

Shape testing in quantile varying coefficient models with heteroscedastic error

Peer-reviewed author version

Gijbels, Irène; IBRAHIM, Mohammed Abdulkerim & VERHASSELT, Anneleen (2017)

Shape testing in quantile varying coefficient models with heteroscedastic error. In:

Journal of nonparametric statistics, 29 (2), p. 391-406.

DOI: 10.1080/10485252.2017.1303066

Handle: <http://hdl.handle.net/1942/23410>

To appear in the *Journal of Nonparametric Statistics*
Vol. 00, No. 00, Month 20XX, 1–16

Shape testing in quantile varying coefficient models with heteroscedastic error

I. Gijbels^a, M.A. Ibrahim^b and A. Verhasselt^{b*}

^a*Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven, Belgium;* ^b*Censtat, I-BioStat, Universiteit Hasselt, Belgium*

(v4.0 released June 2015)

The interest is in regression quantiles in varying coefficient models for analyzing longitudinal data. The coefficients are allowed to vary with time, and the error variance (the variability function) varies with the covariates to allow for heteroscedasticity.

The functional coefficients are estimated using penalized splines (P-splines), not requiring specification of the error distribution. A likelihood-ratio-type test is considered to test the shape (constancy, monotonicity and/or convexity) of the functional coefficients. Further, testing procedures based on L_1 -norm, L_2 -norm and L_∞ -norm of the differences of the P-splines coefficients are considered to test for constant functional coefficients. These norm based tests perform better than the likelihood-ratio-type test in our simulation study. An extreme value test for testing monotonicity or convexity, also performs better than the likelihood-ratio-type test. The likelihood-ratio-type test is, however, useful when testing the shape of the coefficients in signal and in variability function simultaneously. A real data example demonstrates the testing procedures.

Keywords: Heteroscedasticity; Likelihood-ratio-test; Qualitative shape testing; Quantile regression; Varying coefficient models

AMS Subject Classification: 62G08; 62G10

1. Introduction

Recently, there is a growing interest on quantile regression, due to its robustness to outliers among others. Further, if most of the observations are concentrated, for instance, on 75th percentile of the distribution then it is more appropriate to consider 75% regression quantile than mean regression. Furthermore, if one is interested, for example, in studying severe malnutrition in children then lower quantiles can be of interest. For a linear regression model, Koenker and Bassett Jr (1978) utilized an asymmetric absolute loss function to define regression quantiles. The 50% quantile estimator corresponds to the conditional median estimator, which is more robust than the conventional least squares based conditional mean estimator. In contrast to the square loss function in mean regression, the check loss function of quantile regression is non-differentiable and asymmetric (except for the median). Yet, the regression quantiles are still easily computable by linear programming techniques proposed by Portnoy and Koenker (1997). They make use of a Frisch-Newton interior point algorithm, which is as fast as the least squares estimation.

A flexible linear model that allows the regression coefficients to be smooth functions

*Corresponding author. Email: anneleen.verhasselt@uhasselt.be

is a varying coefficient model (VCM). Allowing the coefficients to vary with one or more variables is helpful to depict dynamic behaviors and to overcome the limitation of a linear model. He and Shi (1994) used B-splines for estimating a smooth regression quantile. Then, to avoid over-fitting, Bollaerts, Eilers, and Aerts (2006); Andriyana, Gijbels, and Verhasselt (2014); Gijbels, Ibrahim, and Verhasselt (2016) make use of P-splines.

Investigating several regression quantiles is important in order to get a nuanced picture for the relationship between a response and covariates. However, this will be time consuming if the variance of the errors (the variability function) is of a homoscedastic structure (since it is sufficient, in this case, to consider one regression quantile to check the relationship). Gijbels et al. (2016) studied several structures of the variability function, like power and exponential functions; proposing also a likelihood-ratio-type test to choose between two variability structures. In this study, we want to test the constancy of the coefficients in the variability function as well as in the signal part of the model (median function). We investigate also other type of tests in addition to the likelihood-ratio-type test proposed in Gijbels et al. (2016).

Furthermore, we discuss shape testing for the coefficient of a specific covariate, such as a monotonicity test to check whether a covariate has a non-decreasing/non-increasing effect over the evolution of another covariate. An estimation of monotone B-spline smoothing in a simple non-parametric regression was pioneered by He and Shi (1998). He and Ng (1999) extended these results to a general framework for constrained L_1 -norm regression. The extended framework includes monotonicity, convexity/concavity, periodicity and pointwise constraints. As a consequence, they developed an algorithm called COBS (constrained B-splines smoothing) to S-plus users. Since the above results were for the case of univariate smoothing, where the unknown function is a function of a single variable, Kim (2006) extended them to varying coefficients. Bollaerts et al. (2006) proposed, for the univariate case, an estimation technique for non-parametric monotone quantile regression using P-splines. However, to the best of our knowledge there is no literature on how to test the monotonicity of a coefficient. This testing procedure is important in a real life application, for example, to test whether the weight of a child is significantly decreasing, at some points in time with age of the child.

The rest of the paper is organized as follows. In Section 2, we discuss the considered model, and the estimation method of the signal part and of the variability part of the model. Section 3 deals with testing procedures for constancy, monotonicity and convexity/concavity. Simulation studies are carried out in Section 4. The testing procedures in Section 3 are applied on a real data example in Section 5. Finally, Section 6 presents a further discussion and conclusions.

2. Framework

2.1. *Heteroscedastic model*

We investigate quantile regression in varying coefficient models with a heteroscedastic error structure. We consider a longitudinal observational data scheme, where each subject has measurements at different time points. In this data scheme, one typically assumes that the measurements are independent for different subjects but measurements at different time points within each subject can be correlated.

We consider the varying coefficient model:

$$Y(T) = \mathbf{X}^T(T)\boldsymbol{\beta}(T) + V(\mathbf{X}(T), T)\epsilon(T), \quad (1)$$

where, we have longitudinal observations $(Y(t_{ij}), \mathbf{X}(t_{ij}), t_{ij})$ of $(Y(T), \mathbf{X}(T), T) \in \mathbb{R} \times \mathbb{R}^{p+1} \times \mathbb{R}$ with $\mathbf{X}(T) = (X^{(0)}(T), X^{(1)}(T), \dots, X^{(p)}(T))^T$ and $X^{(0)}(T) \equiv 1$, and errors $\epsilon(t_{ij})$ of $\epsilon(T) \in \mathbb{R}$ for $i = 1, \dots, n$, $j = 1, \dots, N_i$, and $p \geq 1$, 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, $Y(t_{ij})$ and $\mathbf{X}(t_{ij}) = (X^{(0)}(t_{ij}), X^{(1)}(t_{ij}), \dots, X^{(p)}(t_{ij}))^T$ are the observed outcome and covariates of the i th subject at time point t_{ij} . Here $\mathbf{A}^T$ denotes the transpose of a vector or matrix $\mathbf{A}$. The time points t_{ij} take values in a bounded space $\mathcal{T} \subset \mathbb{R}$ and that without loss of generality we can take this space to be $[0, 1]$, and $\beta(T) = (\beta_0(T), \beta_1(T), \dots, \beta_p(T))^T$ is the vector of unknown regression coefficient functions at time T , with $\beta_0(T)$ the baseline effect.

The τ th conditional quantile of $Y(T)$ given $(\mathbf{X}(T), T)$, $\beta_k(T)$ (with $k = 0, \dots, p$) and $V(\mathbf{X}(T), T)$ are estimated with P-splines. The variability function of the following form will be considered (the “full” model in Gijbels et al. (2016)):

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

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

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

2.2. Estimation Methods

The above model (1) consists of three parts, the signal $\mathbf{X}^T(T)\beta(T)$, the variability function $V(\mathbf{X}(T), T)$ and the error term $\epsilon(T)$. Estimating all the components simultaneously can be tricky (due to identifiability issues). We use the Adaptive He (AHe) approach (Andriyana and Gijbels 2016; Andriyana, Gijbels, and Verhasselt 2016; Gijbels et al. 2016) to deal with the identifiability problem. This approach has two advantages: it avoids crossings of the estimated quantile regression curves when estimating several quantiles and the variability function can be estimated. Furthermore, since the AHe approach relates all quantile functions through the conditional median, substantial savings in computation costs can be achieved when multiple quantiles of high-dimensional data are needed.

Without loss of generality, the AHe approach is based on the following assumptions:

(H1): The conditional median quantile of the error term equals zero: $q_{0.5}\{\epsilon(T)|\mathbf{X}(T), T\} = 0$.

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

The AHe procedure consists of the following three steps (see also Gijbels et al. (2016)):

Step 1: Assuming (H1), the median quantile function of $Y(T)$ is given by:

$$q_{0.5}(Y(T)|\mathbf{X}(T), T) = \mathbf{X}^T(T)\beta(T) + V(\mathbf{X}(T), T)q_{0.5}(\epsilon(T)|\mathbf{X}(T), T) = \mathbf{X}^T(T)\beta(T).$$

We approximate $\beta_k(T)$ via normalized B-splines ($\mathbf{B}_k(T; \nu_k) = (B_{k1}(T; \nu_k), \dots, B_{km_k}(T; \nu_k))^T$) of degree ν_k , for $k = 0, 1, \dots, p$:

$$\beta_k(T) \approx \sum_{l=1}^{m_k} \alpha_{kl} B_{kl}(T; \nu_k) = \boldsymbol{\alpha}_k^T \mathbf{B}_k(T; \nu_k), \quad (2)$$

where $\boldsymbol{\alpha}_k = (\alpha_{k1}, \dots, \alpha_{km_k})^T$ denotes the coefficient vector of the associated B-splines

$\mathbf{B}_k(T; \nu_k)$, and $m_k = u_k + \nu_k$, where we have $u_k + 1$ equidistant knots for the k th component ($X^{(k)}(T)$).

The estimated coefficient vectors $\hat{\boldsymbol{\alpha}}_k = (\hat{\alpha}_{k1}, \dots, \hat{\alpha}_{km_k})^T$ ($k = 0, \dots, p$) are obtained by minimizing,

$$S(\boldsymbol{\alpha}) = \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5} \left(Y(t_{ij}) - \sum_{k=0}^p X^{(k)}(t_{ij}) \boldsymbol{\alpha}_k^T \mathbf{B}_k(t_{ij}; \nu_k) \right) + \sum_{k=0}^p \lambda_k \| \mathbf{D}_{m_k}^{d_k} \boldsymbol{\alpha}_k \|_1 \quad (3)$$

with respect to $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_p^T)^T$, where $\rho_\tau(z) = z[\tau - I(z < 0)]$ is the check loss function with $I(A)$ the indicator of A , $\lambda_k > 0$ is the smoothing parameter for the k th component, d_k is the differencing order in the penalty term, and (for $d_k = 1$)

$$\mathbf{D}_{m_k}^1 = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(m_k-1) \times m_k}.$$

Step 2: From model (1) and since $V(\mathbf{X}(T), T) \geq 0$, assuming **(H2)**, we can estimate $V(\cdot, \cdot)$ based on the following equation,

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

with $\theta_0(T) = \ln\{\gamma(T)\}$. Then $\theta_k(T)$ can be approximated in a basis (of size m_k^V) normalized B-splines of degree ν_k^V ($\theta_k(T) \approx \sum_{l=1}^{m_k^V} \alpha_{kl}^V B_{kl}^V(T; \nu_k^V)$), for $k = 0, 1, \dots, p$.

The estimated parameters are obtained by solving

$$\min_{\boldsymbol{\alpha}_1^V, \dots, \boldsymbol{\alpha}_p^V} \left\{ \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5} \left(\ln(|Y(t_{ij}) - \mathbf{X}^T(t_{ij}) \hat{\boldsymbol{\beta}}(t_{ij})|) - \sum_{k=0}^p (\boldsymbol{\alpha}_k^V)^T \mathbf{B}_k^V(t_{ij}; \nu_k^V) X^{(k)}(t_{ij}) \right) + \sum_{k=0}^p \sum_{l=d_k^V+1}^{m_k^V} \lambda_k^V |\Delta^{d_k^V} \alpha_{kl}^V| \right\}, \quad (4)$$

where $\boldsymbol{\alpha}_k^V = (\alpha_{k1}^V, \dots, \alpha_{km_k^V}^V)^T$ ($k = 0, \dots, p$), $\mathbf{B}_k^V(T; \nu_k^V) = (B_{k1}^V(T; \nu_k^V), \dots, B_{km_k^V}^V(T; \nu_k^V))^T$, $\lambda_k^V > 0$ is the smoothing parameter for the k th component, d_k^V is the differencing order in the penalty term. Note that we replaced $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$ in (4). Hence, the efficiency of the estimated variability function relies on the performance of the estimated signal.

Step 3: The τ th conditional quantile regression estimate is obtained, for various values of $\tau \in (0, 1)$, using the estimates from the previous two steps,

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

We approximate the unknown conditional τ th quantile of the error term $\epsilon(T)$ ($a^\tau(T) =$

$q_\tau(\epsilon(T)|\mathbf{X}(T), T)$) using $m^q = (u^q + \nu^q)$ B-spline basis functions of degree ν^q with $u^q + 1$ equidistant knots ($a^\tau(T) \approx \sum_{l=1}^{m^q} \alpha_l^q B_l^q(T; \nu^q)$), and the coefficients $(\alpha_1^q, \alpha_2^q, \dots, \alpha_{m^q}^q)$ are estimated by minimizing:

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

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

The implementation of the estimation procedure and the choice of the smoothing parameters are discussed in Gijbels et al. (2016). The consistency of the estimators for the conditional quantile as well as for the variability function is also shown in Gijbels et al. (2016).

3. Testing

For the signal, the hypotheses are defined as:

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

where $\mathcal{M}_0$ and $\mathcal{M}_1$ are sets of functions defined as,

$$\begin{aligned} \mathcal{M}_0 &= \{q_{0.5}(\cdot) | q_{0.5}(\cdot) \text{ equals the signal under } H_0\}, \text{ and} \\ \mathcal{M}_1 &= \{q_{0.5}(\cdot) | q_{0.5}(\cdot) \text{ equals the signal under } H_1\}. \end{aligned}$$

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

3.1. Testing constancy of the coefficient functions

In this section, four different types of testing procedures for the signal as well as the variability function are discussed. They are the likelihood-ratio-type (LRT) (considered by Kim (2007); Gijbels et al. (2016)), the L_1 , the L_2 and the $L_{\max}$ (discussed in Section 3.1.2 below) tests.

3.1.1. The LRT test

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

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

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

The null hypothesis is rejected when G_m is too large, as in Gijbels et al. (2016). The p-value is obtained by a re-sampling subject bootstrap:

- (1) Re-sample n subjects with replacement from $i = 1, \dots, n$, to obtain the b th bootstrap sample

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

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

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

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

- (2) Repeat the above sampling procedure B times (hence $b = 1, \dots, B$).
- (3) Calculate the test statistic G_m^b for each b th bootstrap sample to obtain its empirical distribution.
- (4) Get the p-value using the empirical probability of $G_m^b \geq G_m$.

Similarly, we can conduct the same procedure for the variability function (see Gijbels et al. (2016)). The test statistic is also defined similarly but changing $(Y(t_{ij}), \hat{m}_0(t_{ij})$ and $\hat{m}_1(t_{ij}))$ to $(\ln |Y(t_{ij}) - \mathbf{X}^T(t_{ij})\hat{\beta}(t_{ij})|, \hat{v}_0(t_{ij})$ and $\hat{v}_1(t_{ij}))$, where $\hat{v}_0(\cdot)$ and $\hat{v}_1(\cdot)$ are the estimated variability functions under H_0 and under H_1 , respectively. The p-value is obtained by a re-sampling subject bootstrap, similar to the test for the signal, but now we obtain the bootstrap sample

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

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

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

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

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

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

We discuss in this section three types of test statistics for testing that a coefficient function is constant or varying. When the coefficients α_k are a constant vector, the corresponding spline $\beta_k(\cdot)$ is constant. From de Boor (2001), we have the first order derivative of $\beta_k(T)$ in (2), with distance $1/u_k$ between the equidistant knots (for $T \in [0, 1]$), given by

$$\beta'_k(T) = u_k \sum_{j=1}^{m_k-1} \Delta^1 \alpha_{k(j+1)} \mathbf{B}_k(T; \nu_k - 1) = u_k \mathbf{B}_k(T; \nu_k - 1)^T \mathbf{D}_{m_k}^1 \alpha_k. \quad (5)$$

Degree one B-splines have support at three knots, and equal zero at the two end knot points and one at the middle knot point. Therefore, for quadratic B-splines and by (5), $\beta'_k(t) \neq 0$ for all $t \in \mathcal{T}$ if and only if $\mathbf{D}_k^1 \boldsymbol{\alpha}_k \neq \mathbf{0}$.

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

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

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

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

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

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

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

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

3.2. Monotonicity tests

By (5), as u_k and $\mathbf{B}_k(T; \nu_k - 1)$ are all positive by definition, restricting $\Delta^1 \alpha_{k(j+1)}$ to be positive (resp. negative) is a sufficient condition for $\beta'_k(T)$ to be positive (resp. negative). Hence, to restrict $\beta_k(T)$ to be monotone increasing or decreasing, we use the objective function considered by Bollaerts et al. (2006), which adds asymmetric weights ω_{kj} to the objective function in (3)

$$S_m(\boldsymbol{\alpha}) = \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5} \left(Y(t_{ij}) - \sum_{k=0}^p X^{(k)}(t_{ij}) \boldsymbol{\alpha}_k^T \mathbf{B}_k(t_{ij}; \nu_k) \right) + \sum_{k=0}^p \lambda_k \|\mathbf{D}_{m_k}^{d_k} \boldsymbol{\alpha}_k\|_1 + \sum_{k=0}^p \kappa_k \sum_{j=2}^{m_k-1} \omega_{kj} |\Delta^1 \alpha_{kj}| \quad (6)$$

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

$$\omega_{kj} = \begin{cases} 0 & \text{if } \Delta^1 \alpha_{kj} \geq 0 \text{ for increasing (resp. } \Delta^1 \alpha_{kj} \leq 0 \text{ for decreasing),} \\ 1 & \text{otherwise.} \end{cases} \quad (7)$$

The implementation of the estimation method for the objective function in (6), for a model with only one covariate, is discussed in Section 3 of Bollaerts et al. (2006). The implementation can be extended, for a model with several covariates, by adding the additional constraints in (7) for each coefficient of the covariates.

To test the monotonicity of β_k , we discuss two types of tests. The first one is the likelihood-ratio-type test (LRT) used by Gijbels et al. (2016) (defined in Section 3.1.1 of the current paper). The second test is the algorithm used by Ahkim, Gijbels, and Verhasselt (2016) for mean regression, which we will call Extreme value test (EVT).

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

$$H_0 : \beta'_k(t) \geq 0 \text{ for all } t \in \mathcal{T} \quad \text{Versus} \quad H_1 : \neg H_0. \quad (8)$$

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

$$H_0 : \beta'_k(t) \leq 0 \text{ for all } t \in \mathcal{T} \quad \text{Versus} \quad H_1 : \neg H_0. \quad (9)$$

Degree one B-splines have support at three knots, and equal zero at the two end knot points and one at the middle knot point. Therefore, for quadratic B-splines and by (5), β_k is non-decreasing if and only if $\mathbf{D}_k^1 \boldsymbol{\alpha}_k \geq \mathbf{0}$. Then, the algorithm for the testing procedure reads as follows:

- (1) $d_{\min} = \min_{1 \leq j \leq m_k-1} (\Delta^1 \hat{\alpha}_{kj})$
- (2) If $d_{\min} \geq 0$, the p-value equals one
- (3) If $d_{\min} < 0$, determine the distribution of $d_{\min}$ under H_0
- (4) Get the p-value using the empirical probability of $d_{\min}^b \leq d_{\min}$, where $d_{\min}^b$ is the test statistic based on the b th bootstrap sample obtained under the null hypothesis.

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

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

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

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

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

where $\hat{\beta}_k^m(t_{ij})$ and $\hat{\beta}_k(t_{ij})$ are the monotone constrained (solution of (6)) and unconstrained estimated coefficients, respectively.

When we use B-splines of degree three or more, our test is based on $d_{\min} = \min_{t \in \mathcal{T}} (\mathbf{B}_k(t; \nu_k - 1)^T \mathbf{D}_{m_k}^1 \boldsymbol{\alpha}_k)$. The test for hypothesis in (9) can be carried out similarly with the obvious adjustment of the inequality and now using $d_{\max} = \max_{1 \leq j \leq m_k-1} (\Delta^1 \hat{\alpha}_{kj})$.

3.3. Convexity/Concavity tests

The second derivative of $\beta_k(T)$ in (2), with distance $1/u_k$ between the equidistant knots (for $T \in [0, 1]$) is given by

$$\beta_k''(T) = u_k^2 \sum_{j=1}^{m_k-2} \Delta^2 \alpha_{k(j+2)} \mathbf{B}_k(T; \nu_k - 2).$$

Hence, to restrict $\beta_k(T)$ to be convex or concave, we use the following objective function, which adds asymmetric weights ϑ_{kj} to the objective function in (3)

$$\begin{aligned} S_c(\boldsymbol{\alpha}) = \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_{0.5} \left(Y(t_{ij}) - \sum_{k=0}^p X^{(k)}(t_{ij}) \boldsymbol{\alpha}_k^T \mathbf{B}_k(t_{ij}; \nu_k) \right) &+ \sum_{k=0}^p \lambda_k \|\mathbf{D}_{m_k}^{d_k} \boldsymbol{\alpha}_k\|_1 \\ &+ \sum_{k=0}^p \kappa_k^c \sum_{j=2}^{m_k-1} \vartheta_{kj} |\Delta^2 \alpha_{kj}| \end{aligned} \quad (10)$$

with κ_k^c a user-defined constraint parameter and

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

The implementation of the estimation method for the objective function in (10) can also be written similarly as that of monotone quantile regression but now using the second order differences.

Here also we can use the two tests discussed in Section 3.2, but now we work with the second order derivative (for EVT test). The hypothesis for testing β_k is a convex function reads as:

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

The EVT test, for cubic splines, is given as $d_{\min} = \min_{1 \leq j \leq m_k-2} (\Delta^2 \hat{\alpha}_{kj})$. When splines of degree four or more are used the test is based on $d_{\min} = \min_{t \in \mathcal{T}} (\mathbf{B}_k(t; \nu_k - 2)^T \mathbf{D}_{m_k}^2 \boldsymbol{\alpha}_k)$.

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

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

3.4. Shape testing in both signal and variability function

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

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

The test statistic is defined as:

$$G = 2 \left\{ \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \left[\rho_{0.5} \{Y(t_{ij}) - \hat{m}_0(t_{ij})\} - \rho_{0.5} \{Y(t_{ij}) - \hat{m}_1(t_{ij})\} \right. \right. \\ \left. \left. + \rho_{0.5} \{ \ln |Y(t_{ij}) - \mathbf{X}^T(t_{ij}) \hat{\beta}(t_{ij})| - \hat{v}_0(t_{ij}) \} - \rho_{0.5} \{ \ln |Y(t_{ij}) - \mathbf{X}^T(t_{ij}) \hat{\beta}(t_{ij})| - \hat{v}_1(t_{ij}) \} \right] \right\},$$

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

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

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

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

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

where $\hat{\beta}_k^c(\cdot)$ and $\hat{\beta}_k(\cdot)$ are, respectively, the convex constrained (solution of (10)) and un-constrained estimated coefficients for the signal, and $\hat{v}_2(\cdot)$ is an estimate for the most complex variability function.

4. Simulation study

In this section we conduct simulation studies to investigate the performances of the testing procedures discussed in Section 3. Three simulation settings are considered (i) testing constancy of the coefficients in the signal as well as in the variability function; (ii) monotonicity test for the coefficients in the signal; (iii) testing, simultaneously, for convexity of the functional coefficient of a covariate in the signal and constancy of the functional coefficient of a covariate in the variability function. To analyze the datasets B-splines of degree two with five equidistant knots on the time interval and differencing order 1 are used.

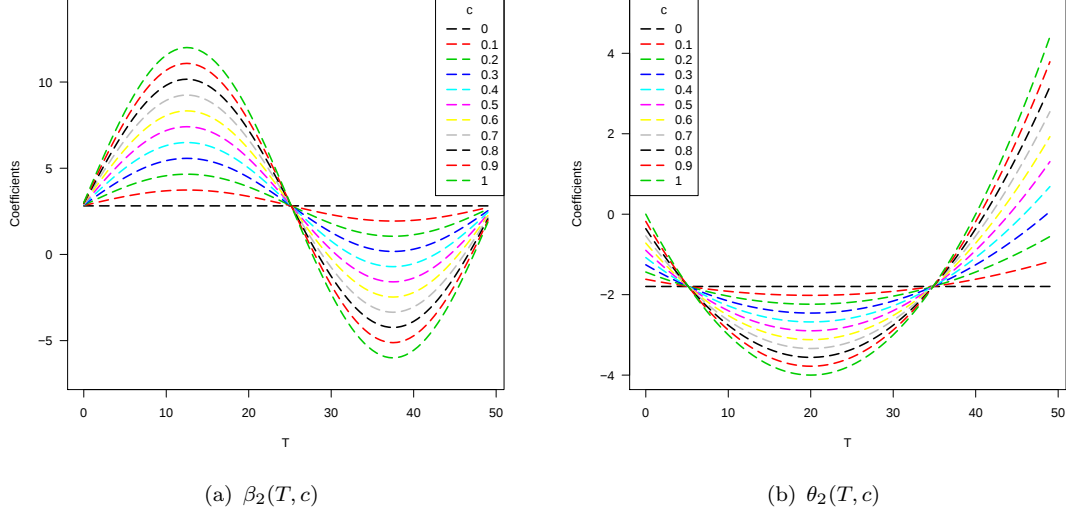
4.1. Simulation study: Constancy

We investigate the performances of the testing procedures in Section 3.1. We further compare the testing procedures with the rank score test (RS) of Wang, Zhu, and Zhou (2009) (using compound symmetry error structure, and cubic splines for B-spline estimators). A simulation setting (defined in Table 1) of the following type is used:

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

Table 1. The description of the coefficients.

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

Figure 1. The true coefficients of $X^{(2)}(T)$ for the signal (left panel) and for the variability function (right panel) for various values of c .

To have equivalent complexity in the data for the simulated samples, the coefficients are formulated such that the signal to noise ratio (SNR),

$$SNR = \frac{\text{Sample variance of } \sum_{k=0}^2 \beta_k(T) X^{(k)}(T)}{\text{Sample variance of } V(\mathbf{X}(T), T) \epsilon(T)},$$

is approximately seven.

The error term is generated from a transformed multivariate normal distribution given by (such that we attain the two assumptions **(H1)** and **(H2)**):

$$\epsilon(T) = \frac{(\zeta(T) - q_{0.5}(\zeta(T)))}{q_{0.5}(|\zeta(T) - q_{0.5}(\zeta(T))|)},$$

where $\zeta(T) \sim N(0, 0.6^2)$ and $\text{Cov}(\zeta(t_{ij}), \zeta(t_{ik})) = 0.6^2/5$, $1 \leq i \leq n$, $1 \leq j, k \leq N_i$, $j \neq k$.

The following time dependent covariates are considered

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

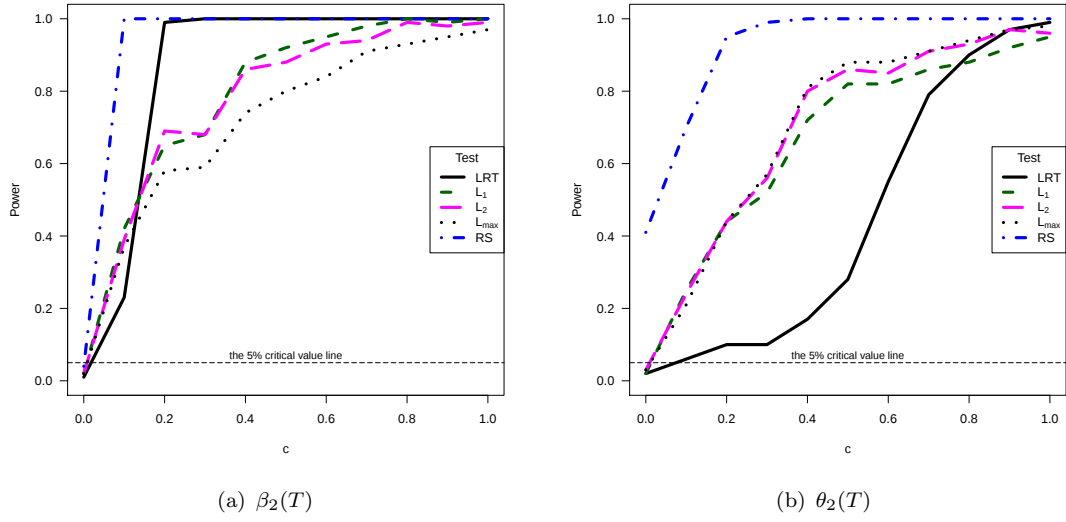


Figure 2. The power curves for $\beta_2(T)$ and $\theta_2(T)$.

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

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

to have them on a comparable scale, which is important to determine the individual smoothing parameters λ_k 's, chosen via the Schwarz information criterion in Gijbels et al. (2016).

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

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

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

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

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

For $\beta_2(T)$, we see in Figure 2 (a) that all the tests preserve the nominal significance level for $c = 0$ and the power increases to 1 for large values of c . Note that the power

Table 2. Proportion of significant tests.

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

curve of the RS test outperforms these of the other tests. Among these, the L_1 , L_2 and $L_{\max}$ tests perform better than the LRT test for small values of c , whereas the latter test performs best for moderate to large values of c . Since the power curves of the L_1 , L_2 , $L_{\max}$ and LRT tests are crossing, there is no uniformly best test among them. For the variability function (see Figure 2(b)) we see that all tests other than the RS test preserve the nominal level. Among the tests that best keep the nominal level, we observe that the power properties of the L_1 , L_2 and $L_{\max}$ tests are substantially better than these of the LRT test, for a large range of c -values. The power further increases to 1 for c large for all tests.

4.2. Simulation study: Monotonicity

In this section we conduct a simulation study to investigate the performances of the testing procedures discussed in Section 3.2, for the signal. We consider a homoscedastic varying coefficient model with $V(\cdot, \cdot) = 0.4$. We use the coefficient functions $\beta_0(T) = 0.25 + 2T$ and $\beta_2(T) = -0.5 + 10(T - 0.5)^2$. We consider four simulation settings to test the hypothesis in (8) for $\beta_1(T)$, which are based on the following two functions,

- $f_{1,a}(T) = -2 + 2(1 + T - a \exp(-50(T - 0.5)^2))$ and
- $f_2(T) = 1.1$.

The first three simulation settings are based on $f_{1,a}(T)$, which is taken from Bowman, Jones, and Gijbels (1998). This function is strictly increasing for $a = 0.15$, a dip appears when $a = 0.3$ and the dip appears more profoundly with $a = 0.45$. The fourth simulation setting is based on $f_2(T)$. The coefficients are formulated such that the signal to noise ratio is approximately seven.

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

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

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

To avoid the violation of the monotonicity constraint we use $\kappa_2 = 1000$.

We present the proportion of significant tests for each testing procedure under the four simulation settings in Table 2. Under the monotone functions $f_2(T)$ and $f_{1,0.15}(T)$, we see that the $P(\text{rejecting } H_0 | H_0)$ is around the nominal significance level for both tests. Whereas, under the non-monotone functions, $f_{1,0.3}(T)$ and $f_{1,0.45}(T)$, the power of the test is around one, except for LRT having less power under $f_{1,0.3}(T)$. Here we see that $d_{\min}$ test perform better than the LRT test.

Table 3. P-values for constancy.

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

4.3. Simulation study: convex in signal and constant in variability

Using the LRT testing procedure, as discussed in Section 3.4, we test whether a covariate has a coefficient function that is convex for the signal and is constant for the variability function. We consider the simulation setting of the previous section and take $\beta_1(T)$ equal to $f_{1,0.45}(T)$. The proportion of significant tests, respectively, are 0.01 and 0.97 for $X^{(0)}(T)$ ($H_0 : \beta_0''(t) \geq 0$ & $\theta_0'(t) = 0$ for all $t \in \mathcal{T}$) and $X^{(1)}(T)$ ($H_0 : \beta_1''(t) \geq 0$ & $\theta_1'(t) = 0$ for all $t \in \mathcal{T}$). These results show that the test has good performance, since the nominal significance level was respected and the power is close to one.

5. Real data example

In this section we illustrate the use of the testing procedures in Section 3 on a real data example. To analyze the data set, B-splines of degree two with five equidistant knots on the time interval and differencing order one are used.

The data example is taken from the National Longitudinal Survey of Youth, collected in USA. It was originally reported in Murnane, Willett, and Boudett (1999). It consists of 888 individuals of age 14 to 17 years old. Each individual has a different number of observations (N_i) ranging between 1 and 13. The response variable $Y(t_{ij})$ is the hourly wage, with the time variable t_{ij} denoting the duration of the work experience (in years), $i = 1, \dots, 888$ and $j = 1, \dots, N_i$. We have two predictor variables. The first one is the *race* variable with three levels (*black*, *Hispanic* and *white*), hence generating two dummy variables – $(X^{(1)}(T), X^{(2)}(T)) = (1, 0)$, $(X^{(1)}(T), X^{(2)}(T)) = (0, 1)$ and $(X^{(1)}(T), X^{(2)}(T)) = (0, 0)$ standing for *black*, *Hispanic* and *white*, respectively. The second one is $X^{(3)}(T)$ (*hgc*) the highest grade completed by the individual. The covariates are standardized as in Section 4.1, so that the estimated coefficients have an equivalent scale.

We check the constancy of the coefficients for the signal and/or the variability function (Table 3) using the four tests in Section 3.1. The four tests lead to similar conclusions, with exception of the LRT test for testing constancy of the coefficient $\theta_0(T)$. Our conclusion is based on the L_1 , L_2 and $L_{\max}$ tests results, since these tests performed better than the LRT test in the simulation study. The aforementioned tests indicate that all the covariates have non-varying coefficients with respect to both signal and variability function.

Further, we check the monotonicity or convexity of $\beta_0(T)$ in Table 4. Using both tests in Sections 3.2 and 3.3, we do not have enough arguments to reject the hypothesis that the coefficient function is non-decreasing or convex. Similar result was obtained by Ahkim et al. (2016) for mean regression, where they have shown that the coefficient is non-decreasing. This means, for *white* individuals who have lower *hgc* level, that as the

Table 4. P-values for monotonicity and convexity.

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

duration of work experience increases the hourly wage is non-decreasing.

6. Discussion

We have considered constancy checking for the coefficients in the general heteroscedastic model proposed by Gijbels et al. (2016), using several types of tests. We compared the performance of the LRT proposed by Gijbels et al. (2016) and RS test of Wang et al. (2009) with the other three tests proposed in this paper. These tests are applied on the signal (median function) as well as on the variability function. We applied the testing procedures in a simulation study and a real data example. For the signal, the simulation study in Section 4.1 shows that the three tests (L_1 , L_2 and $L_{\max}$) studied in this paper have a performance somewhat comparable to that of the other two tests, although none is dominating (cf the crossing power curves). For the variability function however, the proposed tests have better overall (size and power) performances.

Further, we proposed two testing procedures for testing for monotonicity or convexity. Both tests perform well in the simulation study (Sections 4.2). Furthermore, we investigated a coefficient's convexity in the signal and its constancy in the variability function (Section 4.3), using the LRT test procedure in Section 3.4. This test can be generalized to any other shapes either in the signal or the variability function. For instance, we can test monotonicity of a coefficient function in the signal and constancy of a coefficient function in the variability function, simultaneously.

Although the proposed testing procedures have good performances in the simulation studies, it is of interest to prove their consistency. This is out of scope of this paper and is left for future research.

Acknowledgements

The authors gratefully acknowledge support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy). I. Gijbels acknowledges support from the KU Leuven Research Council (GOA/12/014). A. Verhasselt acknowledges support from the FWO research grant 1518917N. M.A. Ibrahim and A. Verhasselt acknowledge support from the Special Research Fund (BOF) of Hasselt University [14NO6VHAA]. The infrastructure of the VSC - Flemish Supercomputer Center, funded by the Hercules Foundation and the Flemish Government - department EWI, was used for the simulations.

References

- Ahkim, M., Gijbels, I., and Verhasselt, A. (2016), 'Shape Testing in in Varying Coefficient Models', *Test*, to appear. Online at <http://link.springer.com/article/10.1007/s11749-016-0518-y>.
Andriyana, Y., and Gijbels, I. (2016), 'Quantile regression in heteroscedastic vary-

- ing coefficient models', *AStA Advances in Statistical Analysis*, to appear. Online at <http://link.springer.com/article/10.1007/s10182-016-0284-x>.
- Andriyana, Y., Gijbels, I., and Verhasselt, A. (2014), 'P-splines quantile regression estimation in varying coefficient models', *Test*, 23, 153–194.
- Andriyana, Y., Gijbels, I., and Verhasselt, A. (2016), 'Quantile regression in varying coefficient models: non-crossingness and heteroscedasticity', *Statistical Papers*, to appear. Online at <http://link.springer.com/article/10.1007/s00362-016-0847-7>.
- Bollaerts, K., Eilers, P.H., and Aerts, M. (2006), 'Quantile regression with monotonicity restrictions using P-splines and the L1-norm', *Statistical Modelling*, 6, 189–207.
- Bowman, A., Jones, M., and Gijbels, I. (1998), 'Testing monotonicity of regression', *Journal of computational and Graphical Statistics*, 7, 489–500.
- de Boor, C. (2001), *A Practical Guide to Splines*, Springer, New York.
- Gijbels, I., Ibrahim, M.A., and Verhasselt, A. (2016), 'Testing the heteroscedastic error structure in quantile varying coefficient models', Submitted.
- He, X., and Ng, P. (1999), 'COBS: qualitatively constrained smoothing via linear programming', *Computational Statistics*, 14, 315–338.
- He, X., and Shi, P. (1994), 'Convergence rate of b-spline estimators of nonparametric conditional quantile functions', *Journal of Nonparametric Statistics*, 3, 299–308.
- He, X., and Shi, P. (1998), 'Monotone B-spline smoothing', *Journal of the American statistical Association*, 93, 643–650.
- Kim, M.O. (2006), 'Quantile regression with shape-constrained varying coefficients', *Sankhyā: The Indian Journal of Statistics*, 68, 369–391.
- Kim, M.O. (2007), 'Quantile regression with varying coefficients', *The Annals of Statistics*, 35, 92–108.
- Koenker, R., and Bassett Jr, G. (1978), 'Regression quantiles', *Econometrica*, 46, 33–50.
- Murnane, R.J., Willett, J.B., and Boudett, K.P. (1999), 'Do male dropouts benefit from obtaining a GED, postsecondary education, and training?', *Evaluation Review*, 23, 475–503.
- 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.
- Wang, H.J., Zhu, Z., and Zhou, J. (2009), 'Quantile regression in partially linear varying coefficient models', *The Annals of Statistics*, 37, 3841–3866.