Constrained Regression Using mboost: An Application to Body Fat Prediction

Benjamin Hofner [1]

Institut für Medizininformatik, Biometrie und Epidemiologie (IMBE)

Friedrich-Alexander-Universität Erlangen-Nürnberg

Statistical Computing 2010

42nd Workshop

Reisensburg - 2010

[1] benjamin.hofner@imbe.med.uni-erlangen.de

Body Fat

Impact and Effect

Overweight (WHO: BMI > 25) and obesity (WHO: BMI > 30)

- are widespread problems
- increase the risk for
  - cardiovascular diseases (e.g., stroke, myocardial infarction)
  - diabetes mellitus (type 2)
  - cancer
  - ...

$\Rightarrow$ have an huge impact on public health.

Measurement of Body Fat

Gold Standards

- hydrostatic weighing (i.e., underwater weighing):
  - time consuming
  - "wet"
  - ...
- DXA (dual energy X-ray absorptiometry) measurement:
  - (quite) expansive
  - exposure to radiation
  - ...

Anthropometric methods

combine measures such as skinfold (SF) thickness and body circumferences (CF) and provide an easy and cheap alternative.

Garcia et al. [2005] propose a linear model to predict DXA body fat measurements (DEXfat) by anthropometric measurements:

```
R> lm(DEXfat ~ hipcirc + anthro3a + kneebreadth)
```

Results for Women

$$\begin{aligned}\text{DEXfat} = & -75.2 + 0.5\ \text{hipcirc} \\ & + 8.9 \left\{\log(\text{chinSF}) + \log(\text{tricepsSF}) + \log(\text{subscapularSF})\right\} \\ & + 1.9\ \text{kneebreadth}\end{aligned}$$

Data Set [Garcia et al., 2005]

DEXfat	body fat measured by DXA, response variable
age	age in years
waistcirc	waist circumference
hipcirc	hip circumference
elbowbreadth	breadth of the elbow
kneebreadth	breadth of the knee
anthro3a	sum of logarithm of three anthropometric measurements
anthro3b	sum of logarithm of three anthropometric measurements
anthro3c	sum of logarithm of three anthropometric measurements
anthro4	sum of logarithm of three anthropometric measurements

Garcia et al. [2005] propose a linear model to predict DXA body fat measurements (DEXfat) by anthropometric measurements:

```
R> lm(DEXfat ~ hipcirc + anthro3a + kneebreadth)
```

Results for Women

$$\begin{aligned} \text{DEXfat} = & -75.2 + 0.5 \text{ hipcirc} \\ & + 8.9 \left\{\log(\text{chinSF}) + \log(\text{tricepsSF}) + \log(\text{subscapularSF})\right\} \\ & + 1.9 \text{ kneebreadth} \end{aligned}$$

Data Set [Garcia et al., 2005]

DEXfat	body fat measured by DXA, response variable
age	age in years
waistcirc	waist circumference
hipcirc	hip circumference
elbowbreadth	breadth of the elbow
kneebreadth	breadth of the knee
anthro3a	sum of logarithm of three anthropometric measurements
anthro3b	sum of logarithm of three anthropometric measurements
anthro3c	sum of logarithm of three anthropometric measurements
anthro4	sum of logarithm of three anthropometric measurements

- **Idea:**
  Enhance the linear model by using smooth effects.
- **Problem:**
  Most covariates are assumed to have a monotonic increasing effect.
- **Questions:**
  - Can we combine smooth effects with monotonic constraints?
  - Can we reduce the set of body measures to a relevant subset?
    (= Variable Selection)

Short Review of P-splines

[Eilers and Marx, 1996]

Smooth function can be expressed with B-splines as

$$f(x) = \sum_{j=1}^{J} \beta_j B_j(x; l) = \boldsymbol{\beta}^\top \mathbf{B}(x),$$

where $B_j(\cdot; l)$ is the j-th B-spline basis function of **degree** l defined on a **grid of knots** ξ_k.

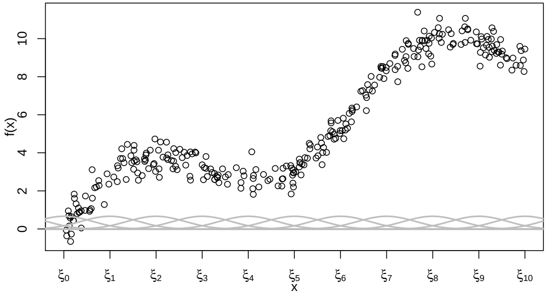
Short Review of P-splines

[Eilers and Marx, 1996]

Smooth function can be expressed with B-splines as

$$f(x) = \sum_{j=1}^{J} \beta_j B_j(x; l) = \boldsymbol{\beta}^\top \mathbf{B}(x),$$

where $B_j(\cdot; l)$ is the j-th B-spline basis function of **degree** l defined on a **grid of knots** ξ_k.



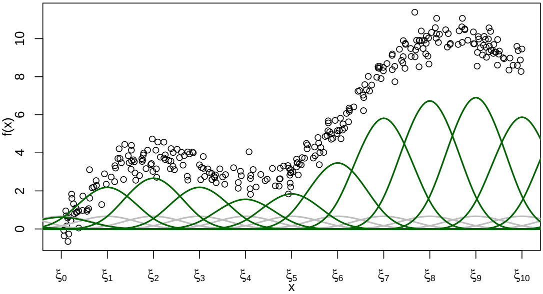
Short Review of P-splines

[Eilers and Marx, 1996]

Smooth function can be expressed with B-splines as

$$f(x) = \sum_{j=1}^{J} \beta_j B_j(x; l) = \boldsymbol{\beta}^\top \mathbf{B}(x),$$

where $B_j(\cdot; l)$ is the j-th B-spline basis function of **degree** l defined on a **grid of knots** ξ_k.



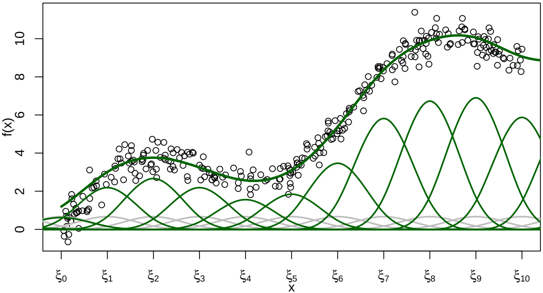
Short Review of P-splines

[Eilers and Marx, 1996]

Smooth function can be expressed with B-splines as

$$f(x) = \sum_{j=1}^{J} \beta_j B_j(x; l) = \boldsymbol{\beta}^\top \mathbf{B}(x),$$

where $B_j(\cdot; l)$ is the j-th B-spline basis function of **degree** l defined on a **grid of knots** ξ_k.



- Smoothness enforced by additional **penalty on adjacent B-splines**:

$$\mathcal{J}(\beta; d) = \sum_{j=d+1}^{J} (\Delta^d \beta_j)^2,$$

  where d is the order of the difference penalty.
- **Difference operator** Δ^d is defined as:

$$\Delta\beta_j = \Delta^1\beta_j = (\beta_j - \beta_{j-1})$$
$$\Delta^2\beta_j = \Delta(\Delta\beta_j) = \beta_j - 2\beta_{j-1} + \beta_{j-2}$$

- **Difference matrix** $\mathbf{D}_{(d)}$

$$\mathbf{D}_{(1)} = \begin{pmatrix} -1 & 1 & 0 & \dots \\ 0 & -1 & 1 & \dots \\ 0 & 0 & \ddots & \ddots \end{pmatrix} \quad \mathbf{D}_{(2)} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots \\ 0 & 1 & -2 & 1 & \dots \\ 0 & 0 & \ddots & \ddots & \ddots \end{pmatrix}$$

Estimation: Penalized least squares criterion

$$\mathcal{Q}(\beta) = (\mathbf{y} - \mathbf{B}\beta)^\top(\mathbf{y} - \mathbf{B}\beta) + \lambda\mathcal{J}(\beta; d)$$

- Smoothness enforced by additional **penalty on adjacent B-splines**:

$$\mathcal{J}(\boldsymbol{\beta}; d) = \sum_{j=d+1}^{J} (\Delta^d \beta_j)^2 = \boldsymbol{\beta}^\top \mathbf{D}_{(d)}^\top \mathbf{D}_{(d)} \boldsymbol{\beta},$$

  where d is the order of the difference penalty.
- **Difference operator** Δ^d is defined as:

$$\Delta\beta_j = \Delta^1\beta_j = (\beta_j - \beta_{j-1})$$
$$\Delta^2\beta_j = \Delta(\Delta\beta_j) = \beta_j - 2\beta_{j-1} + \beta_{j-2}$$

- **Difference matrix** $\mathbf{D}_{(d)}$

$$\mathbf{D}_{(1)} = \begin{pmatrix} -1 & 1 & 0 & \dots \\ 0 & -1 & 1 & \dots \\ 0 & 0 & \ddots & \ddots \end{pmatrix} \quad \mathbf{D}_{(2)} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots \\ 0 & 1 & -2 & 1 & \dots \\ 0 & 0 & \ddots & \ddots & \ddots \end{pmatrix}$$

Estimation: Penalized least squares criterion

$$\mathcal{Q}(\beta) = (\mathbf{y} - \mathbf{B}\beta)^\top(\mathbf{y} - \mathbf{B}\beta) + \lambda\mathcal{J}(\beta; d)$$

- Smoothness enforced by additional **penalty on adjacent B-splines**:

$$\mathcal{J}(\boldsymbol{\beta}; d) = \sum_{j=d+1}^{J} (\Delta^d \beta_j)^2 = \boldsymbol{\beta}^\top \mathbf{D}_{(d)}^\top \mathbf{D}_{(d)} \boldsymbol{\beta},$$

  where d is the order of the difference penalty.

- **Difference operator** Δ^d is defined as:

$$\Delta\beta_j = \Delta^1\beta_j = (\beta_j - \beta_{j-1})$$
$$\Delta^2\beta_j = \Delta(\Delta\beta_j) = \beta_j - 2\beta_{j-1} + \beta_{j-2}$$

- **Difference matrix** $\mathbf{D}_{(d)}$

$$\mathbf{D}_{(1)} = \begin{pmatrix} -1 & 1 & 0 & \dots \\ 0 & -1 & 1 & \dots \\ 0 & 0 & \ddots & \ddots \end{pmatrix} \quad \mathbf{D}_{(2)} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots \\ 0 & 1 & -2 & 1 & \dots \\ 0 & 0 & \ddots & \ddots & \ddots \end{pmatrix}$$

Estimation: Penalized least squares criterion

$$\mathcal{Q}(\beta) = (\mathbf{y} - \mathbf{B}\beta)^\top (\mathbf{y} - \mathbf{B}\beta) + \lambda \mathcal{J}(\beta; d)$$

Monotonic P-splines

- ✓ **Smoothness Constraint:**
  Difference penalty on adjacent coefficients of B-splines
- **Monotonicity Constraint:**

$$f'(x) = \frac{\partial}{\partial x} \sum \beta_j B_j(x; l) = \frac{1}{h} \sum \Delta^1 \beta_{j+1} B_j(x; l-1)$$

(distance of knots $h > 0$, B-spline basis $B_j(x; l-1) \geq 0$)

$\Rightarrow$ depends only on the first differences of the adjacent coefficients

- Monotonic increasing function:

$$f'(x) > 0 \quad \Leftrightarrow \quad \Delta^1 \beta_{j+1} > 0 \quad \forall j$$

- Monotonic decreasing function:

$$f'(x) < 0 \quad \Leftrightarrow \quad \Delta^1 \beta_{j+1} < 0 \quad \forall j$$

- **Convexity / Concavity Constraint:**

$$f''(x) = \frac{1}{h^2} \sum \Delta^2 \beta_{j+2} B_j(x; l-2)$$

Monotonic P-splines

✓ **Smoothness Constraint:**
Difference penalty on adjacent coefficients of B-splines

- **Monotonicity Constraint:**

$$f'(x) = \frac{\partial}{\partial x} \sum \beta_j B_j(x; l) = \frac{1}{h} \sum \Delta^1 \beta_{j+1} B_j(x; l-1)$$

(distance of knots $h > 0$, B-spline basis $B_j(x; l-1) \geq 0$)

$\Rightarrow$ depends only on the first differences of the adjacent coefficients

- Monotonic increasing function:

$$f'(x) > 0 \quad \Leftrightarrow \quad \Delta^1 \beta_{j+1} > 0 \quad \forall j$$

- Monotonic decreasing function:

$$f'(x) < 0 \quad \Leftrightarrow \quad \Delta^1 \beta_{j+1} < 0 \quad \forall j$$

- **Convexity / Concavity Constraint:**

$$f''(x) = \frac{1}{h^2} \sum \Delta^2 \beta_{j+2} B_j(x; l-2)$$

Monotonic P-splines

- ✓ **Smoothness Constraint:**
  Difference penalty on adjacent coefficients of B-splines
- **Monotonicity Constraint:**

$$f'(x) = \frac{\partial}{\partial x} \sum \beta_j B_j(x; l) = \frac{1}{h} \sum \Delta^1 \beta_{j+1} B_j(x; l-1)$$

(distance of knots $h > 0$, B-spline basis $B_j(x; l-1) \geq 0$)

$\Rightarrow$ depends only on the first differences of the adjacent coefficients

- Monotonic increasing function:

$$f'(x) > 0 \quad \Leftrightarrow \quad \Delta^1 \beta_{j+1} > 0 \quad \forall j$$

- Monotonic decreasing function:

$$f'(x) < 0 \quad \Leftrightarrow \quad \Delta^1 \beta_{j+1} < 0 \quad \forall j$$

- **Convexity / Concavity Constraint:**

$$f''(x) = \frac{1}{h^2} \sum \Delta^2 \beta_{j+2} B_j(x; l-2)$$

- "Monotonic coefficients" can be achieved by (additional) asymmetric difference penalties on adjacent coefficients Eilers [2005]:

$$\mathcal{J}_{\text{asym}}(\boldsymbol{\beta}; c) = \sum_{j=c+1}^{J} v_j (\Delta^c \beta_j)^2, = \boldsymbol{\beta}^\top \mathbf{D}_{(c)}^\top \mathbf{V} \mathbf{D}_{(c)} \boldsymbol{\beta},$$

  where c is the order of the difference penalty.

- Important difference to P-spline penalty are weights v_j, which are specified as

$$v_j = \begin{cases} 0 & \text{if } \Delta^c \beta_j > 0 \\ 1 & \text{if } \Delta^c \beta_j \leq 0, \end{cases} \text{ monotonic increasing}$$

  matrix notation: $\mathbf{V} = \text{diag}(\mathbf{v})$, and difference matrices as above.

- Monotonicity constraint if $c = 1$
- (Convex / concave constraint if $c = 2$)

- "Monotonic coefficients" can be achieved by (additional) asymmetric difference penalties on adjacent coefficients Eilers [2005]:

$$\mathcal{J}_{\text{asym}}(\boldsymbol{\beta}; c) = \sum_{j=c+1}^{J} v_j (\Delta^c \beta_j)^2, = \boldsymbol{\beta}^\top \mathbf{D}_{(c)}^\top \mathbf{V} \mathbf{D}_{(c)} \boldsymbol{\beta},$$

  where c is the order of the difference penalty.

- Important difference to P-spline penalty are weights v_j, which are specified as

$$v_j = \begin{cases} 0 & \text{if } \Delta^c \beta_j > 0 \\ 1 & \text{if } \Delta^c \beta_j \leq 0, \end{cases} \text{ monotonic increasing}$$

  matrix notation: $\mathbf{V} = \text{diag}(\mathbf{v})$, and difference matrices as above.

- Monotonicity constraint if $c = 1$
- (Convex / concave constraint if $c = 2$)

- "Monotonic coefficients" can be achieved by (additional) asymmetric difference penalties on adjacent coefficients Eilers [2005]:

$$\mathcal{J}_{\text{asym}}(\boldsymbol{\beta}; c) = \sum_{j=c+1}^{J} v_j (\Delta^c \beta_j)^2, = \boldsymbol{\beta}^\top \mathbf{D}_{(c)}^\top \mathbf{V} \mathbf{D}_{(c)} \boldsymbol{\beta},$$

  where c is the order of the difference penalty.

- Important difference to P-spline penalty are weights v_j, which are specified as

$$v_j = \begin{cases} 0 & \text{if } \Delta^c \beta_j < 0 \\ 1 & \text{if } \Delta^c \beta_j \geq 0, \end{cases} \text{ monotonic decreasing}$$

  matrix notation: $\mathbf{V} = \text{diag}(\mathbf{v})$, and difference matrices as above.

- Monotonicity constraint if $c = 1$
- (Convex / concave constraint if $c = 2$)

- "Monotonic coefficients" can be achieved by (additional) asymmetric difference penalties on adjacent coefficients Eilers [2005]:

$$\mathcal{J}_{\text{asym}}(\boldsymbol{\beta}; c) = \sum_{j=c+1}^{J} v_j (\Delta^c \beta_j)^2, = \boldsymbol{\beta}^\top \mathbf{D}_{(c)}^\top \mathbf{V} \mathbf{D}_{(c)} \boldsymbol{\beta},$$

  where **c is the order of the difference penalty**.

- Important difference to P-spline penalty are weights v_j, which are specified as

$$v_j = \begin{cases} 0 & \text{if } \Delta^c \beta_j < 0 \\ 1 & \text{if } \Delta^c \beta_j \geq 0, \end{cases} \text{monotonic decreasing}$$

  matrix notation: $\mathbf{V} = \text{diag}(\mathbf{v})$, and difference matrices as above.

- Monotonicity constraint if $c = 1$
- (Convex / concave constraint if $c = 2$)

Estimation of Constrained P-splines

Weights v_j depend on $\boldsymbol{\beta}$. Thus:

1) Start with standard P-spline estimate
2) Compute weights v_j for $\hat{\boldsymbol{\beta}}$
3) Minimize

$$\mathcal{Q}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{B}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{B}\boldsymbol{\beta}) + \lambda_1 \mathcal{J}(\boldsymbol{\beta}; d) + \lambda_2 \mathcal{J}_{\text{asym}}(\boldsymbol{\beta}; c),$$

by **penalized least squares** estimate

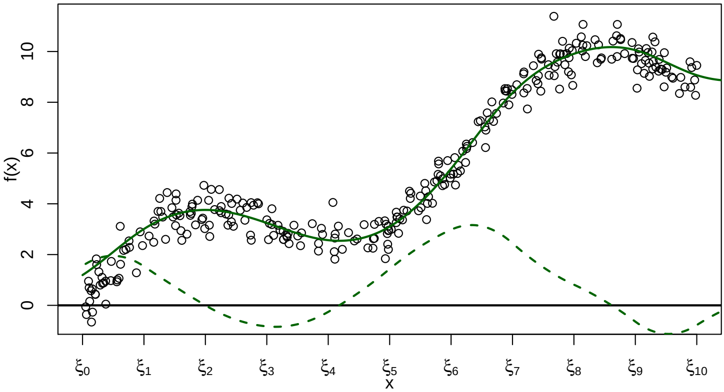
$$\hat{\boldsymbol{\beta}} = (\mathbf{B}^\top\mathbf{B} + \lambda_1 \mathbf{D}_{(d)}^\top \mathbf{D}_{(d)} + \lambda_2 \mathbf{D}_{(c)}^\top \mathbf{V} \mathbf{D}_{(c)})^{-1} \mathbf{B}^\top \mathbf{y}$$

4) Recompute weights v_j with $\hat{\boldsymbol{\beta}}$
5) Iterate 3) and 4) until no more changes in v_j (usually after 2-3 steps)

Smoothing parameter λ_2 is chosen quite large (e.g., 10^6), where larger values are associated with a stronger impact of the monotonic constraint on the estimation

B-Splines, P-Splines, Monotonic Splines

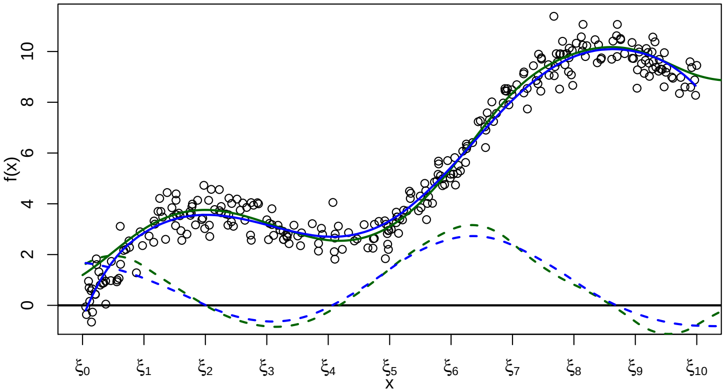
and their Derivatives



- B-spline estimate and derivative (dashed)
- P-spline estimate and derivative (dashed)
- Monotonic spline estimate and derivative (dashed)

B-Splines, P-Splines, Monotonic Splines

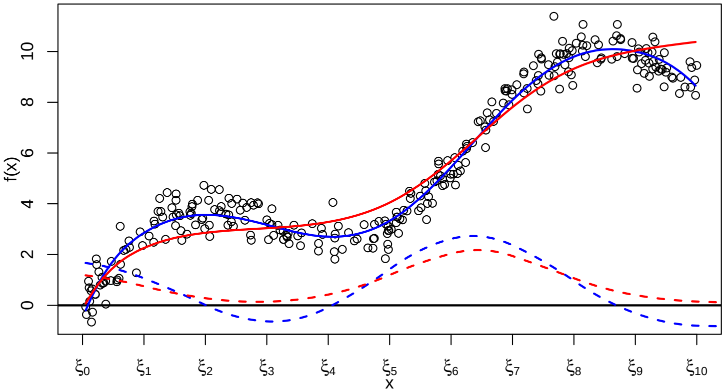
and their Derivatives



- B-spline estimate and derivative (dashed)
- P-spline estimate and derivative (dashed)
- Monotonic spline estimate and derivative (dashed)

B-Splines, P-Splines, Monotonic Splines

and their Derivatives



- B-spline estimate and derivative (dashed)
- P-spline estimate and derivative (dashed)
- Monotonic spline estimate and derivative (dashed)

How . . .

- . . . can we estimate generalized additive models with monotonicity constraint?
- . . . can we include variable selection?

Boosting offers an easy way to combine both tasks.

Furthermore, it offers the possibility to include further types of effects as categorical effects, spatial effects, random effects etc. (not shown in this talk).

Boosting (in a Nutshell)

Structured Additive Model

$$\mu_i = \mathbb{E}(y|\mathbf{x}_i) = h(\eta_i(\mathbf{x}_i))$$

with response function h and additive predictor

$$\eta_i(\mathbf{x}_i) = \beta_0 + \sum_{j=1}^{J} f_j(\mathbf{x}_i),$$

where f_j is (e.g.) a smooth or monotonic function.

- Model fitting aims at **minimizing the expected loss** with appropriate loss function ρ, e.g.,
  squared error loss $(y - \eta(\mathbf{x}))^2$ for "least squares" models
  negative log-likelihood for GLMs
- In practice: Minimization of the **empirical risk**

$$\mathcal{R} = n^{-1} \sum_{i=1}^{n} \rho(y_i, \eta_i(\mathbf{x}_i))$$

Boosting

- minimizes empirical risk (e.g., negative log likelihood)
- in a stagewise fashion
- via functional gradient descent (FGD).

In each iteration m

- negative gradient of the loss function $u_i^{[m]} = -\left.\frac{\partial \rho(y_i, \eta)}{\partial \eta}\right|_{\eta = \hat{\eta}_i^{[m-1]}}$ is estimated via **(penalized) least squares** base-learners $(\hat{\mathbf{u}}^{[m]} = \hat{g}_j(\mathbf{x}) \quad \forall\, j)$
- update only model term corresponding to best-fitting base-learner $\hat{g}_{j^*}$: add a **small fraction ν of the estimate** $\hat{g}_{j^*}$ (e.g., 10%) to the model

$\Rightarrow$ variable selection is achieved

Practical notes

- Base-learners represent functions $f_j(\cdot)$ from structured additive predictor (in the simplest case)
- We get an interpretable model similar to models from MLE
- Regularization via base-learner selection and shrinkage

Prediction of Body Fat

Monotonic Regression using Boosting

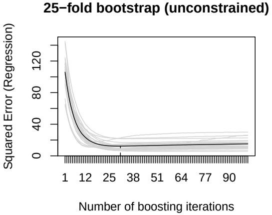
Two models are fitted using `gamboost()` [*R* package **mboost**: Hothorn et al., 2010].

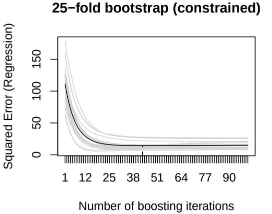
- P-spline Smoothing:

```
DEXfat ~ bbs(age) + bbs(waistcirc) + bbs(hipcirc) +
     bbs(elbowbreadth) + bbs(kneebreadth) + bbs(anthro3a) +
     bbs(anthro3b) + bbs(anthro3c) + bbs(anthro4)
```

- Constrained P-spline Smoothing:

```
DEXfat ~ bbs(age) + bmono(waistcirc, constraint = "inc") +
     bmono(hipcirc, constraint = "inc") + bmono(elbowbreadth,
     constraint = "inc") + bmono(kneebreadth, constraint = "inc") +
     bmono(anthro3a, constraint = "inc") + bmono(anthro3b,
     constraint = "inc") + bmono(anthro3c, constraint = "inc") +
     bmono(anthro4, constraint = "inc")
```

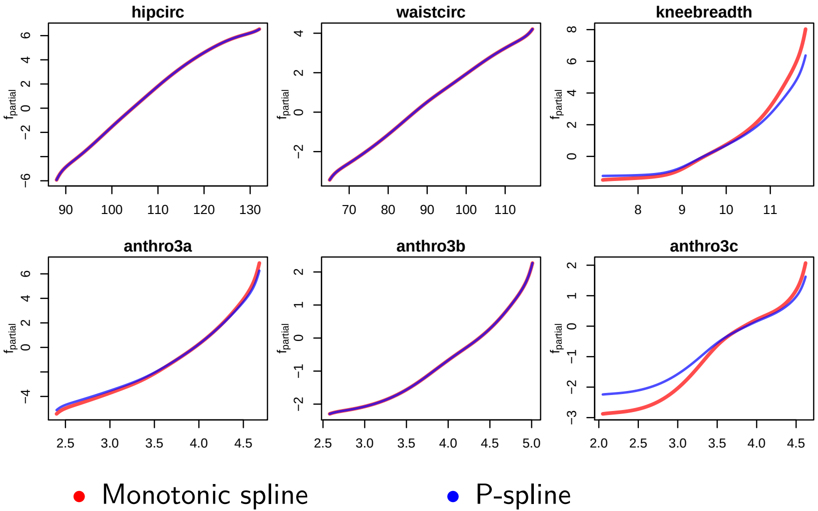




- **Unconstrained Model** (mstop = 31)
  - bbs(waistcirc)
  - bbs(hipcirc)
  - bbs(kneebreadth)
  - bbs(anthro3a)
  - bbs(anthro3b)
  - bbs(anthro3c)
- **Constrained Model** (mstop = 43)
  - bmono(waistcirc, constraint = "inc")
  - bmono(hipcirc, constraint = "inc")
  - bmono(kneebreadth, constraint = "inc")
  - bmono(anthro3a, constraint = "inc")
  - bmono(anthro3b, constraint = "inc")
  - bmono(anthro3c, constraint = "inc")

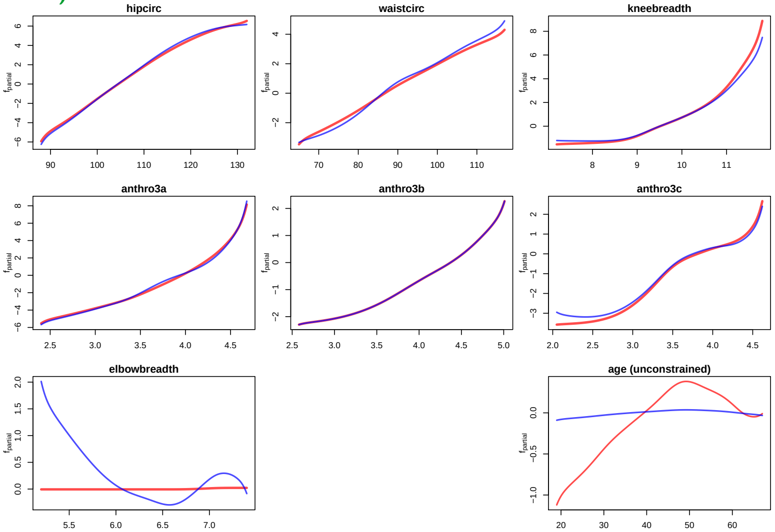
Effect Estimates

(early stopping)



Effect Estimates

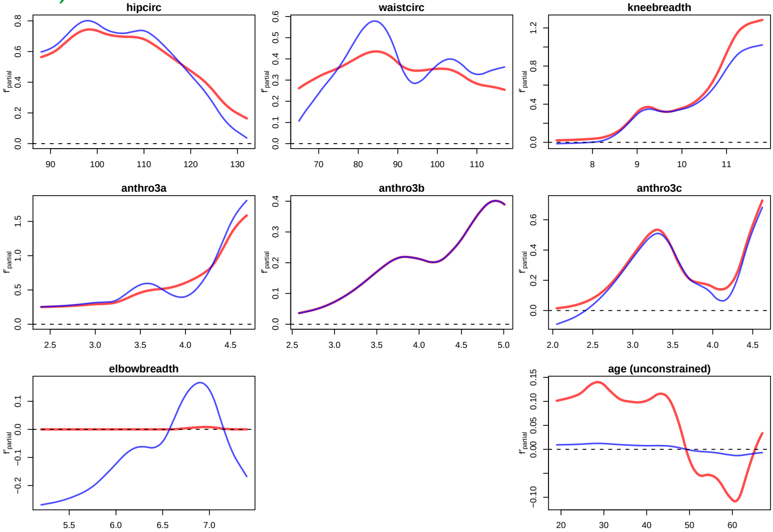
($m_{\text{stop}} = 100$)



- Monotonic spline
- P-spline

Derivatives

($m_{\text{stop}} = 100$)



- Monotonic spline
- P-spline

Messages

Summary & Outlook

- ✓ Monotonic P-splines enforce monotonicity
- ✓ True effects monotone
  $\Rightarrow$ (almost) same estimates as unconstrained estimation
- ✓ Estimated effects can be nicely interpreted
- ✓ Estimated effects are smooth
  (in contrast to other approaches: de Leeuw et al. [2009] (PAVA), Dette et al. [2006])
- ✓ Fast estimation procedure
  (same magnitude of speed as P-splines)
- ✓ Idea can be transferred to estimate monotonic effects for ordinal covariates
- ✓ Suits nicely in the boosting framework:
  - In essence penalized least squares estimate (in contrast to procedures based on L_1 penalties as COBS [He and Ng, 1999])
  - No need to change the boosting procedure itself [cf. "Monotonic Boosting": Leitenstorfer and Tutz, 2007, Tutz and Leitenstorfer, 2007]

Literature

J. de Leeuw, K. Hornik, and P. Mair. Isotone optimization in r: Pool-Adjacent-Violators algorithm (PAVA) and active set methods. Journal of Statistical Software, 32(05), 2009.

Holger Dette, Natalie Neumeyer, and Kay F. Pilz. A simple nonparametric estimator of a strictly monotone regression function. Bernoulli, 12(3):469–490, 2006.

Paul H. C. Eilers. Unimodal smoothing. Journal of Chemometrics, 19:317–328, 2005. doi: 10.1002/cem.935.

Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with B-splines and penalties. Statistical Science, 11:89–121, 1996.

Ada L. Garcia, Karen Wagner, Torsten Hothorn, Corinna Koebnick, Hans-Joachim F. Zunft, and Ulrike Trippo. Improved prediction of body fat by measuring skinfold thickness, circumferences, and bone breadths. Obesity Research, 13:626–634, 2005.

X. He and P. Ng. COBS: qualitatively constrained smoothing via linear programming. Computational Statistics, 14(3):315–338, 1999.

Torsten Hothorn, Peter Bühlmann, Thomas Kneib, Matthias Schmid, and Benjamin Hofner. mboost: Model-Based Boosting, 2010. URL `http://r-forge.r-project.org/projects/mboost`. R package version 2.1-0.

F. Leitenstorfer and G. Tutz. Generalized monotonic regression based on b-splines with an application to air pollution data. Biostatistics, 8(3):654, 2007.

G. Tutz and F. Leitenstorfer. Generalized smooth monotonic regression in additive modeling. Journal of Computational and Graphical Statistics, 16(1):165–188, 2007.