Shape testing for coefficient functions in varying coefficient models

Irène Gijbels

KU Leuven
Department of Mathematics and Leuven Statistics Research Centre
Belgium

January 30, 2018

Joint work with M. Akhim, M. Ibrahim and A. Verhasselt, UHasselt

Outline

- Introduction: varying coefficient models
- Unconstrained estimation in varying coefficient models
- Constrained estimation in varying coefficient models
- Shape testing in varying coefficient models
- Quantile regression in varying coefficient models

... particular shape testing ...

varying coefficient regression model

Y response variable $X^{(1)}, \ldots, X^{(p)}$ covariates

multiple linear regression model: $Y = \beta_0 + \beta_1 X^{(1)} + \ldots + \beta_p X^{(p)} + \varepsilon$

complex data

flexible modelling $\longrightarrow$ **varying coefficient regression model**:

$$Y(\mathbf{T}) = \beta_0(\mathbf{T}) + \beta_1(\mathbf{T}) X^{(1)}(\mathbf{T}) + \ldots + \beta_p(\mathbf{T}) X^{(p)}(\mathbf{T}) + \varepsilon(T)$$

$(Y(T), X^{(1)}(T), \ldots, X^{(p)}(T), T)$ random vector

T takes values in $[0, 1]$ (without loss of generality)

Hastie & Tibshirani (1993), Hoover *et al.* (1998), ... , Honda (2004), Kim (2006), ..., Wang *et al.* (2008), ..., Antoniadis, G. & Verhasselt (2012a), Andriyana (2015), Xie *et al.* (2015), ...

$$
\begin{aligned}
Y(t) &= \beta_0(t) + \beta_1(t)X^{(1)}(t) + \ldots + \beta_p(t)X^{(p)}(t) + \varepsilon(t) \\
&= \mathbf{X}(t)^{\mathsf{T}}\boldsymbol{\beta}(t) + \varepsilon(t)
\end{aligned}
$$

where $\mathbf{X}(t) = \left(\underbrace{X^{(0)}(t)}_{\equiv 1}, X^{(1)}(t), \ldots, X^{(p)}(t) \right)^{\mathsf{T}}$

$\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \ldots, \beta_p(t))^{\mathsf{T}}$

vector of $(p+1)$ unknown **univariate** regression coefficients at time t

$\beta_0(t)$ is the baseline effect

assume that $\varepsilon(t)$ is a **mean zero** stochastic process at time t

first aim: estimate the **mean regression function**

$$
E(Y(t)|\mathbf{X}(t), t) = \beta_{\mathbf{0}}(\mathbf{t}) + \beta_{\mathbf{1}}(\mathbf{t})X^{(1)}(t) + \ldots + \beta_{\mathbf{p}}(\mathbf{t})X^{(p)}(t)
$$

observational setting: longitudinal data setup

n independent subjects/individuals

for each individual i: measurements repeated over a time period

measurements at time points $t_{i1}, \ldots, t_{iN_i}$

N_i different measurements for response and all explanatory variables:

$Y(t_{ij}) = Y_{ij}$

$X^{(k)}(t_{ij}) = X_{ij}^{(k)} \quad k = 1, \ldots, p \Longrightarrow \mathbf{X}(t_{ij}) \stackrel{\text{not.}}{=} \mathbf{X}_{ij} = (X_{ij}^{(0)}, \ldots, X_{ij}^{(p)})^\mathsf{T}$

total number of observations over all individuals:

$$N = \sum_{i=1}^{n} N_i$$

$$E(Y(t)|\mathbf{X}(t),t) = \boldsymbol{\beta_0(t)} + \boldsymbol{\beta_1(t)}X^{(1)}(t) + \ldots + \boldsymbol{\beta_p(t)}X^{(p)}(t)$$

suppose: each unknown function $\beta_k(t)$, $k = 0, \ldots, p$, can be represented by a B-spline basis expansion

$$\beta_k(t) = \alpha_{k1}B_{k1}(t;\nu_k) + \ldots + \alpha_{km_k}B_{km_k}(t;\nu_k) = \sum_{\ell=1}^{m_k} \alpha_{k\ell}B_{k\ell}(t;\nu_k)$$

$$= \boldsymbol{\alpha}_k^T \mathbf{B}_k(t;\nu_k)$$

$\boldsymbol{\alpha_k} = (\alpha_{\mathbf{k1}}, \ldots, \alpha_{\mathbf{km_k}})^\mathsf{T}$ $\quad$ $\mathbf{B}_k(t;\nu_k) = (B_{k1}(t;\nu_k), \ldots, B_{km_k}(t;\nu_k))^\mathsf{T}$

$m_k = u_k + \nu_k$ $\quad$ $u_k + 1 =$ number of knot points

where $\{B_{k\ell}(\cdot;\nu_k) : \ell = 1, \ldots, u_k + \nu_k = m_k\}$ is the ν_k-th degree B-spline basis with $u_k + 1$ equidistant knots for the k-th component

normalized B-splines: $\displaystyle\sum_{\ell=1}^{m_k} B_{k\ell}(t;\nu_k) = 1$

$$\boxed{\beta_k(t_{ij}) = \sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij};\nu_k)}$$

$\alpha_{k\ell}$ unknown coefficients

the **B-spline estimates** of the coefficients: minimize

$$S(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left(Y_{ij} - \sum_{k=0}^{p} \underbrace{\sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij};\nu_k)}_{=\beta_k(t_{ij})} X_{ij}^{(k)} \right)^2$$

with respect to $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0^\mathsf{T}, \ldots, \boldsymbol{\alpha}_p^\mathsf{T})^\mathsf{T}$, where $\boldsymbol{\alpha}_k = (\alpha_{k1}, \ldots, \alpha_{km_k})^\mathsf{T}$

what is the solution to this minimization problem ?

it is better to write all this in matrix notation

$$S(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left(Y_{ij} - \sum_{k=0}^{p} \sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij};\nu_k) X_{ij}^{(k)} \right)^2$$
$$= \sum_{i=1}^{n} (\mathbf{Y}_i - \mathbf{U}_i \boldsymbol{\alpha})^T \mathbf{W}_i (\mathbf{Y}_i - \mathbf{U}_i \boldsymbol{\alpha})$$

$$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})^{\mathsf{T}}$$

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

$$\mathbf{U}_{ij}^{\mathsf{T}} = \mathbf{X}_{ij}^{\mathsf{T}} \mathbf{B}(t_{ij}) \in I\!R^{1 \times m_{\text{tot}}} \qquad \mathbf{X}_{ij} = \left(1, X^{(1)}(t_{ij}), \ldots, X^{(p)}(t_{ij})\right)^{\mathsf{T}}$$

$$\mathbf{U}_i = (\mathbf{U}_{i1}^{\mathsf{T}}, \ldots, \mathbf{U}_{iN_i}^{\mathsf{T}})^{\mathsf{T}} \in I\!R^{N_i \times m_{\text{tot}}} \quad \text{and} \quad m_{\text{tot}} = \sum_{k=0}^{p} m_k$$

$$\mathbf{W}_i = \text{diag}\left(N_i^{-1}, \ldots, N_i^{-1}\right) \in I\!R^{N_i \times N_i}$$

(a diagonal matrix with N_i times N_i^{-1} on the diagonal)

$$S(\boldsymbol{\alpha}) = \sum_{i=1}^{n} (\mathbf{Y}_i - \mathbf{U}_i\boldsymbol{\alpha})^T \mathbf{W}_i (\mathbf{Y}_i - \mathbf{U}_i\boldsymbol{\alpha})$$

if $\sum_{i=1}^{n} \mathbf{U}_i^T \mathbf{W}_i \mathbf{U}_i$ is invertible then $S(\boldsymbol{\alpha})$ has a unique minimizer

$$\boxed{\widehat{\boldsymbol{\alpha}} = \Big(\sum_{i=1}^{n} \mathbf{U}_i^T \mathbf{W}_i \mathbf{U}_i\Big)^{-1} \sum_{i=1}^{n} \mathbf{U}_i^T \mathbf{W}_i \mathbf{Y}_i}$$

where $\widehat{\boldsymbol{\alpha}} = (\widehat{\boldsymbol{\alpha}}_0^\mathsf{T}, \ldots, \widehat{\boldsymbol{\alpha}}_p^\mathsf{T})^\mathsf{T}$ and $\widehat{\boldsymbol{\alpha}}_k = (\widehat{\alpha}_{k1}, \ldots, \widehat{\alpha}_{km_k})^\mathsf{T}$ for $k = 0, \ldots, p$

the **B-spline estimate** of $\boldsymbol{\beta}(t)$ is then

$$\widehat{\boldsymbol{\beta}}(t) = \mathbf{B}(t)\widehat{\boldsymbol{\alpha}} = (\widehat{\beta}_0(t), \ldots, \widehat{\beta}_p(t))^\mathsf{T} \quad \text{with} \quad \widehat{\beta}_k(t) = \sum_{\ell=1}^{m_k} \widehat{\alpha}_{k\ell} B_{k\ell}(t; \nu_k)$$

what about the asymptotic behaviour of this estimator?

notations:

$u^{\max} = \max_{0 \le k \le p} u_k$ maximal number of knot points

we allow $u^{\max}$ to grow with the sample size n, and denote it $u_n^{\max}$

$\rho_n = \inf_{\mathbf{g}^* \in \mathcal{G}} \|\boldsymbol{\beta} - \mathbf{g}^*\|_\infty$ assume: $\rho_n \to 0$ as $n \to \infty$

$$\|\boldsymbol{\beta} - \mathbf{g}^*\|_\infty = \max_{0 \le k \le p} \|\beta_k - g_k^*\|_\infty = \max_{0 \le k \le p} \left(\sup_t |\beta_k(t) - g^*(t)|\right)$$

where $\mathbf{g}^* = (g_0^*, \ldots, g_p^*)^\mathsf{T} \in \mathcal{G}$

$\mathcal{G} = \mathcal{G}_{\nu_0}(\mathcal{K}_0) \times \ldots \times \mathcal{G}_{\nu_p}(\mathcal{K}_p)$ $\mathcal{K}_k$ are sets of knots in $[0,1]$ for $k = 0, \ldots, p$

$\mathcal{G}_\nu(\mathcal{K})$ = space of spline functions of degree ν with set of knots $\mathcal{K}$

$B^r([0,1])$ = set of real-valued functions on $[0,1]$, who have a bounded r-th derivative

e.g. $r = 2$, $\nu = 3$, $\rho_n = O\left((u_n^{\max})^{-2}\right)$

theoretical results

- uniform consistency of $\widehat{\beta}_k(\cdot)$, estimator of $\beta_k(\cdot)$; + rate
- uniform consistency of $\widehat{\beta}_k^{(v)}(\cdot)$, estimator of $\beta_k^{(v)}(\cdot)$; + rate

$v = 0, \ldots, \nu_k$

Corollary.

Suppose $\beta_k(\cdot) \in B^{\nu_k+1}([0,1])$ for $k = 0, \ldots, p$. Then, under Assumptions 1—5,

$$\|\widehat{\boldsymbol{\beta}}^{(v)} - \boldsymbol{\beta}^{(v)}\|_\infty = O_P\left((u_n^{\mathsf{max}})^v \rho_n + (u_n^{\mathsf{max}})^{v-\nu_n^{\mathsf{min}}-1} + (u_n^{\mathsf{max}})^v r_n\right)$$

for $v = 0, \ldots, \nu_n^{\mathsf{min}}$, where $\nu_n^{\mathsf{min}} = \min_{0 \le k \le p} \nu_k$

$$r_n^2 = \frac{(u_n^{\mathsf{max}})^2}{n^2} \sum_{i=1}^{n} \left(\frac{1}{N_i} \left(1 - \frac{1}{u_n^{\mathsf{max}}}\right) + \frac{1}{u_n^{\mathsf{max}}} \right)$$

Assumption 1:

1. The observation times $t_{ij},\ j = 1, \ldots, N_i,\ i = 1, \ldots, n$, are chosen independently according to a distribution function $F_T(t)$ on $[0, 1]$. Moreover, they are independent of the response and the covariate process $\{(Y_i(t), X_i^{(1)}(t), \ldots, X_i^{(p)}(t))\},\ i = 1, \ldots, n$.
   The distribution function $F_T(t)$ has a Lebesgue density $f_T(t)$ that is bounded away from zero and infinity, uniformly over all $t \in [0, 1]$, that is, $\exists$ positive constants M_1 and M_2 such that $M_1 \leqslant f_T(t) \leqslant M_2$ for all $t \in [0, 1]$.
2. The eigenvalues $\eta_0(t), \ldots, \eta_p(t)$ of $\mathbf{\Sigma}(t) = E(\mathbf{X}(t)\mathbf{X}(t)^\mathsf{T})$ are bounded away from zero and infinity, uniformly over all $t \in [0, 1]$, that is, $\exists$ positive constants M_3 and M_4 such that $M_3 \leqslant \eta_0(t) \leqslant \ldots \leqslant \eta_p(t) \leqslant M_4$ for all $t \in [0, 1]$.
3. $\exists$ a positive constant M_5 such that $|X^{(k)}(t)| \leqslant M_5$ for all $t \in [0, 1]$ and $k = 0, \ldots, p$.
4. $\exists$ a positive constant M_6 such that $E(\varepsilon^2(t)) \leqslant M_6 < \infty$ for all $t \in [0, 1]$.
5. $\limsup_{n\to\infty}\left(\frac{\max_k m_k}{\min_k m_k}\right) < \infty$.

example

- data from the National Institute of Mental Health Schizophrenia Collaborative Study
- response variable: 'severity of the illness', measured on a numerical scale from 1 (normal, not ill) to 7 (among most extremely ill)
- most patients were measured at weeks 0, 1, 3 and 6
  a few patients were additionally measured at weeks 2, 4 and 5

  hence, N_i is between 4 and 7
- $n = 437$ patients were randomly assigned to either receive a drug or a placebo
- Drug=binary variable : Drug=1, patient received the drug
  Drug=0, patient received a placebo
- consider a varying coefficient model:

$$Y(\text{week}) = \beta_0(\text{week}) + \beta_1(\text{week})\,\text{Drug} + \varepsilon(\text{week})$$

- number of knots are determined by a 4-fold cross validation (with the number of knots ranging from 1 to 8)

mean fits ($\widehat{E}(Y(t) \mid \mathbf{X}(t), t)$) for the placebo group and the drug group

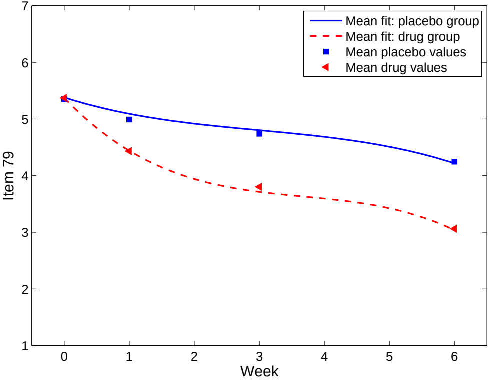
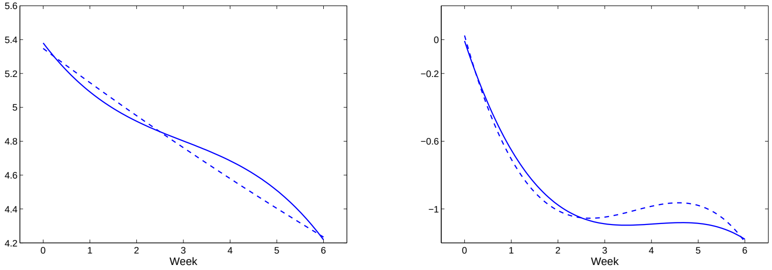


Figure: Schizophrenia data. The mean fits for the placebo and the drug group. The squares and triangles are the mean response measurements at weeks 0, 1, 3 and 6, of the placebo group and drug group, respectively.

- **how does the drug affects the severity of the illness of patients?**
- **how does a possible effect evolve over time?**



(a) (b)

Figure: Schizophrenia data. Estimates of coefficient functions $\beta_0(\cdot)$ and $\beta_1(\cdot)$, using cubic splines (full lines) and splines with degree vector $(3, 2)$ (dashed lines).

- negative $\beta_1(\cdot)$ which is decreasing: drug is effective
- the drug effect drops quickly to reach a steady effect of -1 from week 3 onwards

main goal: testing for various shape constraints on the coefficient functions in a varying coefficient model

- testing

$$\begin{aligned} H_0 : &\quad \beta_k(\cdot) \text{ is monotone increasing} \\ \text{versus } H_1 : &\quad \beta_k(\cdot) \text{ is not monotone increasing} \end{aligned}$$

- testing

$$\begin{aligned} H_0 : &\quad \beta_k(\cdot) \text{ is a convex function} \\ \text{versus } H_1 : &\quad \beta_k(\cdot) \text{ is not a convex function} \end{aligned}$$

- simultaneous testing (e.g.)

$$\begin{aligned} H_0 : &\quad \beta_1(\cdot) \text{ is monotone decreasing} \quad \text{and} \quad \beta_3(\cdot) \text{ is convex} \\ \text{versus } H_1 : &\quad \neg H_0 \end{aligned}$$

- constrained spline estimation: **monotonicity**

which constraints need to be added on the B-spline coefficients to obtain a monotone B-spline estimate ?

- spline function $g(t) = \sum_{\ell=1}^{m} \gamma_\ell B_\ell(t;\nu)$ with distance $1/u$ between equidistant knot points

  the derivative of g is:

$$g'(t) = \sum_{\ell=1}^{m} \gamma_\ell B_\ell^{'}(t;\nu) = u \sum_{\ell=1}^{m-1} \Delta\gamma_{\ell+1} B_\ell(t;\nu-1) \quad \Delta\gamma_{\ell+1} = \gamma_{\ell+1} - \gamma_\ell$$

- in general: **if** $\Delta\gamma_{\ell+1} \geq 0 \quad \forall \ell,$ **then** $g(\cdot)$ is monotone increasing
- **Lemma**

  If $\nu = 2$, then $g'(t) \geq 0$ for all $t \in [0,1]$ **if and only if** $g'(\xi_i) \geq 0$ for $i = 0, 1, \ldots, u$

  hence, monotonicity of $g(t)$ in the knots $\xi_0, \cdots, \xi_u$ is equivalent to monotonicity on the whole domain $[\xi_0, \xi_u]$

- for **quadratic splines**:
  - for $g(t) = \sum_{\ell=1}^{m} \gamma_\ell B_\ell(t;2)$
  - denote the matrix $\mathbf{S} \in I\!R^{(u+1)\times(u+2)}$ which consists of B-spline derivatives *at the knots*; $\mathbf{S}_{ij} = B'_j(\xi_{i-1};2)$
    due to the lemma:

    g is increasing *if and only if*

    $$\mathbf{S}\boldsymbol{\gamma} \geq \mathbf{0} \in I\!R^{u+1} \qquad \text{where } \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^{\mathsf{T}}$$

    Wang & Meyer (2011), Meyer (2012)

- for **cubic splines** ($\nu = 3$): for imposing monotonicity we need to impose quadratic constraints at the knots

  Akhim, G. & Verhasselt (2017)

- or, for cubic or higher order splines: impose general constraint (e.g. via penalty term)

  see e.g. Bollaerts *et al.* (2006), Akhim, G. & Verhasselt (2017)

- testing for **monotonicity**:

H_0 : $\beta_k(\cdot)$ is monotone increasing versus H_1 : $\beta_k(\cdot)$ is not monotone increasing

or equivalently

$$H_0: \quad \beta_k'(t) \geq 0 \quad \forall t \in [0,1] \qquad \text{versus} \qquad H_1: \quad \neg H_0$$

(for testing whether $\beta_k(\cdot)$ is monotone decreasing, replace $X^{(k)}$ by $-X^{(k)}$)

- • using **quadratic spline approximation**
  - translate monotonicity constraint into linear constraint on B-spline coefficients: define $\mathbf{C} = (\mathbf{0}_1, \mathbf{S}, \mathbf{0}_3)$ where

    $\mathbf{0}_1 \in I\!R^{(u_k+1)\times\sum_{j=0}^{k-1} m_j}$ and
    $\mathbf{0}_3 \in I\!R^{(u_k+1)\times\sum_{j=k+1}^{d} m_j}$ are matrices with entries 0

    $\mathbf{S} \in I\!R^{(u_k+1)\times(u_k+2)}$ = matrix of derivatives at the knots of B-splines corresponding to coefficient $\beta_k(\cdot)$: $\mathbf{S}_{ij} = B_{kj}'(\xi_{k,i-1}; 2)$
  - the estimate $\widehat{\beta}_k$ is increasing **if and only if** $\mathbf{C}\widehat{\alpha} \geq 0$

based on this: what would be an appropriate test statistic ?

possible test statistic (Wang & Meyer (2011)) : $\boxed{\min(\mathbf{C}\widehat{\boldsymbol{\alpha}})}$

pseudo algorithm to test the hypothesis H_0 is:

1. determine the unconstrained estimator $\widehat{\boldsymbol{\alpha}}$, and calculate the minimum of the slopes at the knots

$$s_{\min} = \min(\mathbf{C}\widehat{\boldsymbol{\alpha}})$$

2. if $s_{\min}$ is non-negative, do not reject H_0
3. if $s_{\min} < 0$, determine the distribution of $s_{\min}$ under H_0 and calculate the α percentile Q_α
4. if $s_{\min}$ is smaller than the α percentile, then reject H_0

how to access the distribution of $s_{\min}$ under H_0?

two approaches: bootstrap procedure
OR relying on asymptotic normality result

♣ first approach: **bootstrap procedure**

- calculate residuals

$$\widehat{\varepsilon}_{ij} = Y_{ij} - \sum_{k=0}^{p} X_{ij}^{(k)} \widehat{\beta}_k(t_{ij}) \qquad \widehat{\boldsymbol{\beta}}(\cdot) \text{ unconstrained B-spline estimator}$$

- obtain **pseudo responses under H_0**

$$Y_{ij}^{\mathsf{ps}} = \sum_{k=0}^{p} X_{ij}^{(k)} \widehat{\beta}_k^{\mathsf{cs}}(t_{ij}) + \widehat{\varepsilon}_{ij} \quad \text{for } i = 1, \ldots, n \quad \text{and } j = 1, \ldots, N_i$$

where

$$\widehat{\boldsymbol{\beta}}^{\mathsf{cs}} = (\widehat{\beta}_0^{\mathsf{cs}}, \ldots, \widehat{\beta}_p^{\mathsf{cs}})^{\mathsf{T}}$$

is the *constrained estimate* putting the constraint on β_k

bootstrap procedure to determine the distribution of $s_{\min}$ under H_0 is

- Step 1: resample n subjects (with all its repeated measurements) with replacement from

$$\{(Y_{ij}^{\textsf{ps}}, X_{ij}, t_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N_i\}$$

  to obtain the bootstrap sample
  $\{(Y_{ij}^{\textsf{ps}*}, X_{ij}^*, t_{ij}^*) : i = 1, \ldots, n, j = 1, \ldots, N_i^*\}$
- Step 2: repeat the above sampling procedure B times
- Step 3: obtain the test statistic $s_{\min}^*$ from each bootstrap sample and derive the empirical distribution based on all $s_{\min}^*$
- Step 4:
  consider the α percentile $\widehat{Q}_\alpha$ of the empirical distribution in Step 3;

  reject H_0 if $s_{\min} < \widehat{Q}_\alpha$;

  else do not reject H_0

♣ second approach: via **asymptotic normality result**
(see Wang & Meyer (2011))

what about the variance-covariance matrix of the B-spline estimators ?

- the B-splines estimator

$$\widehat{\boldsymbol{\alpha}} = \Big(\sum_{i=1}^{n} \mathbf{U}_i^T \mathbf{W}_i \mathbf{U}_i\Big)^{-1} \sum_{i=1}^{n} \mathbf{U}_i^T \mathbf{W}_i \mathbf{Y}_i = (\mathbf{U}^\intercal \mathbf{W} \mathbf{U})^{-1} \mathbf{U}^\intercal \mathbf{W} \mathbf{Y}$$

  with additional notations

$$\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_n)^\intercal \in I\!R^{N \times m_{\text{tot}}} \quad \mathbf{W} = \text{diag}\Big(\mathbf{W}_1, \ldots, \mathbf{W}_n\Big) \in I\!R^{N \times N}$$

- observations under the model: $\mathbf{Y} \approx \mathbf{U}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$
- denote by $\mathbf{V}$ the variance-covariance matrix of $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n)^\intercal$ a matrix of dimension $N \times N$
- denote $\mathcal{X} = \{(t_{ij}, \mathbf{X}_{ij}) : i = 1, \ldots, n, j = 1, \ldots, N_i\}$
- conditioning on $\mathcal{X}$, one obtains: $E(\widehat{\boldsymbol{\alpha}} \mid \mathcal{X}) \approx \boldsymbol{\alpha}$ and $\text{Cov}(\widehat{\boldsymbol{\alpha}} \mid \mathcal{X}) \approx (\mathbf{U}^\intercal \mathbf{W} \mathbf{U})^{-1} \mathbf{U}^\intercal \mathbf{W} \mathbf{V} \mathbf{W} \mathbf{U} (\mathbf{U}^\intercal \mathbf{W} \mathbf{U})^{-1}$

what now further in case of normal errors?

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top \sim N(\mathbf{0}, \mathbf{V})$$

- recall that we need to evaluate
  $P\left(\min(\mathbf{C}\widehat{\boldsymbol{\alpha}}) \leq r\right) = P\left(s_{\min} \leq r\right), \; r \in \mathbb{R}$
- since $E(\mathbf{Y}|\mathcal{X}) \approx \mathbf{U}\boldsymbol{\alpha}$, we have that $\mathbf{C}\widehat{\boldsymbol{\alpha}}$ is, conditioned on $\mathcal{X}$, approximately normal with mean $\mathbf{C}\boldsymbol{\alpha}$ and variance-covariance matrix

$$\boldsymbol{\Sigma} = \mathbf{C}(\mathbf{U}^\top \mathbf{W}\mathbf{U})^{-1}\mathbf{U}^\top \mathbf{W}\mathbf{V}\mathbf{W}\mathbf{U}(\mathbf{U}^\top \mathbf{W}\mathbf{U})^{-1}\mathbf{C}^\top$$

- we obtain the expression

$$P\left(s_{\min} \leq r\right) = 1 - P\left(s_{\min} > r\right) = 1 - \int \cdots \int_{\{\mathbf{z}|\mathbf{z}-r\mathbf{1}\geq 0\}} \phi(\mathbf{z}; \mathbf{C}\boldsymbol{\alpha}, \boldsymbol{\Sigma}) d\mathbf{z}$$

  where $\mathbf{z}, \mathbf{1} = (1, 1, \ldots, 1)^\top \in \mathbb{R}^{(u_k+1)\times 1}$

  $\phi(\cdot\,; \mathbf{C}\boldsymbol{\alpha}, \boldsymbol{\Sigma})$=multivariate normal density with mean $\mathbf{C}\boldsymbol{\alpha}$ and covariance $\boldsymbol{\Sigma}$
- this probability can only be calculated if $\boldsymbol{\alpha}$ and $\mathbf{V}$ are known ...

consistency of the test, based on asymptotic normality result

probability of committing an error of Type II tends to 0, when $n \to \infty$

Theorem 2

Assume that $u_n^{\text{max}} \rho_n + (u_n^{\text{max}})^{\nu_k} + u_n^{\text{max}} r_n = o(1)$. Under Assumptions 1—5, if $\inf_{t \in [0,1]} \beta_k'(t) = \delta > 0$, then

$$\lim_{n \to \infty} P(s_{\text{min}} < \min(0, \widehat{Q}_\alpha)) = 0$$

- • using **cubic spline approximation**
  - test statistic: $\min_{t \in \text{Grid}} \widehat{\beta}_k'(t)$
  - first approach: bootstrap procedure similar as before
  - second approach: now rely on the asymptotic behaviour of the derivative estimates

- testing for **convexity** testing

$$H_0: \quad \beta_k(\cdot) \text{ is a convex function}$$
$$\text{versus } H_1: \quad \beta_k(\cdot) \text{ is not a convex function}$$

or equivalently

$$H_0: \beta_k''(t) \geq 0 \text{ for all } t \text{ in } [0,1] \qquad \text{versus} \qquad H_1: \neg H_0$$

- similar to before, but now focusing on the estimates of the second derivative function
- here distinction between
  - use of cubic spline approximation
  - use of quartic (or higher order) spline approximation

- **simultaneous testing** : example

$$H_0: \quad \beta_1(\cdot) \text{ is monotone decreasing} \quad \text{and} \quad \beta_3(\cdot) \text{ is convex}$$

$$\text{versus } H_1: \quad \neg H_0$$

- test statistic:

$$\mathbf{s} = \left(\min_{t \in \text{Grid}} \widehat{\beta}_1'(t), \min_{t \in \text{Grid}} \widehat{\beta}_3''(t) \right)$$

- use bootstrap type of procedure
- use Bonferroni type of correction

we looked at: conditional mean

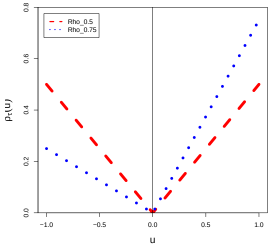
other quantities of interest: **conditional quantiles** (quantile regression)

what about the objective function $S(\alpha)$?

conditional mean	$\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left(Y_{ij} - \sum_{k=0}^{p} \sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij}; \nu_k) X_{ij}^{(k)} \right)^2$
conditional quantile (order τ)	$\sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \rho_\tau \left(Y_{ij} - \sum_{k=0}^{p} \sum_{\ell=1}^{m_k} \alpha_{k\ell} B_{k\ell}(t_{ij}; \nu_k) X_{ij}^{(k)} \right)$

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

check function



in both contexts: **homoscedasticity** $\Longleftrightarrow$ **heteroscedasticity**

general setting model

$$\begin{aligned} Y(T) &= \beta_0(T) + \beta_1(T)X^{(1)}(T) + \cdots + \beta_p(T)X^{(p)}(T) + \widetilde{\varepsilon} \\ &= \mathbf{X}^{\mathsf{T}}(T)\boldsymbol{\beta}(T) + V(\mathbf{X}(T), T)\,\varepsilon(T) \end{aligned}$$

where $\varepsilon(T)$ is independent of $(\mathbf{X}(T), T)$

special settings: | $V(\mathbf{X}(T), T) = V(T)$ simple heteroscedastic setting | $V(\mathbf{X}(T), T) = V$ a constant homoscedastic setting

assumptions to ensure identifiability needed in all settings

Andriyana (2015), Andriyana & G. (2017), Andriyana *et al.* (2017), ...

general heteroscedastic varying coefficient model

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

$$\mathbf{V}\left(\mathbf{X}(\mathbf{T}), \mathbf{T}\right) = \exp\left\{\gamma_{\mathbf{0}}(\mathbf{T}) + \gamma_{\mathbf{1}}(\mathbf{T})\mathbf{X}^{(\mathbf{1})}(\mathbf{T}) + \ldots + \gamma_{\mathbf{p}}(\mathbf{T})\mathbf{X}^{(\mathbf{p})}(\mathbf{T})\right\}$$

from the model and the error structure:

$$Y(T) = \underbrace{\mathbf{X}^{\mathsf{T}}(T)\boldsymbol{\beta}(T)}_{\text{signal part}} + \underbrace{\exp\left\{\mathbf{X}^{\mathsf{T}}(T)\boldsymbol{\gamma}(T)\right\}}_{\text{variability part}}\,\varepsilon(T)$$

where $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \ldots, \beta_p(t))^{\mathsf{T}}$

and $\boldsymbol{\gamma}(t) = (\gamma_0(t), \gamma_1(t), \ldots, \gamma_p(t))^{\mathsf{T}}$

aims: estimate all unknown coefficient functions (in the signal and the variability part!)

estimate all conditional quantiles

shape testing for the coefficient functions, the β_k's and the γ_ℓ's

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

what is the expression for the conditional quantile function?

denote the conditional quantile of order τ $(0 < \tau < 1)$ of $\varepsilon(T)$ given $(\mathbf{X}(T), T)$ by

$$a^{\tau}(T) = \inf\{y : P\{\varepsilon(T) \le y \mid (\mathbf{X}(T), T)\} \ge \tau\} = q_{\tau}(\varepsilon(T) \mid \mathbf{X}(T), T)$$

τ-th conditional quantile of $Y(T)$ given $(\mathbf{X}(T), T)$ is

$$\boxed{q_{\tau}(Y(T)|\mathbf{X}(T), T) = \mathbf{X}^{\mathsf{T}}(T)\boldsymbol{\beta}(T) + V(\mathbf{X}(T), T)\,a^{\tau}(T)}$$

- **estimation methods**
  - for identifiability reasons, and for estimating the variability function: adapt approach of He (1997)
  - basic assumptions:
    - **(H1)**: the conditional median quantile of the error term equals zero: $q_{0.5}\left\{\varepsilon(T) \mid \mathbf{X}(T), T\right\} = 0$
    - **(H2)**: $q_{0.5}\Big\{\ln|\varepsilon(T)| \mid \mathbf{X}(T), T\Big\} = 0$
  - the estimation consists of three steps:
    1. estimate the conditional median function
    2. estimate the variability function $V(\mathbf{X}(T), T)$
    3. estimate the conditional quantile function

various testing problems:

- testing for constancy
- testing for monotonicity
- testing for convexity/concavity
- shape testing for both signal and variability part

tests involving some or all coefficient functions in

the signal part: $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \ldots, \beta_p(t))^\mathsf{T}$

the variability part: $\boldsymbol{\gamma}(t) = (\gamma_0(t), \gamma_1(t), \ldots, \gamma_p(t))^\mathsf{T}$

- ◇ likelihood ratio type of tests
- ◇ other tests: based on looking at differences of B-spline coefficients

References

* Bollaerts, K., Eilers, P.H.C. and Aerts, M. (2006). Quantile regression with monotonicity restrictions using P-splines and the L_1-norm *Statistical Modelling*, **6**, 189—207.
* Andriyana, Y., G., I. and Verhasselt, A. (2017). Quantile regression in varying coefficient models: non-crossingness and heteroscedasticity. *Statistical Papers*, to appear.
* Ahkim, M., G., I. and Verhasselt, A. (2017). Shape testing in varying coefficient models. *Test*, **26**, 429–450.
* G., I., Ibrahim, M. and Verhasselt, A. (2017). Shape testing in quantile varying coefficient models with heteroscedastic error. *Journal of Nonparametric Statistics*, **29**, 391–406.
* He, X. (1997). Quantile curves without crossing *The American Statistician*, **51**, 186—192.
* Kim, M.O. (2007). Quantile regression with varying coefficients. *Annals of Statistics*, **35**, 92—108.
* Meyer, M.C. (2012). Constrained penalized splines. *The Canadian Journal of Statistics*, **40**, 190–206.
* Wang, J.C. and Meyer, M.C. (2011). Testing monotonicity or convexity of a function using regression splines. *The Canadian Journal of Statistics*, **39**, 89–107.