3.2.

## 3 Estimation procedure

The final aim in this paper is to estimate the conditional quantile functions $q_\tau(Y(t)|\mathbf{X}(t), t))$ in (2.5) for various values of $\tau \in (0, 1)$, keeping in mind the desirable property of non-crossing quantile curves in (2.6). There have been a number of papers dealing with developing quantile estimation methods that prevent the estimated quantile curves to cross and violate (2.6). See, among others, the book by Koenker (2005), as well as He (1997), Wu and Liu (2009), Bondell et al. (2010), Liu and Wu (2011), Schnabel and Eilers (2013), Andriyana et al. (2016), and references therein. Andriyana (2015) investigated and compared in detail the performances of various methods. Therefore, in this paper and in this setting of heteroscedastic varying coefficient models, we restrict immediately to adaptation of the methods that appeared 'best' from this extensive study.

We discuss two estimation methods, which are also in line with making different assumptions to ensure identifiability. A basic ingredient throughout all estimation steps is P-splines approximations for univariate functions. Some brief discussion on practical implementation issues for the discussed methods is provided in Sect. 3.3.

### 3.1 Stepwise individual quantile regression estimation

Among the simplest procedures to ensure non-crossingness of estimated quantile regression curves is the stepwise procedure introduced by Wu and Liu (2009). This method starts by estimating a particular quantile curve (for example, the median quantile curve) and then in an upward (respectively, downward) step moves to estimating a higher order (respectively, lower order) quantile curve, including in each of the substeps constraints in the estimation method to prevent the two estimated quantile curves to cross.

Adapting the method of Wu and Liu (2009) to our model setting (2.1) involves constructing the $\tau$-th order conditional quantile of the response $Y(T)$ given $\{\mathbf{X}(T), T = t\}$:

$$q_\tau\left(Y(t)|\mathbf{X}(t), t\right) = \mathbf{X}^{\mathrm{T}}(t)\left[\boldsymbol{\beta}(t) + \boldsymbol{\gamma}(t)a^\tau(t)\right] = \mathbf{X}^{\mathrm{T}}(t)\widetilde{\boldsymbol{\beta}}_\tau(t) \tag{3.1}$$

where $\widetilde{\boldsymbol{\beta}}_\tau(t) = \left(\widetilde{\beta}_{\tau,0}(t), \widetilde{\beta}_{\tau,1}(t), \ldots, \widetilde{\beta}_{\tau,p}(t)\right)^{\mathrm{T}}$ with $\widetilde{\beta}_{\tau,k}(t) = \beta_k(t) + \gamma_k(t)a^\tau(t)$ for $k = 0, \ldots, p$.

The stepwise procedure is designed for estimation of $\widetilde{\boldsymbol{\beta}}_\tau(t) = \left(\widetilde{\beta}_{\tau,0}(t), \widetilde{\beta}_{\tau,1}(t), \ldots, \widetilde{\beta}_{\tau,p}(t)\right)^{\mathrm{T}}$, and from this then subsequently $q_\tau\left(Y(t)|\mathbf{X}(t), t\right)$ from (3.1), and this for

various values of $\tau$, say for $0 < \tau_1 < \cdots < \tau_H < 1$, a set of $H \in I\!N, H \geq 1$ values of $\tau$.

For a given order of the quantile, say $\tau$, the estimation method utilized is that of approximating the unknown univariate functions in $\widetilde{\boldsymbol{\beta}}_\tau(\cdot)$ by a set of basis B-spline functions. More precisely, each of the coefficient functions $\widetilde{\beta}_{\tau,k}(t)$, for $k = 0, \ldots, p$, is approximated by $m_k$ normalized B-splines of degree $v_k$, denoted by $B_{k1}(t; v_k), \ldots, B_{km_k}(t; v_k)$:

$$\widetilde{\beta}_{\tau,k}(t_{ij}) \approx \alpha_{k1} B_{k1}(t_{ij}; v_k) + \cdots + \alpha_{km_k} B_{km_k}(t_{ij}; v_k) = \sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij}; v_k), \quad (3.2)$$

where the dependence on $\tau$ has been suppressed in the right-hand side for notational simplicity.

The P-splines objective function for the individual $\tau$-th order quantile estimation, using an $L_1$-type of penalty function, is then given by

$$\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_\tau \left( Y_{ij} - \sum_{k=0}^{p} \sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij}; v_k) X_{ij}^{(k)} \right) + \sum_{k=0}^{p} \sum_{\ell=d_k+1}^{m_k} \lambda_{\tau,k} \left| \Delta^{d_k} \alpha_{k\ell} \right|,$$
$$(3.3)$$

where $\lambda_{\tau,k} > 0$, for $k = 0, \ldots, p$, are the regularization or smoothing parameters, and where $\Delta^{d_k} \alpha_{k\ell}$ are the differences of order $d_k \in I\!N$, associated with the vector of B-splines coefficients $\boldsymbol{\alpha}_k = (\alpha_{k1}, \ldots, \alpha_{km_k})^{\mathrm{T}}$, i.e. $\Delta^{d_k} \alpha_{k\ell} = \sum_{t=0}^{d_k} (-1)^t \binom{d_k}{t} \alpha_{k(\ell-t)}$. See Eilers and Marx (1996). For simplicity of presentation, we consider the order of differencing not depending on $\tau$, although this is no formal restriction. The global vector containing all B-splines coefficients of all $(p+1)$ unknown univariate functions is given by $\boldsymbol{\alpha}^\tau = (\boldsymbol{\alpha}_0^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_p^{\mathrm{T}})^{\mathrm{T}}$. In the above, the function $\rho_\tau(z) = z I\{z > 0\} + (1 - \tau) I\{z \leq 0\}$ is the so-called check function that is used for quantile regression. Here, $I\{A\}$ denotes the indicator of $A$, i.e. $I\{A\} = 1$ if $A$ holds, and 0 if not. See Koenker (2005) for further background on the check function $\rho_\tau$.

The main steps of the upward and downward stepwise procedure read as follows.

**Step 1: Estimating the median regression curve**
The P-splines median quantile estimator is given by

$$\widehat{\boldsymbol{\alpha}}^{0.5} = \operatorname{argmin}_{\boldsymbol{\alpha}} \left\{ \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5} \left( Y_{ij} - \sum_{k=0}^{p} \sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij}; v_k) X_{ij}^{(k)} \right) \right.$$
$$\left. + \sum_{k=0}^{p} \sum_{\ell=d_k+1}^{m_k} \lambda_{0.5,k} \left| \Delta^{d_k} \alpha_{k\ell} \right| \right\}.$$

**Step 2: Complete up (CU)**
Starting from $\tau_h = 0.5$, the next larger order in the set $\{\tau_1, \ldots, \tau_H\}$ (i.e. $\tau_{h+1} > \tau_h$) is obtained from the following constrained minimization problem:

minimize

$$\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_{h+1}} \left( Y_{ij} - \sum_{k=0}^{p} \sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij}; \nu_k) X_{ij}^{(k)} \right) + \sum_{k=0}^{p} \sum_{\ell=d_k+1}^{m_k} \lambda_{\tau_{h+1},k} \left| \Delta^{d_k} \alpha_{k\ell} \right|$$

(3.4)

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

$$\boldsymbol{\alpha} \geq \widehat{\boldsymbol{\alpha}}^{\tau_h}.$$

(3.5)

The complete upward (CU) stepwise procedure consists of minimizing (3.4) subject to (3.5) for subsequently larger values $\tau_{h+1}$ (for $h$ taking values in a given grid).

**Step 2: Complete down (CD)**
This is similar to the complete upward steps, but now moving downwards from Step 1 (median estimation) on, replacing (3.4) and its constraint (3.5) by

minimize

$$\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{\tau_{h-1}} \left( Y_{ij} - \sum_{k=0}^{p} \sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij}; \nu_k) X_{ij}^{(k)} \right) + \sum_{k=0}^{p} \sum_{\ell=d_k+1}^{m_k} \lambda_{\tau_{h-1},k} \left| \Delta^{d_k} \alpha_{k\ell} \right|$$

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

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

and carrying this out sequentially.

Note that the above stepwise procedure is mainly designed for estimating directly the quantile functions in (2.5) (see also (3.1)). It is not relying on the explicit location-scale model (2.1). It, therefore, does not provide automatically an estimator for the heteroscedasticity function $V(\mathbf{X}(t), t)$ and, hence, does not allow to unravel the impact of covariates on this function. A way to still estimate a heteroscedasticity via this estimation approach is to consider Assumption (A), and estimate the heteroscedasticity function via (2.7) (it is to be noted though that in that case we are relying on the location-scale model again). Unfortunately, Assumption (A) is rather unrealistic and difficult to be justified in practice. In the simulation study in Sect. 4, however, we show the performance of the stepwise method used for the two $\tau$-values, the biggest and smallest quantiles orders considered (i.e $\tau_1$ and $\tau_H$). The advantage of the stepwise estimation procedure is that it does not require any specification of the error term structure (i.e. no location-scale type of modelling is needed).

### 3.2 The AHe approach: an adaptation of He's (1997) approach

We now discuss an estimation approach that is inspired by He (1997) in classical quantile regression. In the current setting and the associated estimation tasks there are three main elements: the coefficient functions for the signal part, $\beta_k(t)$ (for $k = 0, \ldots, p$), the coefficient functions $\gamma_k(t)$ in the heteroscedasticity function (for $k = 0, \ldots, p$), and the $\tau$-th order quantile of the error term, $a^\tau(t)$ for $\tau \in (0, 1)$. In this approach, each of these unknown quantities is estimated separately in three different subsequent

steps. One then finally puts all estimators together to estimate the conditional quantile in (2.5).

The three steps of this procedure read as follows.

*Step 1* Under Assumption (B1), we rely on (2.8) to estimate the $\beta_k(t)$ for $k = 0, \ldots, p$ using P-splines approximation and the 'median' objective function (see (3.2) and (3.3) with $\tau = 0.5$). Denote by $\widehat{\beta}_k(\cdot)$ the obtained estimator of $\beta_k(\cdot)$, and by $\widehat{\boldsymbol{\beta}}(\cdot)$ the resulting estimator for $\boldsymbol{\beta}(\cdot)$.

*Step 2* From (2.9), we know that $V(\mathbf{X}(t), t)$ is the conditional median of $\left|Y(t) - \mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\beta}(t)\right|$. Considering (2.2), we have

$$q_{0.5}\left(\left|Y(t) - \mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\beta}(t)\right|\right) = \gamma_0(t)X^{(0)}(t) + \gamma_1(t)X^{(1)}(t) + \cdots + \gamma_p(t)X^{(p)}(t).$$

Using $\left|Y(t) - \mathbf{X}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}(t)\right|$ as the pseudo response, the coefficient functions $\gamma_k(\cdot)$, for $k = 0, \ldots, p$, can be estimated by means of a P-splines objective function. Firstly, we approximate the coefficients $\gamma_k(t)$ for $k = 0, \ldots, p$ by a B-spline basis of degree $v_k^v$ and dimension $m_k^v$,

$$\gamma_k(t) \approx \alpha_{k1}^v B_{k1}^v(t; v_k^v) + \cdots + \alpha_{km_k^v}^v B_{km_k^v}^v(t; v_k^v) = \sum_{\ell=1}^{m_k^v} \alpha_{k\ell}^v B_{k\ell}^v(t; v_k^v).$$

The P-splines objective function is then given by

$$\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5}\left(\left|Y(t_{ij}) - \mathbf{X}^{\mathrm{T}}(t_{ij})\widehat{\boldsymbol{\beta}}(t_{ij})\right| - \sum_{k=0}^{p}\sum_{\ell=1}^{m_k^v} \alpha_{k\ell}^v B_{k\ell}^v(t_{ij}; v_k^v)X^{(k)}(t_{ij})\right)$$
$$+ \sum_{k=0}^{p}\sum_{\ell=d_k^v}^{m_k^v} \lambda_k^v \left|\Delta^{d_k^v}\alpha_{k\ell}^v\right|,$$

where $\lambda_k^v > 0$ are the regularization parameters. Subsequently, the estimator of $V(\mathbf{X}(t), t)$ is obtained by minimizing the objective function with respect to $\boldsymbol{\alpha}^v = \left((\boldsymbol{\alpha}_0^v)^{\mathrm{T}}, \ldots, (\boldsymbol{\alpha}_p^v)^{\mathrm{T}}\right)^{\mathrm{T}}$, where $\boldsymbol{\alpha}_k^v = \left(\alpha_{k1}^v, \ldots, \alpha_{km_k^v}^v\right)^{\mathrm{T}}$. This leads to $\widehat{\boldsymbol{\alpha}}_k^v$, subsequently $\widehat{\gamma}_k(t) = \sum_{\ell=1}^{m_k^v} \widehat{\alpha}_{k\ell}^v B_{k\ell}^v(t; v_k^v)$, and results in

$$\widehat{V}(\mathbf{X}(t), t) = \widehat{\gamma}_0(t)X^{(0)}(t) + \widehat{\gamma}_1(t)X^{(1)}(t) + \cdots + \widehat{\gamma}_p(t)X^{(p)}(t).$$

*Step 3* In this last step, we exploit the estimation results obtained in the two previous steps. From Model (2.1), we have $Y(t) - \mathbf{X}^{\mathrm{T}}(t)\boldsymbol{\beta}(t) = V(\mathbf{X}(t), t)\varepsilon(t)$. Replacing the coefficient functions $\boldsymbol{\beta}(t)$ and $V(\mathbf{X}(t), t)$ by the estimators obtained in Step 1 and Step 2, respectively, we then approximate the unknown (conditional) quantile $a^{\tau_h}(t)$ of the error term $\varepsilon(t)$ by $m_h^q$ B-spline basis functions of degree $v_h^q$,

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

Hence, the coefficients $\alpha_{h,1}^q, \ldots, \alpha_{h,m_h}^q$ can be estimated by means of a P-splines objective function with pseudo-response $Y(t) - \mathbf{X}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}(t)$ and pseudo-covariate $\widehat{V}(\mathbf{X}(t), t)$, as follows:

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

where $\lambda_h^q > 0$ is the penalization parameter and $d_h^q$ is the order of the differencing operator. Having obtained the estimators for $\alpha_{h,1}^q, \ldots, \alpha_{h,m_h}^q$, the estimate of $a^{\tau_h}(t)$ is

$$\widehat{a}^{\tau_h}(t_{ij}) = \sum_{\ell=1}^{m_h} \widehat{\alpha}_{h,\ell}^q B_\ell^q(t_{ij}; v_h).$$

Putting all estimators together we end up with the following conditional quantile estimators:

$$\widehat{q}_{\tau_h}(Y(t)|\mathbf{X}(t), t) = \mathbf{X}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}(t) + \widehat{V}(\mathbf{X}(t), t)\widehat{a}^{\tau_h}(t). \tag{3.6}$$

To ensure the non-negativity of the estimated heteroscedasticity function in Step 2, we proceed as follows. Before we start, we transform the observed covariate values as to make them non-negative. We do this by subtracting from each $X^{(k)}(t)$, the smallest observed $X^{(k)}(t_{ij})$, i.e. by considering the transformed $\{X^{(k)}(t) - \min_{1 \le i \le n 1 \le j \le N_i} \{X^{(k)}(t_{ij})\}\}$. Since the function values of normalized B-splines are also non-negative, it then suffices to impose that the coefficients are non-negative, i.e.

$$\alpha_{k\ell}^v \ge 0 \quad \forall k = 0, \ldots, p, \quad \ell = 1, \ldots, m_k^v,$$

to ensure that

$$\sum_{k=0}^{p} \sum_{\ell=1}^{m_k^v} \alpha_{k\ell}^v B_{k\ell}^v(t_{ij}; v_k^v) X^{(k)}(t_{ij}) \ge 0.$$

We, thus, add this constraint in the optimization problem.

### 3.3 Practical implementation

As mentioned already in Sect. 3.2, the covariates are transformed such that they are non-negative. This is a way to ensure the non-negativity of the estimated heteroscedasticity function $V(\mathbf{X}(t), t)$, but the transformation also decreases considerably the number of constraints in (3.5) (and alike) in the stepwise procedure of Sect. 3.1.

Approximation by B-splines, with a large number of basis functions, and then introducing a penalty term to prevent for overfitting are employed in all estimation procedures. In practical applications, we fix the degree of the B-splines as well as the degree of the differencing operator [$d_k$ in (3.3)], and use the following data-driven method to choose the regularization parameters $\lambda_{\tau,k}$.

*Step 1* First take $\lambda_{\tau,k} = \lambda$ for all $k = 0, \ldots, p$. From a given grid of $\lambda$-values, choose the $\lambda$ that minimizes the Schwarz Information Criterion

$$\log \left( \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_\tau \left( Y_{ij} - \widehat{q}_\tau (Y_{ij}|\mathbf{X}_{ij}, t_{ij}) \right) \right) + \frac{\log(N)}{2N} p_\lambda,$$

where $p_\lambda$ is the size of the elbow set $\mathcal{E}_\lambda$

$$\mathcal{E}_\lambda = \left\{ (i, j) : Y_{ij} - \widehat{q}_\tau (Y_{ij}|\mathbf{X}_{ij}, t_{ij}) = 0 \right\},$$

i.e. the set of all fits which led to a perfect fitted value $\widehat{q}_\tau (Y_{ij}|\mathbf{X}_{ij}, t_{ij})$ for the observed response value $Y_{ij}$. Denote the resulting choice of $\lambda$ by $\widehat{\lambda}$.

*Step 2* Using $\widehat{\lambda}$, we obtain $\lambda_{\tau,k}$ for all $k = 0, \ldots, p$, from

$$\widehat{\lambda}_{\tau,k} = \widehat{\lambda} \left( \mathcal{R}(\widehat{\beta}_{\tau,k}^B(\cdot)) \right)^{-\kappa}$$

where $\widehat{\beta}_{\tau,k}^B(t)$ is the quantile regression estimator of $\beta_{\tau,k}(t)$ using B-splines (so with the objective function putting $\lambda_{\tau,k} = 0$), $\mathcal{R}(\widehat{\beta}_{\tau,k}^B(\cdot))$ is the range of all values $\widehat{\beta}_{\tau,k}^B((t_{ij}))$, and $\kappa > 0$ is a given number (in Sects. 4 and 5 we took $\kappa = 0.5$).

Each optimization problem discussed in this paper can be translated into a linear programming problem, that is then translated into a dual problem, which is solved using a Frisch-Newton interior-point algorithm. For details, we refer to Andriyana (2015). The necessary computer codes have been developed and are collected in a freely available R Package QRegVCM that has been developed by the first author (Y. Andriyana).

## 4 Simulation study

In this section, we investigate the performances of the stepwise procedure (Sect. 3.1) and the AHe approach (Sect. 3.2) on four simulation models, involving three covariates.

Measurements can happen only at the time points $\{0, 1, 2, \ldots, 49\}$. From this fixed set of possible time points, each time point (except for the starting time point 0)

**Table 1** Coefficient functions in Models 1–4

| Coefficient functions | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| $\beta_0(t)$ | $2\sqrt{t}$ | $\frac{2\sqrt{t}}{3}$ | $0$ | $0$ |
| $\beta_1(t)$ | $\frac{(\pi t + 10)}{15}$ | $\frac{\cos(t-5)\pi}{20}$ | $\cos\left(\frac{(t-25)\pi}{1000}\right)$ | $\cos\left(\frac{(t-25)\pi}{100}\right)$ |
| $\beta_2(t)$ | $\sin\left(\frac{\pi t}{30}\right) + 3$ | $\frac{\sin(\pi t)}{20} + 5$ | $\sin\left(\frac{\pi t}{30}\right) + 3$ | $\sin\left(\frac{\pi t}{30}\right) + 13$ |
| $\beta_3(t)$ | $\frac{(20-t)^2}{1000} - 4$ | $\frac{|25-t|^{5/2}}{200}$ | $-4 + \frac{(20-t)^2}{1000}$ | $-4 + \frac{(20-t)^2}{100}$ |
| $\gamma_0(t)$ | $\frac{\beta_0(t)}{8}$ | $\sqrt{\beta_0(t)}$ | $0.2\sqrt{t}$ | $\frac{(25-t)^2 + 3t}{100}$ |
| $\gamma_1(t)$ | $\frac{\sqrt{|\beta_1(t)|}}{2}$ | $\frac{\beta_1^2(t)}{200}$ | $0$ | $0$ |
| $\gamma_2(t)$ | $\frac{|\beta_2(t)|}{50}$ | $\frac{\sqrt{\beta_2(t)}}{100}$ | $0$ | $0$ |
| $\gamma_3(t)$ | $\frac{|\beta_3(t)|}{50}$ | $\frac{\beta_3(t)}{200}$ | $0$ | $0$ |

has a probability of 40% to be skipped. This creates different numbers of repeated measurements $N_i$ for each subject $i = 1, \ldots, n$. The actual measurement times are generated by adding a $U[0, 0.5]$ random variable to the non-skipped scheduled times. The error term $\varepsilon(t)$ is generated from a transformed multivariate normal distribution. The transformation is needed to fulfill certain model assumptions. Firstly, we consider $\eta(t_{ij})$, where $\eta(\cdot)$ follows a multivariate normal distribution with covariance structure $\text{Cov}(\eta(t_{ij}), \eta(t_{ij'})) = 30 \exp(-|j - j'|)$. Secondly, the generated $\eta(t_{ij})$ are transformed such that the error term has a desired median (in this case $q_{0.5}(\varepsilon(t)) = 0$ and $q_{0.5}(|\varepsilon(t)|) = 1$) by the following transformation:

$$\varepsilon(t) = \frac{\eta(t) - q_{0.5}(\eta(t))}{q_{0.5}\left(|\eta(t) - q_{0.5}(\eta(t))|\right)}.$$

The varying coefficients, $\beta_k(\cdot)$ and $\gamma_k(\cdot)$ for $k = 0, 1, 2, 3$ in, respectively, the main conditional quantile function and the heteroscedasticity function, for the four models are presented in Table 1.

All elements of a model are such that the signal-to-noise ratio

$$\frac{\text{Sample variance of all } \left\{\beta_0(t_{ij}) + \sum_{k=1}^3 \beta_k(t_{ij}) X^{(k)}(t_{ij})\right\}}{\text{Sample variance of all } \left\{V(\mathbf{X}(t_{ij}), t_{ij}) \varepsilon(t_{ij})\right\}}$$

is approximately 7.

Throughout Sects. 4 and 5, we utilize B-splines of degree 3 with 10 equidistant knot points on the time interval (leading to $13 - 1 = 12$ basis functions) and a penalty in which we take differencing order 1. All covariates are transformed to be in $[0, 1]$, which puts them on a comparable scale. More precisely, we work with $\left\{X^{(k)}(t) - \min_{1 \le i \le n 1 \le j \le N_i} \left\{X^{(k)}(t_{ij})\right\}\right\} / \left\{\max_{1 \le i \le n 1 \le j \le N_i} \left\{X^{(k)}(t_{ij})\right\} - \min_{1 \le i \le n 1 \le j \le N_i} \left\{X^{(k)}(t_{ij})\right\}\right\}$.
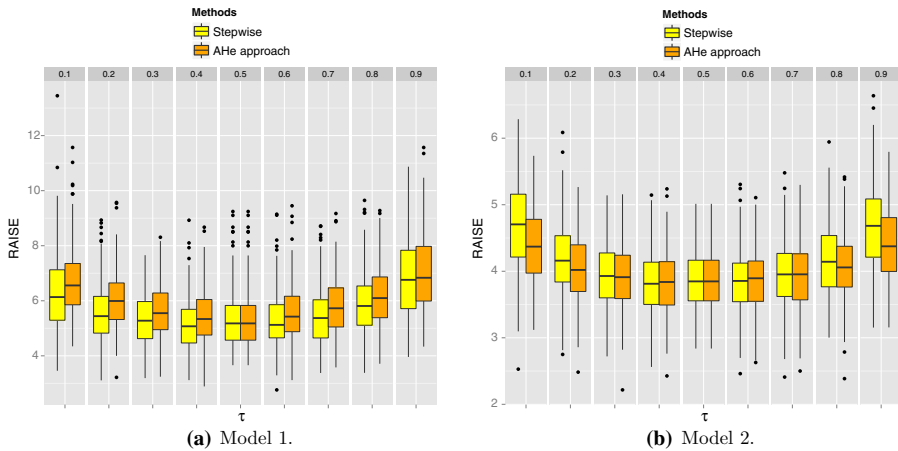
**Fig. 3** *Boxplots* of RAISE($\widehat{q}_{\tau_h}(Y(t)|\mathbf{X}(t), t)$) using the stepwise procedure and the AHe approach

For each model, we draw 200 samples of size $n = 100$. For a given simulated sample $s$ ($s = 1, \ldots, 200$), with the dependence on the sample indicated through the superscript "$(s)$", the performance of a quantile estimation method is evaluated via the empirical root approximate integrated squared error, defined as:

$$\text{RAISE}(\widehat{q}_\tau^{(s)}(\cdot)) = \left( \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left( \widehat{q}_\tau^{(s)}(Y_{ij}|\mathbf{X}_{ij}, t_{ij}) - q_\tau(Y_{ij}|\mathbf{X}_{ij}, t_{ij}) \right)^2 \right)^{1/2}. \quad (4.1)$$

We do not write the dependence on the sample $s$ in the second term within brackets for not making the notation too heavy, but it is good to realize that due to the conditioning on the covariates and time, this term also changes with the sample.

Note that Models 3 and 4 are models for which $V(\mathbf{X}(t), t) = \gamma_0(t)$ and, hence, the error term does not vary with the covariates. For these models, the methods developed in Andriyana et al. (2016) would suffice. These two models are included in the simulation study to see the possible loss using, in this simpler setting, the too sophisticated estimation method of, for example, Sect. 3.2. In Sect. 4.1, the simulation results for Models 1 and 2 are summarized. These, for Models 3 and 4, can be found in Sect. 4.2.

## 4.1 Simulation results for Models 1 and 2

Figure 3 shows boxplots of RAISE($\widehat{q}_{\tau_h}^{(s)}(\cdot)$) over 200 simulations for the two investigated methods. Overall, both methods have a comparable performance, with the stepwise procedure performing slightly better than the AHe approach for Model 1.

A representative plot of the estimated conditional quantile curves is provided next. For each method, we focus on a sample for which median performance of the estimators across the 200 simulations was obtained, i.e. a sample for which the evaluation criterion in (4.1) delivered the 50th-percentile of $\overline{\text{RAISE}}_H\left(\widehat{q}^{(s)}(\cdot)\right)$, where
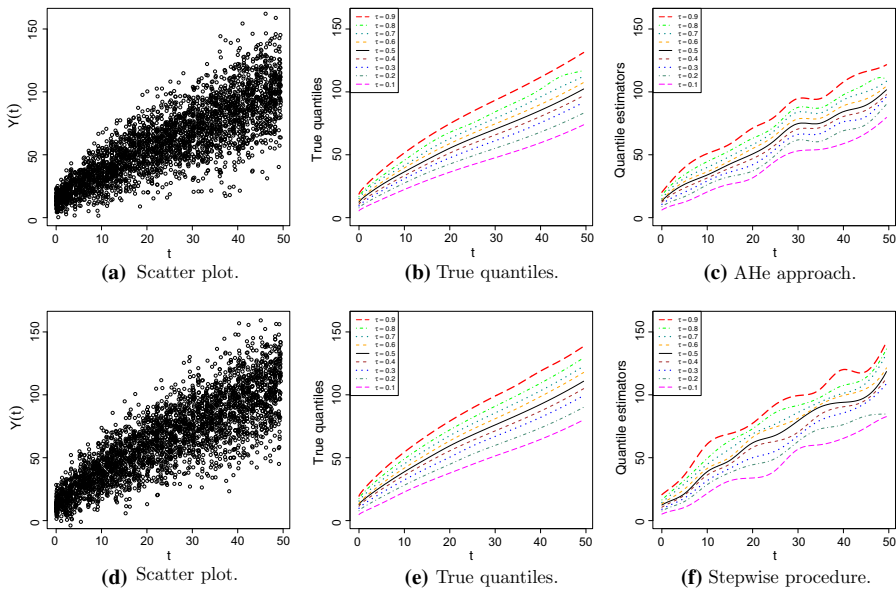
**Fig. 4** Model 1. **a**, **d** Scatter plots of the sample giving median performance of $\widehat{q}_\tau(.)$; **b**, **e** true conditional quantiles at the maximum values of the covariates; **c**, **f** $\widehat{q}_\tau(Y(t)|\mathbf{X}(t), t)$. Results for the AHe approach (*top panels*) and the stepwise procedure (*bottom panels*)

$$\overline{\mathrm{RAISE}}_H\left(\widehat{q}^{(s)}(.)\right) = \frac{1}{H}\sum_{h=1}^{H}\mathrm{RAISE}\left(\widehat{q}_{\tau_h}^{(s)}(.)\right).$$

Consequently, we present estimated curves with median performance in each method.

The resulting representative estimated quantile curves for Model 1 are depicted in Fig. 4. For graphical presentation purpose, we focus on the maximum values of the covariates at each time point. To give a visual impression of the quality of the conditional quantile estimation, we provide the associated scatterplot, the true quantiles curves, and the estimated quantile curves, for the chosen representative sample (which might be different for each method). In Fig. 5, we present the results for Model 2, but only for the method following the AHe Approach. Results for the stepwise procedure are similar, but slightly more variable (as already indicated by the boxplots in Fig. 3).

Next, we investigate the finite-sample performances of the estimators of $\widehat{V}(\mathbf{X}(t), t)$ via the criterion (suppressing from now on, for notational simplicity, the dependence "$(s)$" on the sample)

$$\mathrm{RAISE}\left(\widehat{V}(\cdot, \cdot)\right) = \left(\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\widehat{V}(\mathbf{X}_{ij}, t_{ij}) - V(\mathbf{X}_{ij}, t_{ij})\right)^2\right)^{1/2}. \qquad (4.2)$$

The boxplots of $\mathrm{RAISE}\left(\widehat{V}(.)\right)$ over the 200 simulations are given in Fig. 6. Note that the stepwise procedure is less variable than the AHe approach, but the latter has a better performance in case of Model 2.
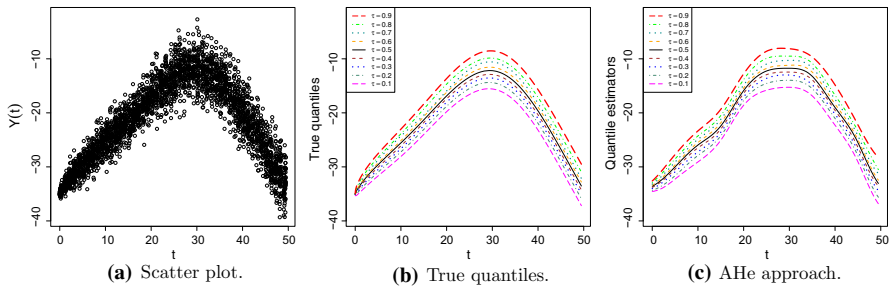
**(a)** Scatter plot.          **(b)** True quantiles.          **(c)** AHe approach.

**Fig. 5** Model 2. **a** Scatter plot of the sample giving median performance of $\widehat{q}_\tau(.)$; **b** true conditional quantiles at the maximum values of the covariates; **c** $\widehat{q}_\tau(Y(t)|\mathbf{X}(t), t)$. Results for the AHe approach
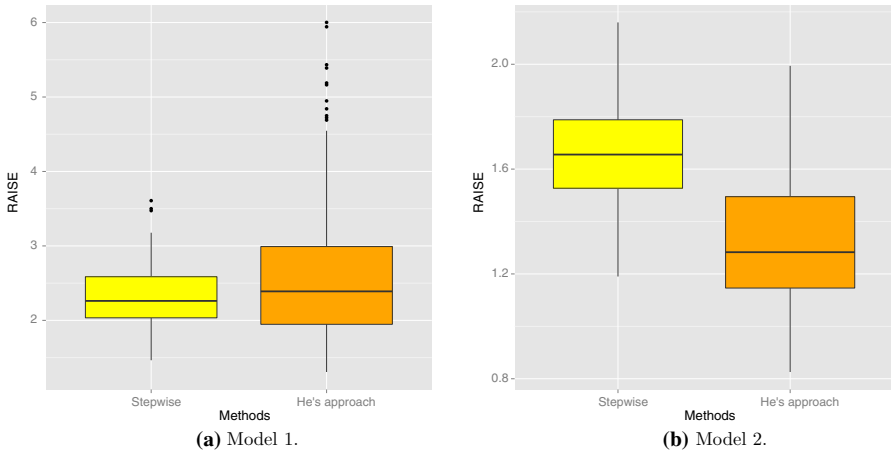


**(a)** Model 1.                                           **(b)** Model 2.

**Fig. 6** *Boxplots* of RAISE($\widehat{V}(\cdot, \cdot)$) using the stepwise procedure and the AHe approach (respectively, the *first* and the *second boxplots* in **a**, **b**)

## 4.2 Simulation results for Models 3 and 4

Since for Models 3 and 4 (see Table 1) $V(\mathbf{X}(t), t) = \gamma_0(t)$, an error structure as in (2.1) is not needed, and the methods developed by Andriyana et al. (2016) would be sufficient to estimate $V(t) = \gamma_0(t)$. Among these methods is an adaptation of the approach of He (1997) for that simpler setting, and we refer to it as the AHe $V(t)$ approach in the text below.

Figure 7b, c depicts boxplots of the RAISE($\widehat{q}_{\tau_h}(Y(t)|\mathbf{X}(t), t)$) values for Models 3 and 4 using the method of Sect. 3.2 and the AHe $V(t)$ approach. For comparison purpose, we also include the boxplot of the results for both methods for Model 1 (see Fig. 7a). As can be seen from the boxplots in Fig. 7b, c, both methods perform comparable, with only a small price to pay in terms of a slightly increased variability for the method of Sect. 3.2. This more general method is, thus, also able to adapt to the simpler situations. When using the simpler AHe $V(t)$ approach to Model 1, it is clear that it performs badly [except for the median quantile curve, when both methods
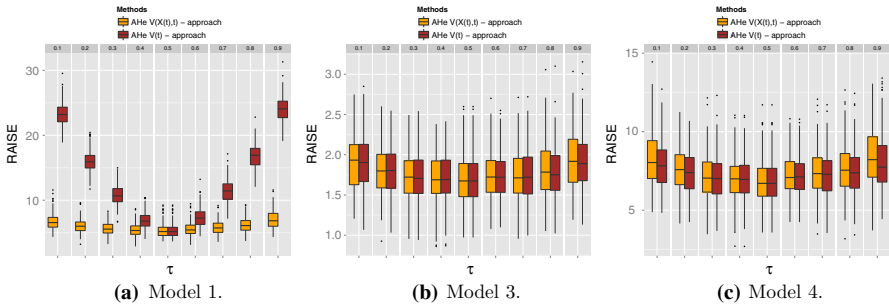
**Fig. 7** *Boxplots* of RAISE($\widehat{q}_{\tau_h}(Y(t)|\mathbf{X}(t), t)$) for the AHe $V(\mathbf{X}(t), t)$ and the AHe $V(t)$ approaches (respectively, *first* and *second boxplots*), for Models 1, 3 and 4
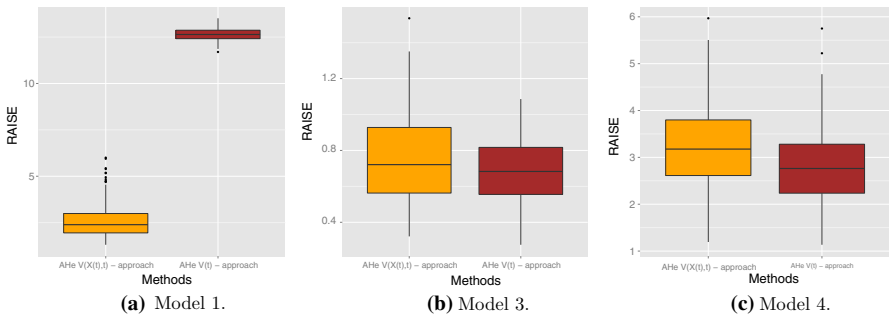


**Fig. 8** *Boxplots* of RAISE($\widehat{V}(\mathbf{X}(t), t)$) for the AHe $V(\mathbf{X}(t), t)$ and the AHe $V(t)$ approaches (respectively, *first* and *second boxplots*), for Models 1, 3 and 4

coincide by Assumption (B1)]. This just illustrates the need for the methods developed in this paper.

The boxplots in Fig. 8b, c summarize the quality of the estimators for the function $V(\mathbf{X}(t), t)$ for Models 3 and 4. A similar conclusion can be drawn as for the estimation of the quantile curves. Both methods, this of Sect. 3.2 and the simpler AHe $V(t)$ approach perform comparable, with a slightly higher variability for the former one. In contrast, when applying the simpler method to Model 1 it fails completely, as seen from the boxplots in Fig. 8a.

Picking up a sample corresponding to median performance of $\widehat{V}(t)$, we plot the estimators $\widehat{V}(\mathbf{X}(t), t)$ and $\widehat{V}(t)$ together with the corresponding standardized residuals in Fig. 9 for Model 3. In particular, the standardized residuals based on $\widehat{V}(\mathbf{X}(t), t)$, i.e. $\left(Y_{ij} - \widehat{q}_{0.5}(X_{ij}, t_{ij})\right)/\widehat{V}(\mathbf{X}(t_{ij}), t_{ij})$, show that the method performs well also in this simpler setting [note the similarity between pictures (e) and (f)]. Figure 10 shows the true $V(t)$ and the estimators $\widehat{V}(\mathbf{X}(t), t)$ and $\widehat{V}(t)$, for a median performing sample, for Model 4. The plots for the standardized residuals are not provided, since they look similar to these for Model 3.

The conditional quantile curves for Model 3, based on a sample exhibiting median quantile performance for the simple AHe $V(t)$ approach, are presented in Fig. 11. This figure confirms the previous conclusions.
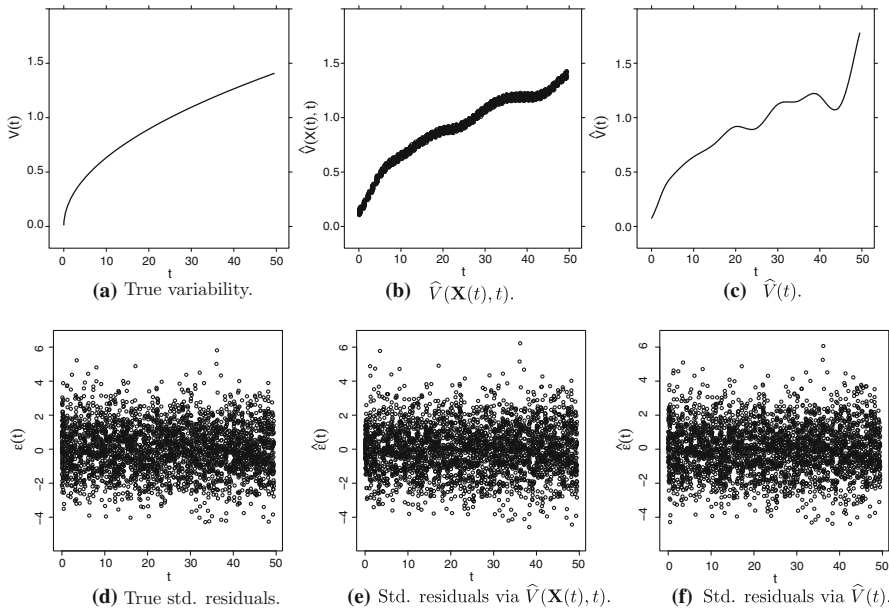
**Fig. 9** Model 3: median performance of $\widehat{V}(t)$; **a** true variability, **b** $\widehat{V}(\mathbf{X}(t), t)$, (c) $\widehat{V}(t)$, **d** The true standardized residuals, **e** standardized residuals via the $\widehat{V}(\mathbf{X}(t), t)$ approach and **f** Standardized residuals via the $\widehat{V}(t)$ approach
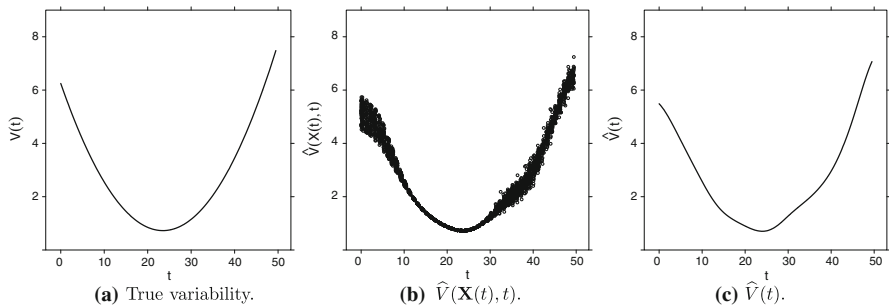


**Fig. 10** Model 4: median performance of $\widehat{V}(t)$; **a** true variability, **b** $\widehat{V}(\mathbf{X}(t), t)$, **c** $\widehat{V}(t)$, **d** the true standardized residuals, **e** standardized residuals via the $\widehat{V}(\mathbf{X}(t), t)$ approach and **f** standardized residuals via the $\widehat{V}(t)$ approach

## 5 Real-data applications

In this section, we illustrate the practical use of the methods of Sect. 3 on a few real-data examples.

We present the estimators of the conditional quantile curves using the stepwise procedure of Sect. 3.1, the AHe $V(\mathbf{X}(t), t)$ approach of Sect. 3.2 and the simple AHe $V(t)$ approach of Andriyana et al. (2016). Differences in results for the first two methods on the one hand and the last method on the other hand will be a possible
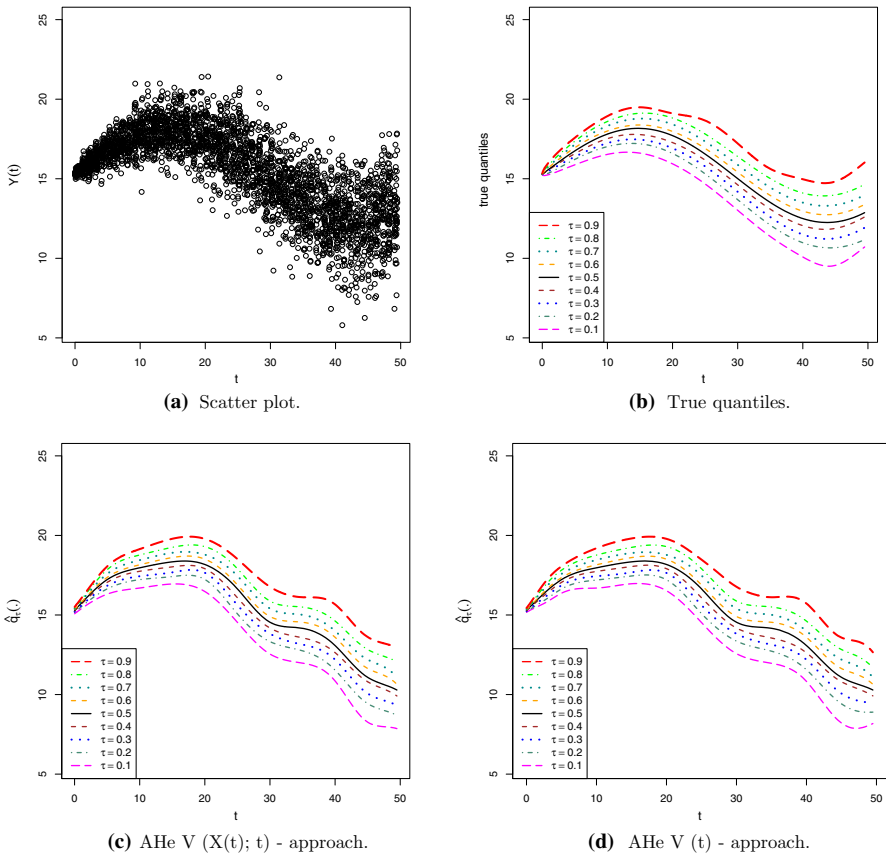
**(a)** Scatter plot.

**(b)** True quantiles.

**(c)** AHe V (X(t); t) - approach.

**(d)** AHe V (t) - approach.

**Fig. 11** Model 3: **a** scatter plot of the sample giving median quantile performance using AHe $V(t)$ approach; **b** true conditional quantiles applied to the maximum values of the covariates; **c** $\widehat{q}_\tau(Y(t)|\mathbf{X}(t), t)$ using AHe $V(X(t), t)$ approach; and **d** $\widehat{q}_\tau(Y(t)|\mathbf{X}(t), t)$ using AHe $V(t)$ approach

indication for which model might be more appropriate. The development of formal statistical goodness-of-fit tests is part of future research. See also Sect. 6.

To analyse all real-data examples, we use B-splines of degree 3 with 10 equidistant knot points on the time interval (leading to $13 - 1 = 12$ basis functions) with differencing order 1. All covariates are transformed to be in $[0, 1]$, which puts them on a comparable scale.

## 5.1 Air Pollution data example (PM10)

This data set originates back to a study regarding air pollution at a road, where traffic volume and meteorological variables are measured. The data considered here are a subsample of 500 observations. 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
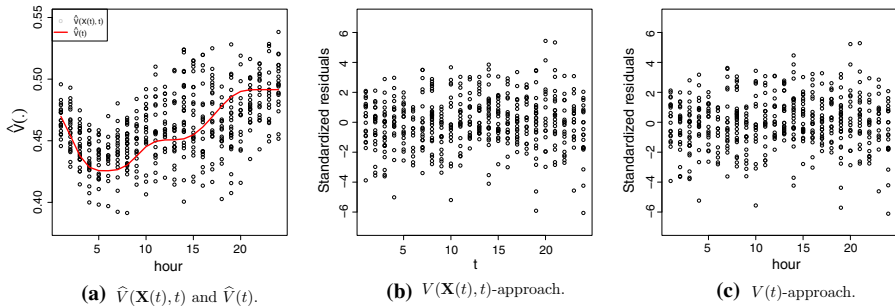
**Fig. 12** Air pollution data. **a** The estimated variability curves $\widehat{V}(\mathbf{X}(t), t)$ and $\widehat{V}(t)$; **b** standardized residuals using the AHe $V(\mathbf{X}(t), t)$ approach and **c** Standardized residuals for the AHe $V(t)$ approach

time points (hours). There are between 1 and 6 measurements per day with 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. Guo et al. (2012) analysed these data using a varying coefficient model including two covariates, but in an i.i.d. setting. These data were also analysed in Andriyana et al. (2016), stepping away from the i.i.d. setting but assuming that the error is not depending on the covariates. In the analysis here, we drop this assumption, and also include an additional covariate in the analysis.

The response variable $Y(t)$ consists of hourly values of the logarithm of the concentration of PM10. This substance is a mixture of solid and liquid droplets with diameter less than 10 μm, and is one of the air pollutants suspected to have a negative effect on human health. See for example Aldrin and Hobaek Haff (2005) and Oftedal et al. (2009) for studies on air pollution, involving particle pollutants such as PM10. The covariates considered in the analysis here are: $X^{(1)}(t)$ is the logarithm of the number of cars per hour; $X^{(2)}(t)$ is the wind speed (in meters/second), and $X^{(3)}(t)$ is the temperature (in degree Celcius, measured two meters above the ground).

Figure 12a depicts $\widehat{V}(\mathbf{X}(t), t)$ as a function of $t$ (the points) as well as $\widehat{V}(t)$ (the solid line). The standardized residuals $\left(Y_{ij} - \widehat{q}_{0.5}(X_{ij}, t_{ij})\right) / \widehat{V}(\mathbf{X}(t_{ij}), t_{ij})$ and $\left(Y_{ij} - \widehat{q}_{0.5}(X_{ij}, t_{ij})\right) / \widehat{V}(t_{ij})$ are plotted in Fig. 12b, c, and little conclusions can be drawn from these.

The influences of each of the covariates on, respectively, the main conditional quantile function and the heteroscedasticity function are obtained via, respectively, the estimated coefficients $\widehat{\beta}_k(\cdot)$ and $\widehat{\gamma}_k(\cdot)$, using the method of Sect. 3.2. These estimated coefficients are depicted in Figs. 13 and 14. Note that for example the covariate *temperature* has the smallest effect on the main conditional quantile, whereas this covariate seems to play a role (although possibly small) in the heteroscedasticity present in the data. The covariates number of *cars* and *wind* have most impact on the main conditional quantile function, with their effect being maximal and nearly constant for a large portion of the day (between approximately 8 am and 7 pm for the first covariate).

The estimated conditional quantile curves applied to the mean and the maximum values of the covariates are presented in Fig. 15, in the top and bottom panels, respectively. The estimated conditional quantile curves show less curvature for mean values of the covariates than for maximum values.
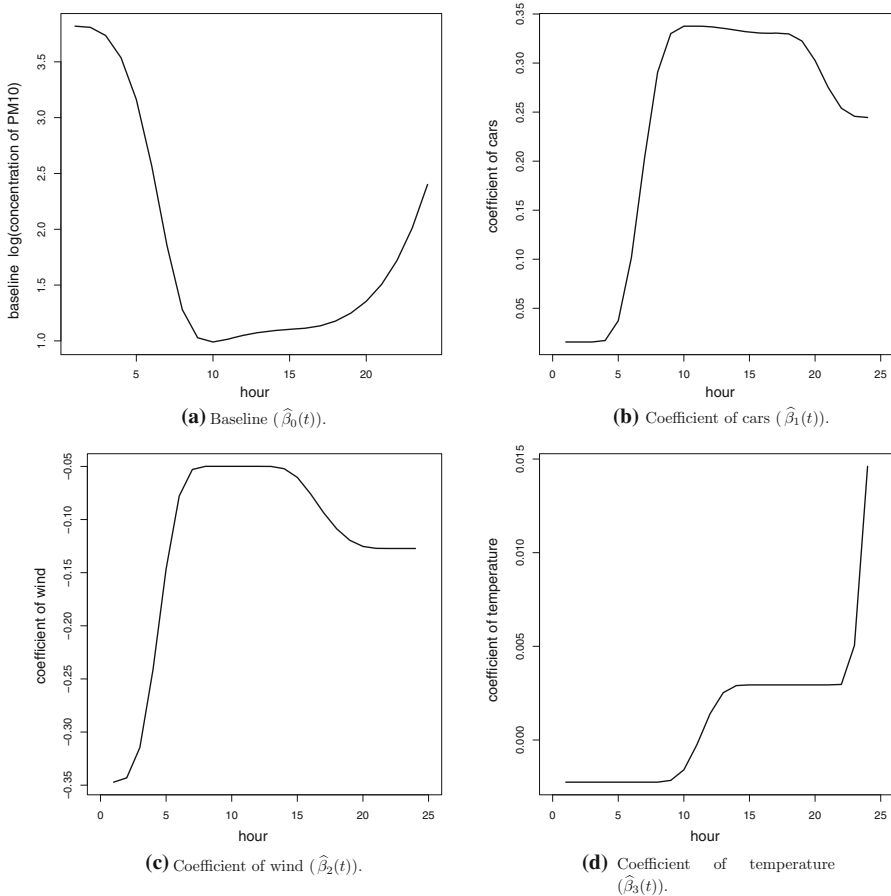
(a) Baseline ($\widehat{\beta}_0(t)$).

(b) Coefficient of cars ($\widehat{\beta}_1(t)$).

(c) Coefficient of wind ($\widehat{\beta}_2(t)$).

(d) Coefficient of temperature ($\widehat{\beta}_3(t)$).

**Fig. 13** Air pollution data. The estimated coefficient functions $\widehat{\beta}_k(t)$, $k = 0, 1, 2, 3$, using the method of Sect. 3.2. **a** Baseline ($\widehat{\beta}_0(t)$). **b** Coefficient of cars ($\widehat{\beta}_1(t)$). **c** Coefficient of wind ($\widehat{\beta}_2(t)$). **d** Coefficient of temperature ($\widehat{\beta}_3(t)$)

## 5.2 UK employment data example

The next data frame contains company accounts from Datastream International, which provide accounts records of employment and remuneration (i.e. wage bill) for all UK quoted companies. The data are available in the `plm` R-package under `EmplUK`. Table 2 briefly describes the variables involved. More detailed information regarding these data can be found in the data Appendix of Arellano and Bond (1991). This study deals with determinants of *employment* in 140 UK firms observed each year from 1976–1984. The data are unbalanced both in the sense that some firms have more observations than others, and also in the sense that these observations correspond to different points in historical time. The range of observations per firm is seven to nine [i.e. $\min(N_i) = 7$ and $\max(N_i) = 9$]. The response variable ($Y(t_{ij})$) is the logarithm of the UK *employment* in company $i$ at time $t_{ij}$, where $i = 1, \ldots, 140$ and $j = 1, \ldots, N_i$.
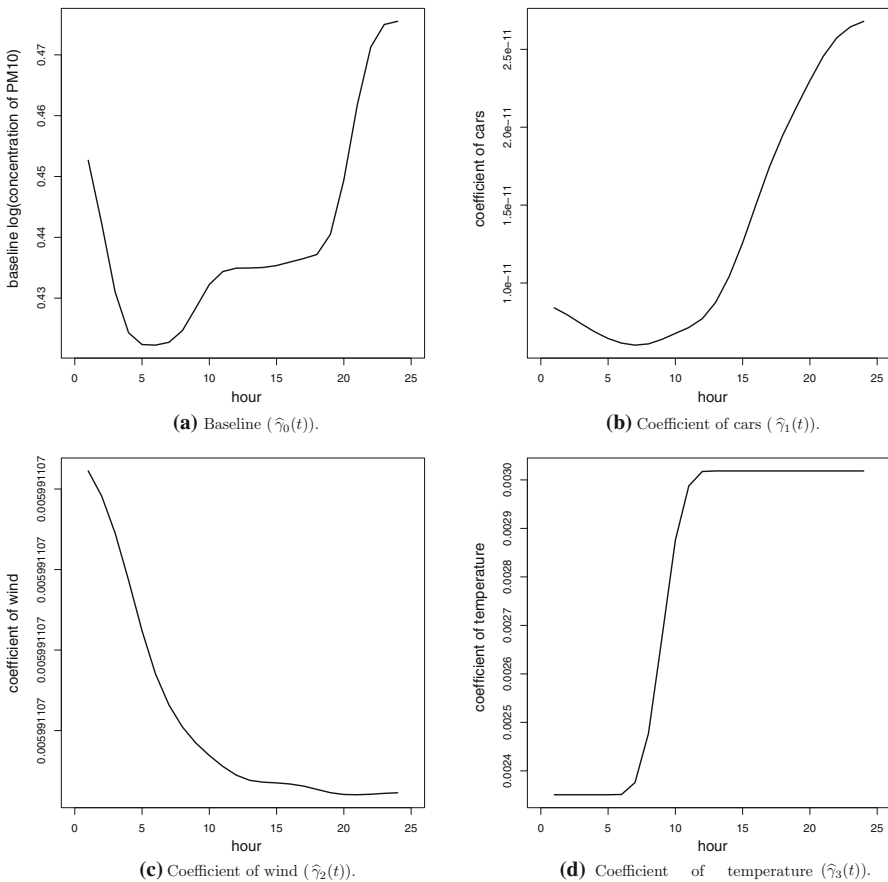
**Fig. 14** Air pollution data. The estimated coefficient functions $\widehat{\gamma}_k(t)$, $k = 0, 1, 2, 3$, using the method of Sect. 3.2. **a** Baseline ($\widehat{\gamma}_0(t)$). **b** Coefficient of cars ($\widehat{\gamma}_1(t)$). **c** Coefficient of wind ($\widehat{\gamma}_2(t)$). **d** Coefficient of temperature ($\widehat{\gamma}_3(t)$)

The three covariates are (see Table 2): the average annual *wage* per employee in the company ($X^{(1)}(t_{ij})$); the *capital* defined as the book value of gross fixed assets ($X^{(2)}(t_{ij})$); and an index of value-added *output* at constant cost ($X^{(3)}(t_{ij})$). Arellano and Bond (1991) and also Kleiber and Zeileis (2008) combine a static model equation including all three covariates using a dynamic model with 2 lagged endogenous terms.

Here, we present a more flexible modelling using varying coefficient models. Firstly, the conditional quantile estimators are plotted in Fig. 1, in the top panels when applied to the mean of the covariate values, and in the bottom panels when applied to the maximum of the covariate values. As can be seen, the estimated conditional quantiles for given mean values of the covariates are almost linear, with a change in direction from the year 1982 onwards. Note that, considering linearity in quantile regression for all quantiles is quite restrictive; since for some quantiles a linearity assumption might be appropriate, while for some other order quantile curves this might not be the case.
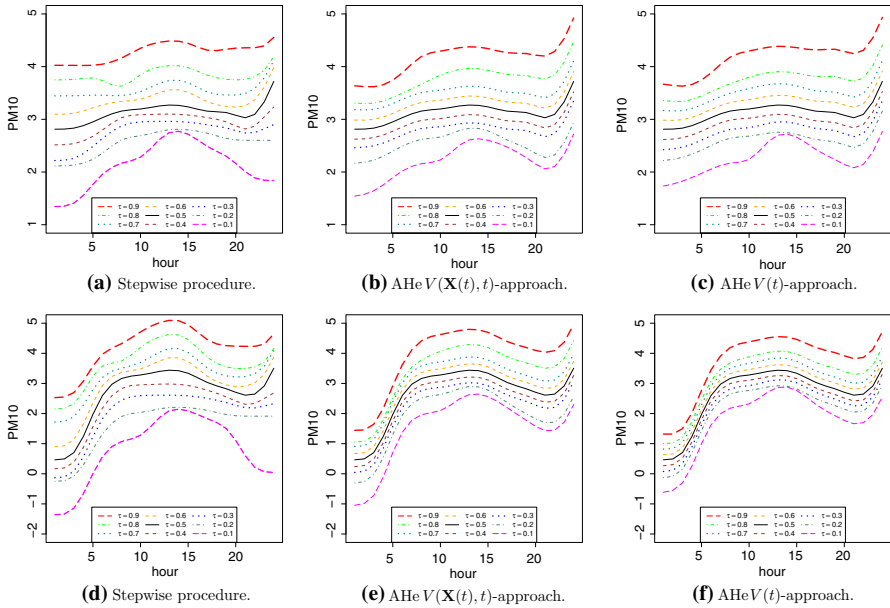
**Fig. 15** Air pollution data. Estimated log(concentration of PM10) quantile curves for $\tau = 0.1, 0.2, \ldots, 0.9$ at the mean values (*top panels*) and at the maximal values (*bottom panels*) of all covariates, using the methods: **a** stepwise procedure, **b** AHe $V(\mathbf{X}(t), t)$ approach, and **c** AHe $V(t)$ approach

**Table 2** The UK employment data

| Variable | Description |
|---|---|
| Employment | The number of employees in 140 companies with operations mainly in UK, whose main activity is manufacturing |
| Real wage | A measure of real average annual remuneration per employee in a company |
| Gross capital stock | The historic cost book value of gross fixed assets |
| Industry output | An index of value-added output at constant factor cost |

See more detailed information about the data in the Appendix of Arellano and Bond (1991)

This is for example well visible for the estimated conditional quantile curves in Fig. 1, using the approach of Sect. 3.1. Most remarkable is the difference in the estimated conditional quantiles in Fig. 1d–f, revealing that possible heteroscedasticity is present in the data.

We next wonder about this heteroscedasticity present in the data. The estimated heteroscedasticity function $\widehat{V}(\mathbf{X}(t), t)$, using the method of Sect. 3.2, is plotted in Fig. 16a, together with the estimator $\widehat{V}(t)$. The associated standardized residuals are plotted in Fig. 16b, c. To be remarked is the skewness in the plot in Fig. 16a for a fixed year, with a high concentration towards lower values of $\widehat{V}(\mathbf{X}(t), t)$, but also a considerable amount of high values for the same time moment.

Finally, Figs. 17 and 18 depict the estimated main quantile coefficient functions and the estimated coefficients in the heteroscedasticity function. See also Fig. 2 in Sect. 1.
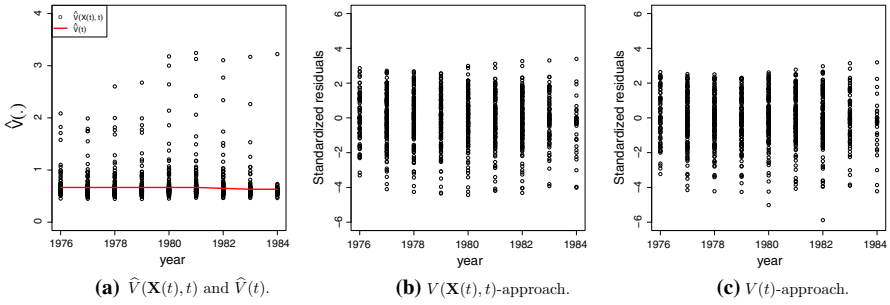
**(a)** $\widehat{V}(\mathbf{X}(t), t)$ and $\widehat{V}(t)$.     **(b)** $V(\mathbf{X}(t), t)$-approach.     **(c)** $V(t)$-approach.

**Fig. 16** UK employment data. **a** Variability estimators and **b**, **c** standardized residuals



**(a)** Baseline   log-employment $(\widehat{\beta}_0(t))$.     **(b)** Coefficient of *wage* $(\widehat{\beta}_1(t))$.     **(c)** Coefficient of *output* $(\widehat{\beta}_3(t))$.
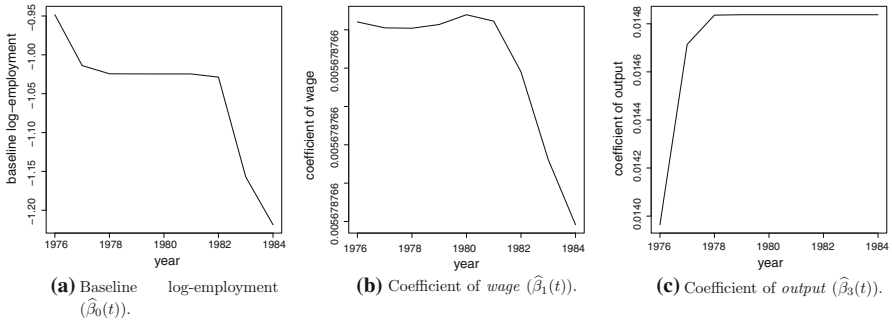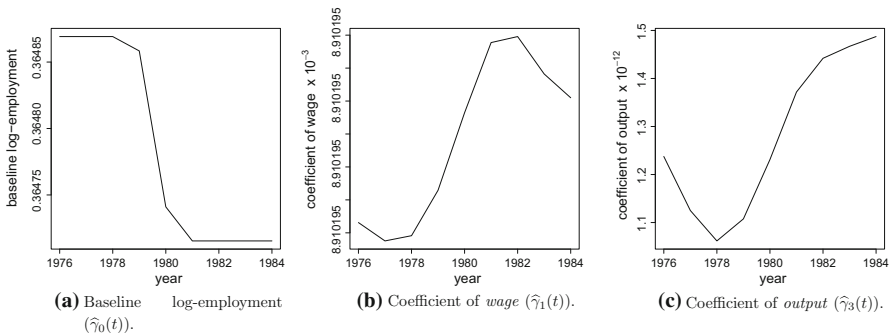
**Fig. 17** UK employment data. The estimated coefficient functions $\widehat{\beta}_k(t)$, $k = 0, 1, 3$ using the method of Sect. 3.2. **a** Baseline log-employment $(\widehat{\beta}_0(t))$. **b** Coefficient of *wage* $(\widehat{\beta}_1(t))$. **c** Coefficient of *output* $(\widehat{\beta}_3(t))$



**(a)** Baseline   log-employment $(\widehat{\gamma}_0(t))$.     **(b)** Coefficient of *wage* $(\widehat{\gamma}_1(t))$.     **(c)** Coefficient of *output* $(\widehat{\gamma}_3(t))$.

**Fig. 18** UK employment data. The estimated coefficient functions $\widehat{\gamma}_k(t)$, $k = 0, 1, 3$ using the method of Sect. 3.2. **a** Baseline log-employment $(\widehat{\gamma}_0(t))$. **b** Coefficient of *wage* $(\widehat{\gamma}_1(t))$. **c** Coefficient of *output* $(\widehat{\gamma}_3(t))$

The covariate *capital* seems to contribute most, both in the main conditional quantile function and in the heteroscedasticity function.

## 6 Further discussion and conclusion

In Andriyana et al. (2016), the heteroscedasticity function was only allowed to change with the time (i.e. $V(t)$) and as was illustrated with the example in the introduction and Sect. 5 this may be too simple in some applications. In this paper, we there-

fore consider a varying coefficient model allowing for a flexible heteroscedastic error structure that may also vary with the covariates, i.e. $V(\mathbf{X}(t), t)$. The interest was in estimating conditional quantile curves, together with the more complex heteroscedasticity function. All univariate coefficient functions appearing in both target quantities are approximated using P-splines. Thanks to this flexible modelling, combined with an appropriate approach (such as the AHe and the stepwise method), these complex estimation tasks become feasible. In the simulation study, it was demonstrated that if one unnecessarily assumes the more complex heteroscedastic error structure, the loss in efficiency, compared to working with the sufficient simpler heteroscedastic error structure, is small.

An interesting question, however, is how to test for an appropriate heteroscedastic structure. Note that the choice between a simple heteroscedastic structure $V(t)$ or a more flexible heteroscedastic structure $V(\mathbf{X}(t), t)$ can, in the setting of this paper, be translated in the univariate coefficient functions $\gamma_k(\cdot)$, for $k = 1, \ldots, p$, to be all equal to zero versus at least one coefficient function is not zero. In Gijbels et al. (2016), testing procedures for testing for various (nested) heteroscedastic structures are developed.

Note that in this paper we assume that $V(\mathbf{X}(T), T) \geq 0$. To ensure positiveness of the estimated heteroscedasticity function, one can proceed as indicated at the end of Sect. 3.2. An alternative would be to model the logarithm (or any other appropriate link function) of the heteroscedasticity function as in (2.2). Both approaches are feasible with rather similar performances, and we do not elaborate further on this.

Varying coefficient models are among flexible regression models, together with, for example, generalized additive models. As was remarked by a referee, one could think of some alternative estimation approaches in varying coefficient models. In case there is only one covariate one could indeed view a varying coefficient model (the most simplistic variant of it) as an additive model. See for example Stasinapoulos and Rigby (2007), and the GAMLSS package in R. An additional limitation, however, is that one needs to specify a specific error distribution.

Yet another viewpoint and alternative estimation approach could be envisaged as follows. Approximating all the univariate covariate functions by a combination of B-spline basis functions, there is a way to write the systematic term in the location-scale model as a linear term involving a design matrix that consists of rows of all B-spline functions evaluated in all observational points. See for example Andriyana et al. (2014). This then opens the way to use Bayesian regression quantile methods (for example using the R package bayesQR). Although this is an alternative approach, the disadvantage is that it does not allow to estimate the heteroscedasticity function. Given the above limitations we did not further elaborate on these alternative approaches.

# References

Aldrin, M., Hobaek Haff, I.: Generalised additive modelling of air pollution, traffic volume and meteorology. Atmos. Environ. **39**, 2145–2155 (2005)

Andriyana, Y.: P-splines quantile regression in varying coefficient models. Doctoral Dissertation, KU Leuven, Department of Mathematics, January 2015

Andriyana, Y., Gijbels, I., Verhasselt, A.: P-splines quantile regression estimation in varying coefficient models. Test **23**, 153–194 (2014)

Andriyana, Y., Gijbels, I., Verhasselt, A.: Quantile regression in varying-coefficient models: non-crossing quantile curves and heteroscedasticity. Stat. Pap. (2016). doi:10.1007/s00362-016-0847-7

Arellano, M., Bond, S.: Some test of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econ. Stud. **58**, 277–297 (1991)

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

Davidian, M., Giltinan, D.M.: Nonlinear Models for Repeated Measurement Data. Chapman & Hall/CRC, New York (1995)

Eilers, P.H.C., Marx, B.: Flexible smoothing with B-splines and penalties. Stat. Sci. **11**, 89–102 (1996)

Fortin, M., Daigle, G., Ung, C.-H., Bégin, J., Archambault, L.: A variance-covariance structure to take into account repeated measurements and heteroscedasticity in growth modelling. Eur. J. For. Res. **126**, 573–585 (2007)

Gijbels, I., Ibrahim, M., Verhasselt, A.: Testing the heteroscedastic error structure in quantile varying coefficient models (2016) (under review)

Guo, J., Tian, M., Zhu, K.: New efficient and robust estimation in varying-coefficient models with heteroscedasticity. Stat. Sinica **22**, 1075–1101 (2012)

Hastie, T., Tibshirani, R.: Varying-coefficient models. J. R. Stat. Soc. Ser. B **55**, 757–796 (1993)

He, X.: Quantile curves without crossing. Am. Stat. **51**, 186–192 (1997)

Honda, T.: Quantile regression in varying coefficient models. J. Stat. Plan. Inference **121**, 113–125 (2004)

Hoover, D.R., Rice, J.A., Wu, C.O., Yang, L.-P.: Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika **85**, 809–822 (1998)

Huang, J.Z., Wu, C.O., Zhou, L.: Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Stat. Sinica **14**, 763–788 (2004)

Kim, M.-O.: Quantile regression with shape-constrained varying coefficients. Indian J. Stat. **68**, 369–391 (2006)

Kim, M.-O.: Quantile regression with varying coefficients. Ann. Stat. **35**, 92–108 (2007)

Kleiber, C., Zeileis, A.: Applied Econometrics with R. Springer, Berlin (2008)

Koenker, R.: Quantile Regression. Cambridge University Press, Cambridge (2005)

Liu, Y., Wu, Y.: Simultaneous multiple non-crossing quantile regression estimation using kernel constrains. J. Nonparametr. Stat. **23**, 415–437 (2011)

Oftedal, B., Walker, S.-E., Gram, F., McInnes, H., Nafstad, P.: Modelling long-term averages of local ambient air pollution in Oslo, Norway: evaluation of nitrogen dioxide, PM10 and PM2.5. Int. J. Environ. Pollut. **36**, 110–126 (2009)

Qu, A., Li, R.: Quadratic inference functions for varying-coefficient models with longitudinal data. Biometrics **62**, 379–391 (2006)

Schnabel, S.K., Eilers, P.H.C.: Simultaneous estimation of quantile curves using quantile sheets. AStA Adv. Stat. Anal. **97**, 77–87 (2013)

Şentürk, D., Müller, H.-G.: Inference for covariates adjusted regression via varying coefficient models. Ann. Stat. **34**, 654–679 (2006)

Şentürk, D., Müller, H.-G.: Functional varying coefficient models for longitudinal data. J. Am. Stat. Assoc. **105**, 1256–1264 (2010)

Stasinopoulos, D.M., Rigby, R.A.: Generalized additive models for location scale and shape (gamlls) in R. J. Stat. Softw. **23**, 1–46 (2007)

Wang, H.J., Zhu, Z., Zhou, J.: Quantile regression in partially linear varying coefficient models. Ann. Stat. **37**, 3841–3866 (2009)

Wu, Y., Liu, Y.: Stepwise multiple quantile regression estimation using non-crossing constraints. Stat. Interface **2**, 299–310 (2009)