Robust B-splines estimators in generalized partly linear regression under monotone constraints

Graciela Boente[1]

[1]Universidad de Buenos Aires and CONICET, Argentina

Based on joint work with

- Daniela Rodriguez[1] and Pablo Vena[1]

Semiparametric generalized partially linear model

- $y_i|(\mathbf{x}_i, z_i) \sim F(., \mu_i)$ canonical exponential family, $z_i \in [0,1]$

$$\exp\left\{\left[y\theta(\mathbf{x}, z) - B\left(\theta(\mathbf{x}, z)\right)\right]/A(\kappa_0) + C(y, \kappa_0)\right\} ,$$

- $\text{VAR}(y_i|(\mathbf{x}_i, z_i)) = A^2(\kappa_0)V(\mu_i)$ with $V : \mathbb{R} \to \mathbb{R}$ known function.

- $\mu_i = \mathbb{E}(y_i|(\mathbf{x}_i, z_i)) = \mu\left(\mathbf{x}_i, z_i\right)$

$$\mu\left(\mathbf{x}, z\right) = H\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(z)\right)$$

  - $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ is an unknown parameter.
  - $\eta_0 : [0,1] \to \mathbb{R}$ is a continuous function.
  - κ_0: nuisance parameter

Semiparametric generalized partially linear model

GPLM

$$\mu(\mathbf{x}, z) = H\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(z)\right)$$

- Partial linear logistic Model
- Partial linear Poisson Model

Semiparametric generalized partially linear model

GPLM

$$\mu(\mathbf{x}, z) = H\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(z)\right)$$

Partial linear Model

- Partial linear logistic Model
- Partial linear Poisson Model

Symmetric errors

Semiparametric generalized partially linear model

GPLM

$$\mu(\mathbf{x}, z) = H\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(z)\right)$$

Partial linear Model

- Partial linear logistic Model
- Partial linear Poisson Model

Symmetric errors

Skewed errors

Log–Gamma Model

Isotonic generalized partially linear model

- We add a **monotone constraint** on the nonparametric component:

We assume that η_0 **is non–decreasing**.

Adding monotonicity to the GPLM

In many applications, monotonicity is a desired property.

- When $\boldsymbol{\beta} = \mathbf{0}$, Ramsay (1988) studied the relation between the incidence of Down's syndrome and the mother's age.
- Leitenstorfer and Tutz (2006) studied the air pollution (São Paulo) to evaluate the association between the number of daily deaths of elderly people for respiratory causes and the concentration of SO_2, CO, PM_{10} and O_3.
- Lu (2014) studied air pollution(Mexico City). The response y was daily death count, the covariates are
  - $z = PM_{10} =$ the daily mean ambient concentration of fine particle air pollutants $< 10\mu m$
  - $\mathbf{x} =$ the daily mean temperature and daily rainfall indicator.

Semi-parametric estimation

When $H(t) = t$

- Huang (2002): LS under constrains.
- Lu (2010): ML estimators based on $B-$splines.
- Wang and Huang (2002): Robust isotonic estimators ($\boldsymbol{\beta} = \mathbf{0}$).
- Álvarez and Yohai (2012): $M-$isotonic regression estimators ($\boldsymbol{\beta} = \mathbf{0}$).
- Du *et al.* (2013): $M-$estimators based on monotone $B-$splines with known scale.

Semi-parametric estimation

When $H(t) = t$

- Huang (2002): LS under constrains.
- Lu (2010): ML estimators based on $B-$splines.
- Wang and Huang (2002): Robust isotonic estimators ($\beta = \mathbf{0}$).
- Álvarez and Yohai (2012): $M-$isotonic regression estimators ($\beta = \mathbf{0}$).
- Du *et al.* (2013): $M-$estimators based on monotone $B-$splines with known scale.

Under a GPLM

- Boente *et al.* (2006): Robust profile kernel based estimators of η and β (*no restrictions on η*)
- Boente and Rodriguez (2010): Robust two–step kernel based estimators of η and β (*no restrictions on η*)
- Lu (2014): Monotone $B-$splines estimators based on the quasi–likelihood.

Spline approaches

$B-$spline approximation

Spline approaches

$B-$spline approximation

- **Monotone Splines**

Spline approaches

$B-$spline approximation

- **Monotone Splines**

- Monotone modification of unconstrained estimators

  Dette, Neumeyer & Pilz(2006) and Neumeyer (2007)

Splines and monotonicity

Consider the knots $\mathcal{Z}_n = \{\xi_i\}_{i=1}^{m_n+2\ell}$ where

$$0 = \xi_1 = \cdots = \xi_\ell < \xi_{\ell+1} < \cdots < \xi_{m_n+\ell+1} = \cdots = \xi_{m_n+2\ell} = 1$$

and denote as $\mathcal{S}_n(\mathcal{Z}_n, \ell)$ the class of splines of order $\ell > 1$ with knots $\mathcal{Z}_n$.

Schumaker (1981)

- There exist a class of $B-$spline basis functions $\{B_j : 1 \leq j \leq k_n\}$, with $k_n = m_n + \ell$, such that $g = \sum_{j=1}^{k_n} a_j B_j$, for any $g \in \mathcal{S}_n(\mathcal{Z}_n, \ell)$.
- The spline g is nondecreasing on $[0, 1]$ if $a_1 \leq \cdots \leq a_{k_n}$.

Robust Estimators

To obtain Robust estimators, combine monotone $B-$splines

Loss function that bounds residuals

Weight function to control the effect of leverage points

$\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$: loss function

$\mathbf{w} : \mathbb{R}^{\mathbf{p}} \rightarrow \mathbb{R}$: weight function to control leverage of $\mathbf{x}$

Robust estimators

- $\widehat{\kappa}$: robust consistent estimator of the nuisance parameter κ_0.

The estimators

$$(\widehat{\boldsymbol{\beta}}, \widehat{\eta}) = (\widehat{\boldsymbol{\beta}}, \sum_{j=1}^{k_n} \widehat{a}_j B_j)$$

where

$$(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{a}}) = \underset{\mathbf{b}\in\mathbb{R}^p, \mathbf{a}\in\mathcal{L}_{k_n}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \phi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b} + \sum_{j=1}^{k_n} a_j B_j(z_i), \widehat{\kappa} \right) w(\mathbf{x}_i),$$

$$\mathcal{L}_{k_n} = \left\{ \mathbf{a} \in \mathbb{R}^{k_n} : a_1 \leq \cdots \leq a_{k_n} \right\}.$$

Loss functions: Bounding the deviances

$$\phi(y, u, \kappa) = \rho_c[d(y; u)] + G(H(u)) \,, \quad c = c(\kappa)$$

- ρ_c odd and bounded nondecreasing function with continuous derivative φ_c.
- c is a tuning parameter.
- G guarantees Fisher–consistency.

$$G'(s) = \int \psi_c[d(y; u)]\, f'(y, s) d\mu(y) = \mathbb{E}_s \left(\psi_c[d(y; u)]\, \frac{f'(y, s)}{f(y, s)} \right) \,,$$

- $\mathbb{E}_s$ expectation taken under $F(\cdot, s)$ and $f'(y, s) = \frac{\partial}{\partial s} f(y, s)$.

Loss functions: Bounding the deviances

$$\phi(y, u, \kappa) = \rho_c[d(y; u)] + G(H(u)) , \quad c = c(\kappa)$$

- ρ_c odd and bounded nondecreasing function with continuous derivative φ_c.
- c is a tuning parameter.
- G guarantees Fisher–consistency.

$$G'(s) = \int \psi_c[d(y; u)]\, f'(y, s) d\mu(y) = \mathbb{E}_s \left(\psi_c[d(y; u)]\, \frac{f'(y, s)}{f(y, s)} \right) ,$$

- $\mathbb{E}_s$ expectation taken under $F(\cdot, s)$ and $f'(y, s) = \frac{\partial}{\partial s} f(y, s)$.

When $y_i|(\mathbf{x}_i, z_i)$ **has a density,** $G(s) \equiv 0$ (Bianco *et al.*, 2005).

The partial linear model: **Symmetric errors**

$$y=\beta_0^{\mathrm{T}}\mathbf{x}+\eta_0(z)+\ u\,,\qquad u\sim G_0(\cdot/\sigma_0)$$

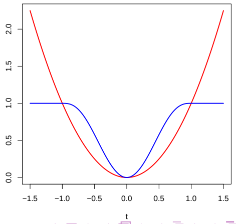
- κ_0 is scale parameter σ_0 and

$$\boldsymbol{\phi(\mathbf{y},\mathbf{s},\kappa)=\rho_{\mathbf{c}}\left(\frac{\mathbf{y}-\mathbf{s}}{\kappa}\right)}\,,$$

  - $\rho_c(t)=\rho(t/c)$ and $\rho:\mathbb{R}\to[0,\infty)$ is a $\rho-$function

ρ: bisquare function

$$\rho_{\mathrm{T},\,c}(t)=\min\left(1-(1-(t/c)^2)^3,1\right)$$



PLM: Symmetric errors

1. Compute an unrestricted $MM-$estimator $(\widehat{\beta}, \widehat{\eta}) = (\widehat{\beta}, \sum_{j=1}^{k_n} \widehat{a}_j B_j)$

$$(\widehat{\beta}, \widehat{\mathbf{a}}) = \underset{\mathbf{b}\in\mathbb{R}^p, \mathbf{a}\in\mathbb{R}^{k_n}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \rho_c \left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}} \mathbf{b} - \sum_{j=1}^{k_n} a_j\, B_j(z_i)}{\widehat{\sigma}} \right),$$

$\widehat{\sigma}$ is the scale related to an $S-$estimator (Yohai, 1987)

PLM: Symmetric errors

1. Compute an unrestricted $MM-$estimator $(\widehat{\beta}, \widehat{\eta}) = (\widehat{\beta}, \sum_{j=1}^{k_n} \widehat{a}_j B_j)$

$$(\widehat{\beta}, \widehat{\mathbf{a}}) \quad = \quad \underset{\mathbf{b}\in\mathbb{R}^p, \mathbf{a}\in\mathbb{R}^{k_n}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \rho_c\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b} - \sum_{j=1}^{k_n} a_j\, B_j(z_i)}{\widehat{\sigma}}\right),$$

$\widehat{\sigma}$ is the scale related to an $S-$estimator (Yohai, 1987)

2. If $\widehat{a}_1^{(0)} \leq \widehat{a}_2^{(0)} \leq \cdots \leq \widehat{a}_{k_n}^{(0)}$, then
   - $\widehat{\beta} = \widehat{\beta}^{(0)}$
   - $\widehat{\eta}(z) = \sum_{j=1}^{k_n} \widehat{a}_j^{(0)} B_j(z)$.

PLM: Symmetric errors

1. Compute an unrestricted $MM-$estimator $(\widehat{\beta}, \widehat{\eta}) = (\widehat{\beta}, \sum_{j=1}^{k_n} \widehat{a}_j B_j)$

$$(\widehat{\beta}, \widehat{\mathbf{a}}) = \underset{\mathbf{b}\in\mathbb{R}^p, \mathbf{a}\in\mathbb{R}^{k_n}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \rho_c\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b} - \sum_{j=1}^{k_n} a_j B_j(z_i)}{\widehat{\sigma}}\right),$$

$\widehat{\sigma}$ is the scale related to an $S-$estimator (Yohai, 1987)

2. If $\widehat{a}_1^{(0)} \leq \widehat{a}_2^{(0)} \leq \cdots \leq \widehat{a}_{k_n}^{(0)}$, then
   - $\widehat{\beta} = \widehat{\beta}^{(0)}$
   - $\widehat{\eta}(z) = \sum_{j=1}^{k_n} \widehat{a}_j^{(0)} B_j(z)$.

3. Otherwise, use an IRWLS that takes into account the given restrictions, that is, we approximate the minimization problem using IRWLS subject to $a_1 \leq \cdots \leq a_{k_n}$ using quadratic programming.

Details

PLM: **Errors with exponential unimodal density**

$$y = \boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{x} + \eta_0(z) + u \,,$$

- Errors density

$$\mathbf{g_0(u, \alpha_0) = Q(\alpha_0) \exp^{\alpha_0 \, \nu(u)}} \,,$$

  - $\alpha_0 > 0$ an unknown parameter
  - ν is a continuous function with unique maximum at u_0
  - Log–Gamma case: $\nu(s) = s - \exp(s)$, $u_0 = 0$

PLM: Errors with exponential unimodal density

$$y = \beta_0^{\mathrm{T}} \mathbf{x} + \eta_0(z) + u \,,$$

Loss function: Bianco, García Ben & Yohai (2005)

$$\boldsymbol{\phi(y, s, \kappa) = \rho \left(\frac{\sqrt{d\,(y - s)}}{\kappa} \right)} \,,$$

- $d(s) = \nu(u_0) - \nu(s)$.
- ρ a $\rho-$function.
- κ: tuning constant related to the parameter α_0.

PLM: Errors with exponential unimodal density

- $MM-$estimator without restrictions

$$\left(\widehat{\beta}^{(0)}, \widehat{\mathbf{a}}^{(0)}\right) = \underset{(\mathbf{b},\mathbf{a})\in\mathbb{R}^{p+k_n}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{\sqrt{d\left(y_i - \left[\mathbf{x}_i^{\mathrm{T}}\mathbf{b} + \mathbf{a}^{\mathrm{T}}\mathbf{B}_i\right]\right)}}{\widehat{\kappa}_n}\right) w(\mathbf{x}_i)\,,$$

$\widehat{\kappa}_n$ is the tuning constant as in Bianco *et al.* (2005).

PLM: Errors with exponential unimodal density

- $MM-$estimator without restrictions

$$\left(\widehat{\boldsymbol{\beta}}^{(0)}, \widehat{\mathbf{a}}^{(0)}\right) = \underset{(\mathbf{b},\mathbf{a})\in\mathbb{R}^{p+k_n}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{\sqrt{d\left(y_i - \left[\mathbf{x}_i^{\mathrm{T}}\mathbf{b} + \mathbf{a}^{\mathrm{T}}\mathbf{B}_i\right]\right)}}{\widehat{\kappa}_n}\right) w(\mathbf{x}_i)\,,$$

$\widehat{\kappa}_n$ is the tuning constant as in Bianco *et al.* (2005).

- If $\widehat{a}_1^{(0)} \leq \widehat{a}_2^{(0)} \leq \cdots \leq \widehat{a}_{k_n}^{(0)}$, then
  - $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{(0)}$
  - $\widehat{\eta}(z) = \sum_{j=1}^{k_n} \widehat{a}_j^{(0)} B_j(z)$.

PLM: Errors with exponential unimodal density

- Otherwise, use a non–linear minimization algorithm with restrictions choosing as initial value $(\widehat{\beta}^{(0)}, \mathbf{a}^{(0)})$, where $\mathbf{a}^{(0)} \in \mathcal{L}_{k_n}$.
  One possible choice for $\mathbf{a}^0$ is $a_1^0 = a_2^0 = 0$ and $a_i^0 = i - 2$ for $i = 3, \ldots, k_n$.

Details

The increasing modification: Dette, Neumeyer & Pilz (2005), Neumeyer (2007)

- $f : [a, b] \to \mathbb{R}$ define

$$\Upsilon(f)(u) = \int_a^b \mathbb{I}_{\{f(z) \leq u\}} dz + a \qquad u \in \mathbb{R}$$

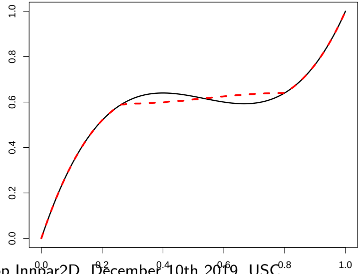
The increasing modification: Dette, Neumeyer & Pilz (2005), Neumeyer (2007)

- $f : [a,b] \to \mathbb{R}$ define

$$\Upsilon(f)(u) = \int_a^b \mathbb{I}_{\{f(z)\leq u\}} dz + a \qquad u \in \mathbb{R}$$

- Given $f : [0,1] \to \mathbb{R}$, the **I**ncreasing **mod**ification $f_{\mathrm{IMOD}} : [0,1] \to \mathbb{R}$ is

$$f_{\mathrm{IMOD}} = \Upsilon\left(\Upsilon(f)\mathbb{I}_{[f(0),f(1)]}\right)\mathbb{I}_{[0,1]}$$



f_{IMOD}

$f(x) =$
$5x^3 + 4x - 8x^2\mathbb{I}_{0\leq x\leq 1}$

The monotone estimator of η

A monotone estimator of $\eta : [0,1] \to \mathbb{R}$ may be constructed as

$$\widehat{\eta}_{\text{IMOD}} = \Upsilon\left(\Upsilon(\widehat{\eta})\mathbb{I}_{[\widehat{\eta}(0),\widehat{\eta}(1)]}\right)\mathbb{I}_{[0,1]}$$

from the unconstrained estimators.

Selection of k_n

As in He and Shi (1996) and He, Zhu & Fung (2002), define

$$BIC(k) = \log\left\{\frac{1}{n}\sum_{i=1}^{n}\rho\left(y_i, \mathbf{x}_i^{\mathrm{T}}\mathbf{b} + \sum_{j=1}^{k}\lambda_j\, B_j(z_i), \widehat{\kappa}\right) w(\mathbf{x}_i)\right\} + \frac{\log n}{2\,n}\, k\,.$$

A possible criterion is to search for the first (i.e. smallest k) local minimum of $BIC(k)$ in the range of

$$\max\left(\frac{n^{1/5}}{2}, 4\right) \le k \le 8 + 2\,n^{1/5}$$

when cubic splines are considered.

Assumptions

- $(y_i, \mathbf{x}_i, z_i)^{\mathrm{T}}$ are i.i.d. observations satisfying a GPLM model with η_0 non-decreasing
- $\eta_0 \in C^r[0,1]$ and $\eta_0^{(r)}$ is Lipschitz continuous
- The maximum spacing of the knots is of order $O(n^{-\nu})$, $0 \leq \nu \leq 1/2$
- $k_n = O(n^{\nu})$ for $1/(2r+2) < \nu < 1/(2r)$
- $\widehat{\kappa} \xrightarrow{a.s.} \kappa_0$

Asymptotic results

Let $\|\eta_0 - \widehat{\eta}\|^2_{L^2(Q)} = \mathbb{E}(\eta_0(t_1) - \widehat{\eta}(t_1))^2$.

- a) $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 + \|\widehat{\eta} - \eta_0\|^2_{L^2(Q)} \xrightarrow{a.s.} 0$.
- b) $\gamma_n \left(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 + \|\widehat{\eta} - \eta_0\|^2_{L^2(Q)}\right) = O_{\mathbb{P}}(1)$, where

$$\gamma_n = n^{\min(r\nu, \frac{1-\nu}{2})}$$

Hence, if $\nu = 1/(1+2r)$, the estimators converge at the optimal rate $n^{r/(1+2r)}$ and $\|\widehat{\eta} - \eta_0\|_\infty \xrightarrow{p} 0$.

Asymptotic results

Let $\|\eta_0 - \widehat{\eta}\|^2_{L^2(Q)} = \mathbb{E}(\eta_0(t_1) - \widehat{\eta}(t_1))^2$.

- a) $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 + \|\widehat{\eta} - \eta_0\|^2_{L^2(Q)} \xrightarrow{a.s.} 0$.
- b) $\gamma_n \left(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 + \|\widehat{\eta} - \eta_0\|^2_{L^2(Q)}\right) = O_{\mathbb{P}}(1)$, where

$$\gamma_n = n^{\min(r\nu, \frac{1-\nu}{2})}$$

Hence, if $\nu = 1/(1+2r)$, the estimators converge at the optimal rate $n^{r/(1+2r)}$ and $\|\widehat{\eta} - \eta_0\|_\infty \xrightarrow{p} 0$.

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N\left(0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0, \kappa_0)\right).$$

Monte Carlo study

- $NR = 1000$ replications,
- samples of size $n = 100$,

The uncontaminated sample, C_0, is generated as follows:

- (x_i, z_i) independent of each other, $x_i \sim \mathrm{N}(0,1)$, $z_i \sim \mathcal{U}(0,1)$.
- $y_i = \beta_0 x_i + \eta_0(z_i) + u_i$,

  $u_i \sim \log(\Gamma(3,1))$, $\beta_0 = 2$
- Two choices for the nonparametric component:

  Model 1 $\eta_{0,1}(t) = \sin(\pi t/2)$
  Model 2 $\eta_{0,2}(t) = \pi\, t + 0.25 \sin(4\pi t)$

Contaminations

We generate a sample $v_i \sim \mathcal{U}(0,1)$ for $1 \leq i \leq n$ and then:

- C_1 **introduces bad high leverage points in the carriers x, without changing the responses already generated**:

$$y_{i,c} = y_i \qquad x_{i,c} = \begin{cases} x_i & \text{if } v_i \leq 0.90 \\ x_i^\star & \text{if } v_i > 0.90\,, \end{cases}$$

  where $x_i^\star \sim N(5, 1/16)$.

Contaminations

We generate a sample $v_i \sim \mathcal{U}(0,1)$ for $1 \leq i \leq n$ and then:

- **C_1 introduces bad high leverage points in the carriers x, without changing the responses already generated**:

$$y_{i,c} = y_i \qquad x_{i,c} = \begin{cases} x_i & \text{if } v_i \leq 0.90 \\ x_i^\star & \text{if } v_i > 0.90\,, \end{cases}$$

  where $x_i^\star \sim N\,(5, 1/16)$.

- **C_2 introduces outlying observations in the responses generated according to the model but with an incorrect carrier x.**

$$y_{i,c} = \begin{cases} y_i & \text{if } v_i \leq 0.90 \\ y_i^\star & \text{if } v_i > 0.90\,, \end{cases} \qquad x_{i,c} = x_i$$

  where $y_i^\star = \beta_0 x_i^\star + \eta_0(z_i) + u_i^\star$ with

$$u_i^\star \sim \log(\Gamma(3,1)) \qquad x_i^\star \sim \mathrm{N}\,(5, 1/16)\,,$$

Contaminations

- C_3 **corresponds to increasing the variance of the carriers** x **and also to introduce large values on the responses**

$$x_{i,c} = \begin{cases} x_i & \text{if } v_i \leq 0.90 \\ \text{a new observation from a N}(0,25) & \text{if } v_i > 0.90, \end{cases}$$

$$y_{i,c} = \begin{cases} y_i & \text{if } v_i \leq 0.90 \\ y_i^\star & \text{if } v_i > 0.90\,, \end{cases}$$

with $y_i^\star = 3\,\log(10) + u_i^\star$ and $u_i^\star \sim \log(\Gamma(3,1))$.

Results under C_0

	Model 1						
		Summary measures for $\widehat{\beta}$					$\mathrm{MISE}(\widehat{\eta})$
	Estimator	Bias	SD	MSE	AS.SE	Cov.Prob	
(a)	CL	0.0002	0.0608	0.0037	0.0568	0.9340	0.0088
	ROB	0.0021	0.0672	0.0045	0.0620	0.9270	0.0096

Results under C_0

	Model 1						
		Summary measures for $\widehat{\beta}$					$\mathrm{MISE}(\widehat{\eta})$
	Estimator	Bias	SD	MSE	AS.SE	Cov.Prob	
(a)	CL	0.0002	0.0608	0.0037	0.0568	0.9340	0.0088
	ROB	0.0021	0.0672	0.0045	0.0620	0.9270	0.0096
(b)	CL	0.0009	0.0613	0.0038	0.0565	0.9280	0.0118
	ROB	-0.0000	0.0921	0.0085	0.0620	0.9060	0.0157

a) Monotone $B-$splines

b) Isotone Modification

$$\mathrm{ISE}(\widehat{\eta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\eta}(t_i) - \eta_0(t_i)\right)^2 .$$

Results under C_0

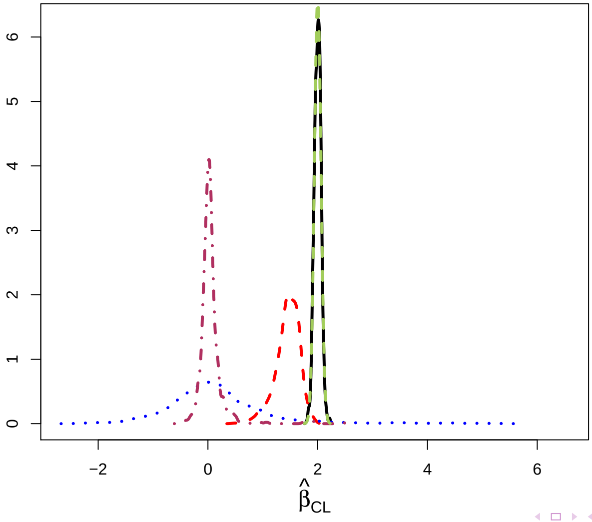
		Model 1					
		Summary measures for $\widehat{\beta}$					MISE($\widehat{\eta}$)
	Estimator	Bias	SD	MSE	AS.SE	Cov.Prob	
(a)	CL	0.0002	0.0608	0.0037	0.0568	0.9340	0.0088
	ROB	0.0021	0.0672	0.0045	0.0620	0.9270	0.0096
(b)	CL	0.0009	0.0613	0.0038	0.0565	0.9280	0.0118
	ROB	-0.0000	0.0921	0.0085	0.0620	0.9060	0.0157

a) Monotone $B-$splines
b) Isotone Modification

$$\text{ISE}(\widehat{\eta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\eta}(t_i) - \eta_0(t_i)\right)^2 .$$

We will only present the results obtained when η_0 is estimated using Monotone $B-$splines

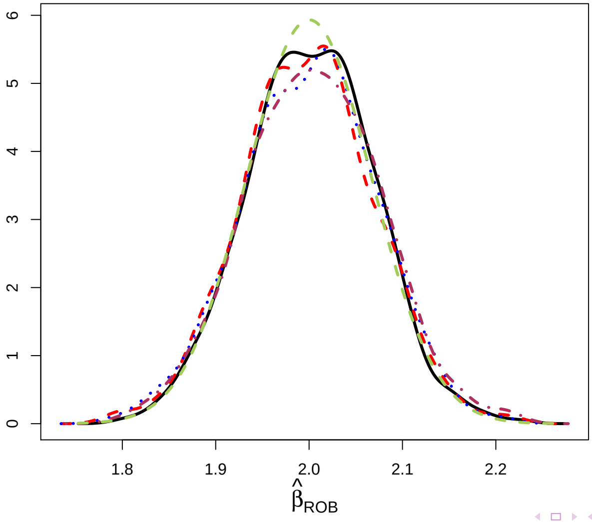
Density estimators of $\widehat{\beta}_{\mathrm{CL}}$, Model 1.



- ——: C_0
- − −: C_1
- . . . : C_2
- − · −: C_3

- − −: $N(0, \widehat{\sigma}^2)$

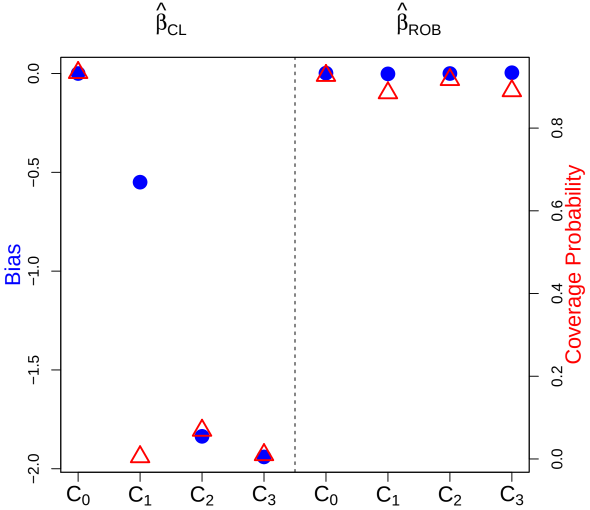
Density estimators of $\widehat{\beta}_{\mathrm{R}}$, Model 1.



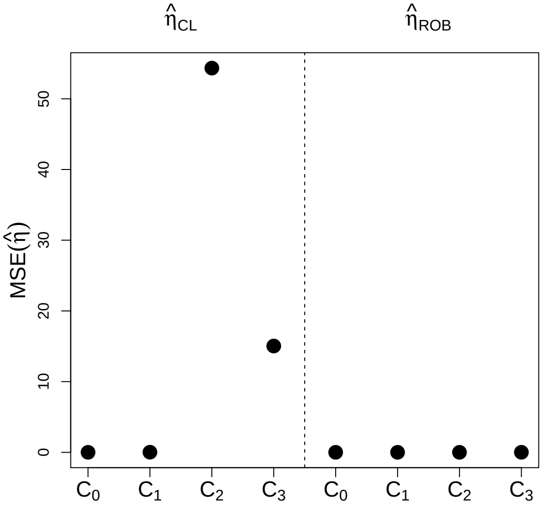
- ——: C_0
- – –: C_1
- . . . : C_2
- – · –: C_3

- – –: $N(0, \widehat{\sigma}^2)$

Performance of $\widehat{\beta}$, Model 1



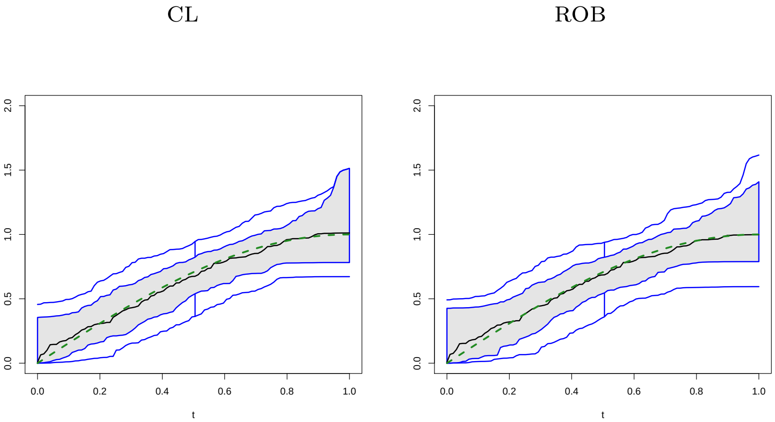
Performance of $\widehat{\eta}$, Model 1



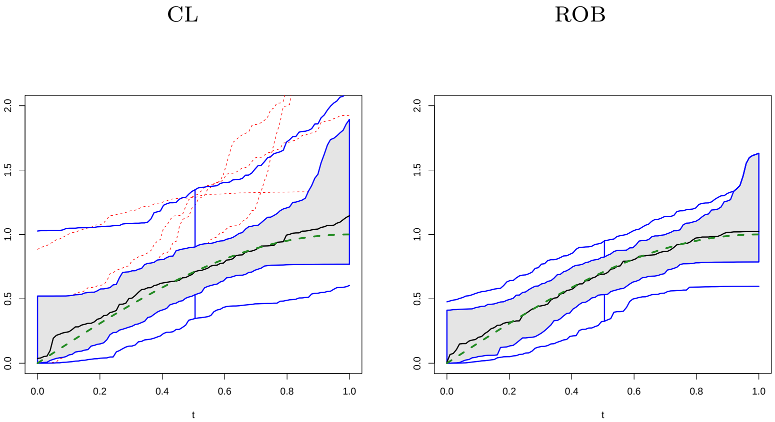
Final Conclusions

Tables

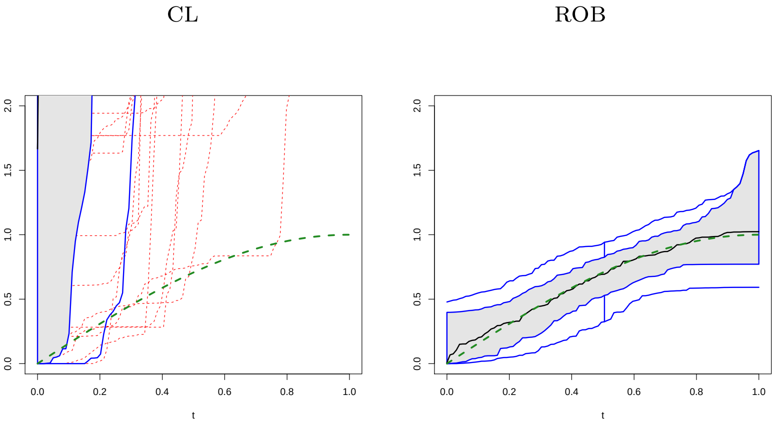
Performance of $\widehat{\eta}$: C_0



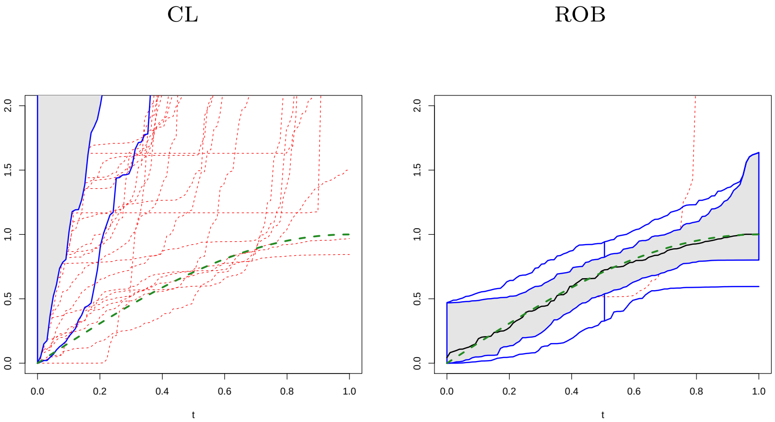
Performance of $\widehat{\eta}$: C_1



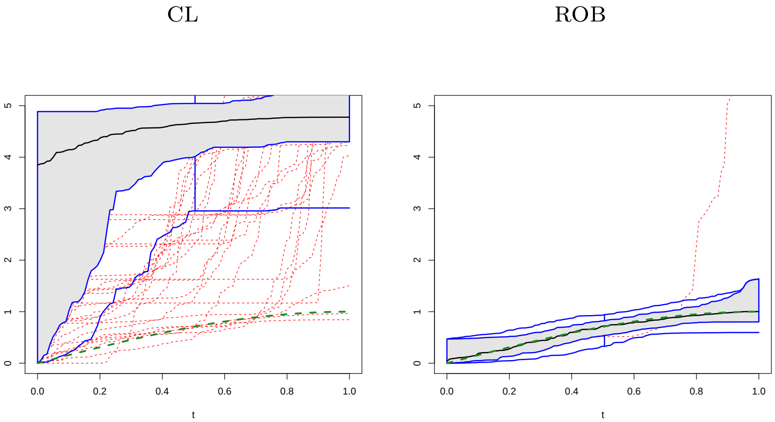
Performance of $\widehat{\eta}$: C_2



Performance of $\widehat{\eta}$: C_3



Performance of $\widehat{\eta}$: C_3



Hospital Costs Data (Marazzi and Yohai, 2004)

The data set corresponds to the costs of 100 patients hospitalized at the Centre Hospitalier Universitaire Vaudois in Lausanne (Switzerland) during 1999 for *medical back problems*.

Aim: Study the relationship between the **hospital cost of stay**, y, and the following **administrative explanatory variables**:

- LOS length of stay in days
- ADM admission type (0 = planned; 1 = emergency)
- INS insurance type (0 = regular; 1 = private)
- AGE years
- SEX (0 = female; 1 = male)
- DEST discharge destination (1 = home; 0 = another institution)

Linear fit approach

Cantoni and Ronchetti (2006) and Bianco *et al.* (2013) fitted a log–Gamma model to the data,

$$w_i|\mathbf{v}_i \sim \Gamma(\alpha, \mu_i) \qquad \log(\mu_i) = \log(\mathbb{E}(z_i|\mathbf{v}_i)) = \boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{v}_i$$

which is equivalent to a linear regression model with asymmetric errors

$$y_i = \log(w_i) = \gamma_0^{\mathrm{T}}\mathbf{v}_i + u_i,$$

- $u_i \sim \log\Gamma(\alpha, 1)$
- $\mathbf{v} = (ADM, INS, AGE, SEX, DEST, \log(LOS), 1)$

Using a robust QL approach Cantoni and Ronchetti (2006) identified 5 outliers ($i = 14, 21, 28, 44$ and 63), affecting the classical estimates of *INS* and the shape parameter.

Our setting

We will not impose a linear relation between $\log(y_i)$ and the $\log(LOS)$.

$$y_i = \beta_0^{\mathrm{T}} \mathbf{x}_i + \eta_0(z_i) + u_i$$

- $u_i \sim \log \Gamma(\alpha, 1)$,
- $\mathbf{x} = (ADM, INS, AGE, SEX, DEST)$, $\quad z = \log(LOS)$.
- $\eta_0 : \mathbb{R} \to \mathbb{R}$ is an increasing function.
- *BIC* criterion:
  - $\widehat{\beta}_{\mathrm{CL}}$ $k_n = 4$
  - $\widehat{\beta}_{\mathrm{R}}$ $k_n = 5$ $\quad c_\rho = 0.3515$

Hospital Costs Data

	ADM	*INS*	*AGE*	*SEX*	*DEST*	$\widehat{\alpha}$
$\widehat{\boldsymbol{\beta}}_{\mathrm{CL}}$	0.2148	**0.0984**	-0.0009	0.1088	-0.1358	**21.0809**
	(0.0497)	(0.0792)	(0.0013)	(0.0529)	(0.0723)	
$\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$	0.1979	-0.0207	-0.0019	0.0615	-0.1673	46.0088
	(0.0339)	(0.0537)	(0.0009)	(0.0358)	(0.0493)	

Hospital Costs Data

	ADM	*INS*	*AGE*	*SEX*	*DEST*	$\widehat{\alpha}$
$\widehat{\boldsymbol{\beta}}_{\mathrm{CL}}$	0.2148	**0.0984**	-0.0009	0.1088	-0.1358	**21.0809**
	(0.0497)	(0.0792)	(0.0013)	(0.0529)	(0.0723)	
$\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$	0.1979	-0.0207	-0.0019	0.0615	-0.1673	46.0088
	(0.0339)	(0.0537)	(0.0009)	(0.0358)	(0.0493)	
$\widehat{\boldsymbol{\beta}}_{\mathrm{CL}}^{-\{5\}}$	0.2172	-0.0324	-0.0016	0.0820	-0.1608	45.7560
	(0.0345)	(0.0575)	(0.0009)	(0.0354)	(0.0489)	

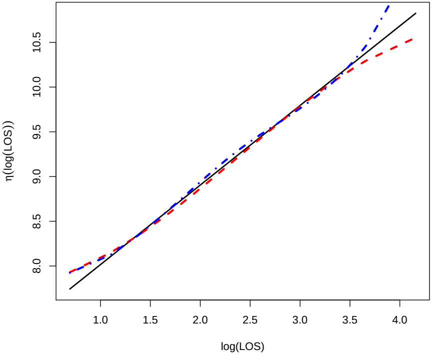
Analysis of Hospital Costs data, between brackets are reported the estimated asymptotic standard deviations of the estimators.

- As in the linear fit, the classical estimator of $\boldsymbol{\beta}$ are highly affected by the 5 outliers, which were also detected in our study.

- After removing these 5 data points, the classical estimators $\widehat{\boldsymbol{\beta}}_{\mathrm{CL}}^{-\{5\}}$ are very similar to those obtained using $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$, showing its good performance in presence of outliers.

Hospital Costs Data

$$\widehat{\eta}(z) = 0.8892\, z + 7.1268$$



$\widehat{\eta}_{\mathrm{CL}}$ in red
$\widehat{\eta}_{\mathrm{R}}$ in blue

- The linear fit (in black) seems to be a good choice for this data set, however, some discrepancies appear near the boundary.
- It is worth noting that in this case, the shape of the classical estimator (in red) is quite close to that of the robust one (in blue).

Summary

- We have defined a robust estimators for the regression parameter and the nonparametric function under the constraint that η_0 monotone.
- Our estimators are consistent and attain the optimal convergence rate.
- The estimators of the regression coefficient are asymptotically normally distributed.
- The simulation study illustrate the bad behaviour of the classical estimator when outliers are present.
- In particular, expected large responses affect the classical estimators of the nonparametric component.

Thanks for your attention.

Algorithm

Denote $\psi = \rho'$ and

$$r_i(\mathbf{b}, \mathbf{a}) = y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b} - \sum_{j=1}^{k_n} a_j\, B_j(z_i)$$

- **Step 1:**
  Let $m = 0$ and $(\mathbf{b}^{(0)}, \mathbf{a}^{(0)}) = (\widehat{\beta}, \widehat{\mathbf{a}})$ the $MM-$estimators computed without restrictions and $\widehat{\sigma}$ the scale given in the S-step.
- **Step 2:**
  - Given m define the weights

$$w_{i,m} = \psi\left(\frac{r_i(\mathbf{b}^{(m)}, \mathbf{a}^{(m)})}{\widehat{\sigma}}\right)\frac{\widehat{\sigma}}{r_i(\mathbf{b}^{(m)}, \mathbf{a}^{(m)})}$$

  - Define

$$y_{w,i} = w_{i,m}^{1/2} y_i \quad , \quad x_{w,i\ell} = w_{i,m}^{1/2} x_{i\ell} \quad , \quad B_{w,i\ell} = w_{i,m}^{1/2} B_\ell(z_i)$$

Algorithm

- **Step 2:**
  - Define

$$y_{w,i} = w_{i,m}^{1/2} y_i \quad , \quad x_{w,i\ell} = w_{i,m}^{1/2} x_{i\ell} \quad , \quad B_{w,i\ell} = w_{i,m}^{1/2} B_\ell(z_i)$$

  - Let $\mathbf{v}_i = (x_{w,i1}, \ldots, x_{w,ip_1}, B_{w,i1}, \ldots, B_{w,ip_2})^{\mathrm{T}}$, $\mathbf{y}_w = (y_{w,1}, \ldots, y_{w,n})^{\mathrm{T}}$ and $\mathbf{d} = \left(\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\lambda}^{\mathrm{T}}\right)^{\mathrm{T}}$. We solve the quadratic problem with monotone restrictions

$$\widehat{\mathbf{d}} \;=\; \min_{\mathbf{b}, a_1 \leq \cdots \leq a_{k_n}} \|\mathbf{y}_w - \mathbf{V}^{\mathrm{T}}\mathbf{d}\|^2 = \min_{\mathbf{b}, a_1 \leq \cdots \leq a_{k_n}} \sum_{i=1}^{n} w_{i,m} r_i^2(\mathbf{b}, \mathbf{a})$$

  - Define $\mathbf{b}^{(m+1)}$ as the first p components of $\widehat{\mathbf{d}}$ and $\mathbf{a}^{(m+1)}$ as the last ones.
- Go to step 2 and iterate until convergence.

Return

Algorithm

- **Step 1.**

Step 1.1 Compute an initial $S-$estimator $\widetilde{\boldsymbol{\nu}} = (\widetilde{\boldsymbol{\beta}}_n, \widetilde{\mathbf{a}}_n)$ as in Bianco *et al.* (2005), i.e.,

$$\widetilde{\boldsymbol{\nu}}_n = \underset{\mathbf{b},\mathbf{a}}{\operatorname{argmin}} \ \sigma_n(\mathbf{b}, \mathbf{a})$$

where

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{\sqrt{d(y_i - \mathbf{b}^{\mathrm{T}}\mathbf{x}_i - \mathbf{a}^{\mathrm{T}}\mathbf{B}_i)}}{\sigma_n(\mathbf{b}, \mathbf{a})}\right) = \frac{1}{2} \ ,$$

$$\widehat{\sigma}_n = \sigma_n(\widetilde{\boldsymbol{\beta}}_n, \widetilde{\mathbf{a}}_n)$$

Algorithm

- **Step 1.2.**

  Let $u \sim \log \Gamma(\alpha, 1)$ and $\sigma^*(\alpha)$ the solution of

$$\mathbb{E}\left[\rho\left(\frac{\sqrt{1-u-\exp(u)}}{\sigma^*(\alpha)}\right)\right] = \frac{1}{2},$$

  Compute

  - $\widehat{\alpha}_n = \sigma^{*-1}(\widehat{\sigma}_n)$
  - $\widehat{\kappa}_n = \max(\widehat{\sigma}_n, C_e(\widehat{\alpha}_n))$.

- Let $\widehat{\boldsymbol{\nu}}_n^{(0)}$ be $WMM-$estimator of $\boldsymbol{\nu}$ defined as

$$\widehat{\boldsymbol{\nu}}_n^{(0)} = \underset{(\mathbf{b},\mathbf{a})}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{\sqrt{d(y_i - \mathbf{b}^{\mathrm{T}}\mathbf{x}_i - \mathbf{a}^{\mathrm{T}}\mathbf{B}_i)}}{\widehat{\kappa}_n}\right) w(\mathbf{x}_i).$$

Algorithm

- **Step 2.**
  - ⋆ If $\widehat{a}_1^{(0)} \leq \widehat{a}_2^{(0)} \leq \cdots \leq \widehat{a}_{k_n}^{(0)}$, the final estimators are $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{(0)}$ and $\widehat{\eta}(t) = \sum_{j=1}^{k_n} \widehat{a}_j^{(0)} B_j(t)$.
  - ⋆ Otherwise, the final estimators are obtained using a standard minimization algorithm with restrictions choosing as initial value $(\widehat{\boldsymbol{\beta}}_n^{(0)}, \mathbf{a}^0)$, where $\mathbf{a}^0 \in \mathcal{L}_{k_n}$.
    One possible choice for $\mathbf{a}^0$ is $a_1^0 = a_2^0 = 0$ and $a_i^0 = i - 2$ for $i = 3, \ldots, k_n$.

Algorithm: Generalised Rosen Algorithm (Jamshidian, 2004)

- Denote $\widehat{\boldsymbol{\nabla}}$ the gradient function and $\widehat{\mathbf{H}}$ the gradient and negative Hessian of the objective function Let $\mathcal{A} = \{i_1, \ldots, i_m\}$ the set of indices such that $a_{i_j}^{(0)} = a_{i_j+1}^{(0)}$. If $m > 0$ define the working matrix as $\mathbf{A} \in \mathbb{R}^{m\times(k_n+p)}$ in which the $j-$th row is the vector with its i_j-th element equal to 1 and the $(i_j+1)-$th element equal to -1, the remaining ones equal to 0.
- Fix an initial value $\boldsymbol{\nu}$ (in the first step, $\boldsymbol{\nu} = (\widehat{\boldsymbol{\beta}}_n^{(0)}, \mathbf{a}^0)$ and denote $\widehat{\mathbf{H}} = \widehat{\mathbf{H}}(\boldsymbol{\nu})$, $\widehat{\boldsymbol{\nabla}} = \widehat{\boldsymbol{\nabla}}(\boldsymbol{\nu})$.
- **S1** Find the feasible direction as

$$\boldsymbol{\eta} = \left(\mathbf{I} - \widehat{\mathbf{H}}^{-1}\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\widehat{\mathbf{H}}^{-1}\mathbf{A}^{\mathrm{T}}\right)^{-1}\mathbf{A}\right)\widehat{\mathbf{H}}^{-1}\widehat{\boldsymbol{\nabla}}$$

Algorithm

- **S2** If $\|\boldsymbol{\eta}\| < \epsilon$ for some $\epsilon > 0$ small enough, compute the Lagrange multipliers

$$\boldsymbol{\mu} = \left(\mathbf{A}\widehat{\mathbf{H}}^{-1}\mathbf{A}^{\mathrm{T}}\right)^{-1}\mathbf{A}\widehat{\mathbf{H}}^{-1}\widehat{\boldsymbol{\nabla}}$$

  Let μ_i be the $i-$th component of $\boldsymbol{\mu}$.
  - If $\mu_i \geq 0$, for all $i \in \mathcal{A}$, then $\widehat{\boldsymbol{\nu}} = \boldsymbol{\nu}$.
  - If there exists at least one $i \in \mathcal{A}$ such that $\mu_i < 0$, determine the index corresponding to the largest μ_i and remove it from $\mathcal{A}$ and go to **S1**.
- **S3** Compute

$$\theta_1 = \min_{\eta_i > \eta_{i+1}, i \notin \mathcal{A}, 1 \leq i \leq k_n - 1} \frac{-(a_{i+1} - a_i)}{\eta_{i+1} - \eta_i}$$

  and find the smallest r such that $L_n(\boldsymbol{\nu} + 2^{-r}\boldsymbol{\eta}) > L_n(\boldsymbol{\nu})$. Then replace $\boldsymbol{\nu}$ by $\widetilde{\boldsymbol{\nu}} = \boldsymbol{\nu} + \min(2^{-r}, \theta_1)\boldsymbol{\eta})$, update $\mathcal{A}$ and $\mathbf{A}$ and go to **S1**.

Return

Results when $\eta_0 = \eta_{0,1}$

		Summary measures for $\widehat{\beta}$			$\mathrm{MISE}(\widehat{\eta})$
	Estimator	Bias	MSE	Cov.Prob	
C_0	CL	0.0002	0.0037	0.9340	0.0088
	ROB	0.0021	0.0045	0.9270	0.0096

Results when $\eta_0 = \eta_{0,1}$

		Summary measures for $\widehat{\beta}$			$\text{MISE}(\widehat{\eta})$
	Estimator	Bias	MSE	Cov.Prob	
C_0	CL	0.0002	0.0037	0.9340	0.0088
	ROB	0.0021	0.0045	0.9270	0.0096
C_1	CL	**-0.5497**	**0.3492**	**0.0050**	0.0265
	ROB	-0.0016	0.0050	0.8850	0.0100

Results when $\eta_0 = \eta_{0,1}$

	Estimator	Summary measures for $\widehat{\beta}$ Bias	MSE	Cov.Prob	$\mathrm{MISE}(\widehat{\eta})$
C_0	CL	0.0002	0.0037	0.9340	0.0088
	ROB	0.0021	0.0045	0.9270	0.0096
C_2	CL	**-1.8359**	**4.2426**	**0.0690**	**54.3390**
	ROB	0.0002	0.0051	0.9170	0.0103

Results when $\eta_0 = \eta_{0,1}$

		Summary measures for $\widehat{\beta}$			$\mathrm{MISE}(\widehat{\eta})$
	Estimator	Bias	MSE	Cov.Prob	
C_0	CL	0.0002	0.0037	0.9340	0.0088
	ROB	0.0021	0.0045	0.9270	0.0096
C_3	CL	**-1.9400**	**3.8376**	**0.0100**	**15.0401**
	ROB	0.0043	0.0053	0.8900	0.0146

Results

		Model 1					
		Summary measures for $\widehat{\beta}$					$\mathrm{MISE}(\widehat{\eta})$
	Estimator	Bias	SD	MSE	AS.SE	Cov.Prob	
C_0	CL	0.0002	0.0608	0.0037	0.0568	0.9340	0.0088
	ROB	0.0021	0.0672	0.0045	0.0620	0.9270	0.0096

Results

		Model 1					
		Summary measures for $\widehat{\beta}$					$\mathrm{MISE}(\widehat{\eta})$
	Estimator	Bias	SD	MSE	AS.SE	Cov.Prob	
C_0	CL	0.0002	0.0608	0.0037	0.0568	0.9340	0.0088
	ROB	0.0021	0.0672	0.0045	0.0620	0.9270	0.0096
C_1	CL	**-0.5497**	**0.2170**	**0.3492**	0.0535	**0.0050**	0.0265
	ROB	-0.0016	0.0706	0.0050	0.0591	0.8850	0.0100

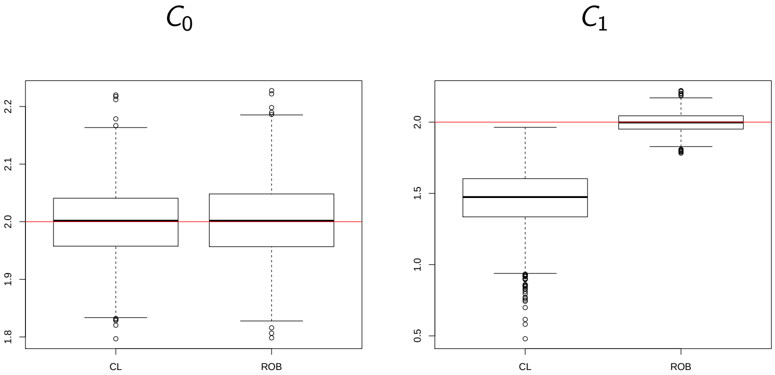
Results

		Model 1					
		Summary measures for $\widehat{\beta}$					$\mathrm{MISE}(\widehat{\eta})$
	Estimator	Bias	SD	MSE	AS.SE	Cov.Prob	
C_0	CL	0.0002	0.0608	0.0037	0.0568	0.9340	0.0088
	ROB	0.0021	0.0672	0.0045	0.0620	0.9270	0.0096
C_2	CL	**-1.8359**	**0.9343**	**4.2426**	**0.3781**	**0.0690**	**54.3390**
	ROB	0.0002	0.0711	0.0051	0.0639	0.9170	0.0103

Results

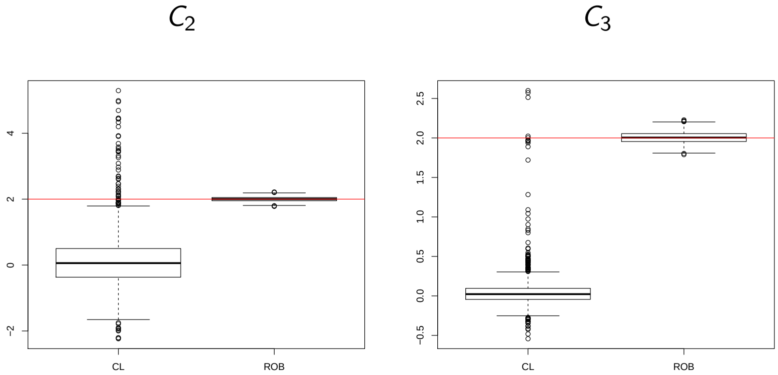
		Model 1					
		Summary measures for $\widehat{\beta}$					$\mathrm{MISE}(\widehat{\eta})$
	Estimator	Bias	SD	MSE	AS.SE	Cov.Prob	
C_0	CL	0.0002	0.0608	0.0037	0.0568	0.9340	0.0088
	ROB	0.0021	0.0672	0.0045	0.0620	0.9270	0.0096
C_3	CL	**-1.9400**	0.2721	**3.8376**	0.1848	**0.0100**	**15.0401**
	ROB	0.0043	0.0727	0.0053	0.0598	0.8900	0.0146

Boxplots for $\widehat{\beta}$, Model 1



Boxplots of $\widehat{\beta}$, under a Gamma Model with $\eta_0 = \eta_{0,1}$, $c_w = \sqrt{\chi^2_{0.975,1}}$.

Boxplots for $\widehat{\beta}$, Model 1



Boxplots of $\widehat{\beta}$, under a Gamma Model with $\eta_0 = \eta_{0,1}$, $c_w = \sqrt{\chi^2_{0.975,1}}$.