Score estimation in monotone single index models

Kim Hendrickx

Hasselt University

The regression parameter α_0 is our parameter of interest in the Single Index Model

The single index model (SIM):

$$Y = \psi_0(\boldsymbol{\alpha}_0^T \boldsymbol{X}) + \varepsilon,$$

where

- Y is a response variable
- $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ is a d-dimensional covariate ($d \geq 1$)
- $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^T$ is a d-dimensional regression parameter
- ψ is an unknown link function
- ε is a random error term with $E(\varepsilon|\boldsymbol{X}) = 0$

The monotone single index model:

- ψ is an unknown **monotone** link function

Identifiability restrictions are needed in Single Index Models

Take $a, b \in \mathbb{R}$ and let ψ^* be the function defined by the relationship

$$\psi^*(a + bt) = \psi_0(t),$$

for all t in the support of $\boldsymbol{\alpha}_0^T \boldsymbol{X}$, then

$$E(Y|\boldsymbol{X}) = \psi_0(\boldsymbol{\alpha}_0^T \boldsymbol{X}) = \psi^*(a + b\boldsymbol{\alpha}_0^T \boldsymbol{X}).$$

Restrictions:

- Location normalization: $\boldsymbol{X}$ cannot contain an intercept
- Scale normalization:

$$\{\boldsymbol{\alpha} \in \mathbb{R}^d : \alpha_1 = 1\} \quad \text{or} \quad \{\boldsymbol{\alpha} \in \mathbb{R}^d : \|\boldsymbol{\alpha}\| = 1, \alpha_1 \geq 0\}.$$

In our monotone SIM we use: $\{\boldsymbol{\alpha} \in \mathbb{R}^d : \|\boldsymbol{\alpha}\| = 1\}$.

The Least Squares Estimator minimizes the sum of squares

Consider the sum of squared errors

$$S_n(\boldsymbol{\alpha}, \psi) = \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i - \psi(\boldsymbol{\alpha}^T \boldsymbol{X}_i) \right\}^2,$$

The Least Squares Estimator (LSE) $(\hat{\boldsymbol{\alpha}}_n, \hat{\psi}_n)$ is defined by

$$(\hat{\boldsymbol{\alpha}}_n, \hat{\psi}_n) = \arg\min_{(\boldsymbol{\alpha}, \psi)} S_n(\boldsymbol{\alpha}, \psi).$$

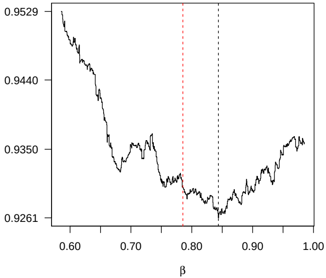
Example via simulation model:

$Y = \exp(X_1/\sqrt{2} + X_2/\sqrt{2}) + \varepsilon,$

$X_1, X_2 \sim U[-1, 1]$ and $\varepsilon \sim N(0, 1)$

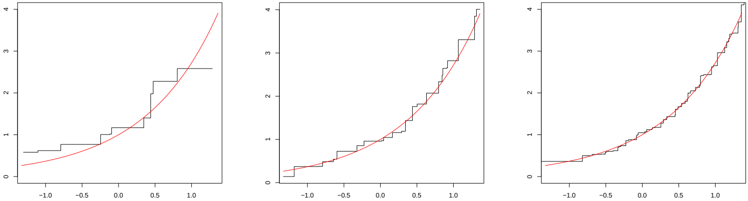
$(\alpha_1, \alpha_2) = (\cos(\beta), \sin(\beta))$

$\beta_0 = \pi/4$



The LSE of the link function is obtained first

Step 1: $\hat{\psi}_{n,\boldsymbol{\alpha}} = \arg\min_{\psi} S_n(\boldsymbol{\alpha}, \psi)$ (under monotonicity)

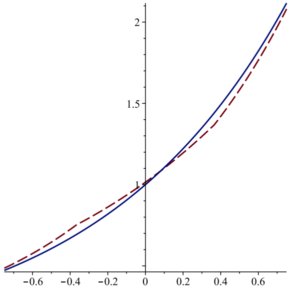


(a) $n = 100$ **(b)** $n = 1,000$ **(c)** $n = 10,000$

Step 2: $\hat{\boldsymbol{\alpha}}_n = \arg\min_{\boldsymbol{\alpha}} S_n(\boldsymbol{\alpha}, \hat{\psi}_{n,\boldsymbol{\alpha}})$

The LSE of the link function converges to ψ_α

Step 1: $\hat{\psi}_{n,\boldsymbol{\alpha}} = \arg\min_\psi S_n(\boldsymbol{\alpha}, \psi) \rightarrow \psi_{\boldsymbol{\alpha}} \stackrel{def}{=} E\left[\psi_0(\boldsymbol{\alpha}_0^T \boldsymbol{X}) | \boldsymbol{\alpha}^T \boldsymbol{X} = \cdot\right]$



The real $\psi_0 = \exp(x)$ (blue, solid) and the function $\psi_{\boldsymbol{\alpha}}$ (red, dashed) for $\alpha_{01} = \alpha_{02} = 1/\sqrt{2}$ and $\alpha_1 = 1/2, \alpha_2 = \sqrt{3}/2$.

Asymptotics:

$$\sup_{\boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta_0)} \int \left\{\hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{x}) - \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{x})\right\}^2 dG(\boldsymbol{x}) = O_p\left((\log n)^2 n^{-2/3}\right).$$

The limiting distribution of the LSE remains an open problem

$$\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0 = O_p(n^{-1/3})$$

$$\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0 = O_p(n^{-1/2})?$$

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) \to N(0, \Sigma)?$$

We parametrize the unit sphere and consider a score approach

Local parametrization:

$\mathbb{S} : \mathbb{R}^{d-1} \to \{\boldsymbol{\alpha} \in \mathbb{R}^d : \|\boldsymbol{\alpha}\| = 1\}$ such that there exists a unique vector $\boldsymbol{\beta} \in \mathbb{R}^{d-1}$ satisfying

$$\boldsymbol{\alpha} = \mathbb{S}(\boldsymbol{\beta}) = \mathbb{S}(\beta_1, \ldots, \beta_{d-1}) = (\mathbb{S}_1(\boldsymbol{\beta}), \ldots \mathbb{S}_d(\boldsymbol{\beta}))^T.$$

Differentiate the Sum of Squares:

Consider the partial derivatives w.r.t. $\beta_1, ...\beta_{d-1}$ of

$$\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \psi(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i)\right\}^2,$$

given by (for $j = 1, \ldots, d-1$)

$$\frac{2}{n}\sum_{i=1}^{n}\left(\frac{\partial \mathbb{S}(\boldsymbol{\beta})}{\partial \beta_j}\right)^T \boldsymbol{X}_i \psi'(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i)\left\{Y_i - \psi(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i)\right\},$$

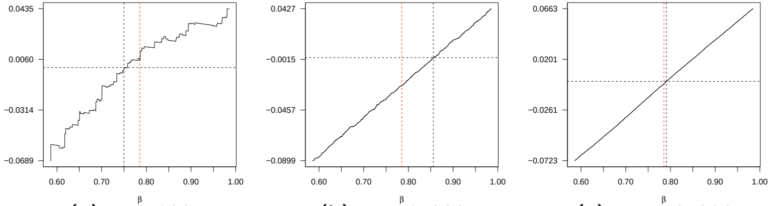
Let's consider the score instead of the minimization approach

Replace ψ by $\hat{\psi}_{n,\boldsymbol{\alpha}} \equiv \hat{\psi}_{n,\mathbb{S}(\boldsymbol{\beta})}$ and solve the score equations:

$$\sum_{i=1}^{n} \left(\frac{\partial \mathbb{S}(\boldsymbol{\beta})}{\partial \beta_j} \right)^T \boldsymbol{X}_i \hat{\psi}'_{n,\boldsymbol{\alpha}}(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i) \left\{ Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i) \right\} = 0.$$

Since the derivative of $\hat{\psi}_{n,\boldsymbol{\alpha}}$ is not defined, simplify the score equations:

$$Z_n(\boldsymbol{\beta}) \stackrel{def}{=} \sum_{i=1}^{n} \left(\frac{\partial \mathbb{S}(\boldsymbol{\beta})}{\partial \beta_j} \right)^T \boldsymbol{X}_i \left\{ Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i) \right\} = 0, \quad j = 1, \ldots, d-1$$



(a) $n = 100$ **(b)** $n = 1,000$ **(c)** $n = 10,000$

The Simple Score Estimator is $\sqrt{n}$-consistent and asymptotically normally distributed

Our Simple Score Estimator (SSE) $\hat{\boldsymbol{\alpha}}_n$ is defined by,

$$\hat{\boldsymbol{\alpha}}_n \stackrel{def}{=} \mathbb{S}(\hat{\boldsymbol{\beta}}_n),$$

where $\hat{\boldsymbol{\beta}}_n$ is a zero-crossing of the function Z_n

$$\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0 = O_p(n^{-1/2})$$
$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) \to_d N_d\left(\mathbf{0}, \boldsymbol{A}^{-}\Sigma\boldsymbol{A}^{-}\right)$$

where $\boldsymbol{A}^{-}$ is the Moore-Penrose inverse of $\boldsymbol{A}$, where

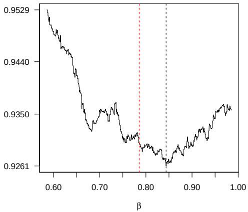
$$\boldsymbol{A} \stackrel{\text{def}}{=} E\Big[\psi_0'(\boldsymbol{\alpha}_0^T\boldsymbol{X})\,\mathrm{Cov}(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\Big],$$

$$\boldsymbol{\Sigma} \stackrel{\text{def}}{=} E\left[\left\{Y - \psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}^2\left\{\boldsymbol{X} - E(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}\left\{\boldsymbol{X} - E(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}^T\right]$$

Score estimators are compared to LSEs and Han's MRCEs

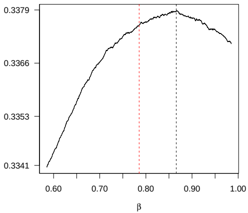
Least Squares Estimator (LSE):

$\hat{\boldsymbol{\alpha}}_n =$
$\arg\min \sum_{i=1}^{n} \left\{ Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{X}_i) \right\}^2$



Maximum Rank Correlation Estimator (MRCE):

$$\hat{\boldsymbol{\alpha}}_n = \arg\max \sum_{i \neq j} I_{\{Y_i > Y_j\}} I_{\{\alpha^T X_i > \alpha^T X_j\}}$$



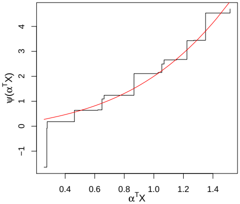
Simulations illustrate good finite sample behavior

Simulation model

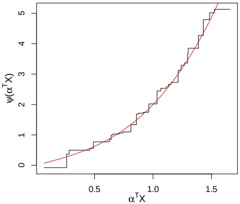
$$Y = \psi_0(\boldsymbol{\alpha}_0^T \boldsymbol{X}) + \varepsilon,$$

where

- $\psi_0(x) = x + x^3$
- $\alpha_{01} = \alpha_{02} = \alpha_{03} = 1/\sqrt{3}$
- $X_1, X_2, X_3 \overset{i.i.d}{\sim} U[0,1]$
- $\varepsilon \sim N(0,1)$
- $n = 100; 1,000; 10,000$



(a) $n = 100$



(b) $n = 1,000$

Score estimation is done in two steps

Step 1: Parametrization: Spherical coordinate system

$$\mathbb{S} : [0, 2\pi] \times [0, \pi] \mapsto \mathcal{S}_2 \subset \mathbb{R}^3 :$$
$$(\beta_1, \beta_2) \mapsto (\cos(\beta_1)\sin(\beta_2), \sin(\beta_1)\sin(\beta_2), \cos(\beta_2))$$

e.g. $\mathbb{S}(\pi/4, \arctan(\sqrt{2})) = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$

Step 2: Score equations:

$$\sum_{i=1}^{n} \left(\frac{\partial \mathbb{S}(\boldsymbol{\beta})}{\partial \beta_j} \right)^T \boldsymbol{X}_i \left\{ Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i) \right\} = 0.$$

$$\begin{cases} \sum_{i=1}^{n}(-\sin(\beta_1)\sin(\beta_2)X_{i1} + \cos(\beta_1)\sin(\beta_2)X_{i2})\{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{X}_i)\} = 0, \\ \sum_{i=1}^{n}(\cos(\beta_1)\cos(\beta_2)X_{i1} + \sin(\beta_1)\cos(\beta_2)X_{i2} - \sin(\beta_2)X_{i3}) \\ \qquad \{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{X}_i)\} = 0. \end{cases}$$

The average estimates converge to the true parameter values

Method	n	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
LSE	100	0.567232	0.566927	0.566318
	1,000	0.576605	0.575703	0.576451
	10,000	0.577146	0.577374	0.577146
	∞	0.577350	0.577350	0.577350
SSE	100	0.587614	0.541965	0.532872
	1,000	0.574333	0.576839	0.579007
	10,000	0.576838	0.577328	0.577704
	∞	0.577350	0.577350	0.577350
MRCE	100	0.567500	0.567568	0.568074
	1,000	0.576239	0.576586	0.576801
	10,000	0.577441	0.577291	0.577097
	∞	0.577350	0.577350	0.577350

The covariance matrix converges to the asymptotic variance

Method	n	$\hat{\sigma}_{11}$	$\hat{\sigma}_{22}$	$\hat{\sigma}_{33}$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{13}$	$\hat{\sigma}_{23}$	$d(\hat{\Sigma}, \Sigma)$
LSE	100	1.201293	1.203355	1.208558	-0.577588	-0.592399	-0.577087	-
	1,000	1.276503	1.271816	1.249995	-0.645524	-0.628822	-0.618845	-
	10,000	1.506928	1.473131	1.471213	-0.755098	-0.751517	-0.718229	-
	∞	-	-	-	-	-	-	
SSE	100	1.544919	1.772945	4.386728	-0.955064	-1.496068	0.644483	25.896840
	1,000	0.695695	0.753258	0.700203	-0.368076	-0.322984	-0.381043	2.976074
	10,000	0.679286	0.708114	0.700672	-0.342635	-0.335454	-0.365785	2.891139
	∞	0.692042	0.692042	0.692042	-0.346021	-0.346021	-0.346021	
MRCE	100	1.075928	1.137504	1.097447	-0.529870	-0.517154	-0.557786	5.485893
	1,000	0.926655	0.931801	0.938936	-0.458336	-0.465244	-0.472174	4.499525
	10,000	0.836557	0.856753	0.859607	-0.416975	-0.419253	-0.439996	4.048597
	∞	0.789576	0.789576	0.789576	-0.394788	-0.394788	-0.394788	

($\hat{\sigma}_{ij} = n \cdot \mathrm{cov}(\hat{\alpha}_{in}, \hat{\alpha}_{jn}), i, j = 1, 2, 3$, $\psi_0(x) = x + x^3$, $\alpha_{01} = \alpha_{02} = \alpha_{03} = 1/\sqrt{3}$, $\varepsilon \sim N(0, 1)$, $X_i \sim U[0, 1]$)

The covariance matrix for the LSE does not behave consistently

Consider again $\psi_0(x) = x + x^3$, $\alpha_{01} = \alpha_{02} = \alpha_{03} = 1/\sqrt{3}$, $\varepsilon \sim N(0,1)$ but generate $X_i \sim N(0,1)$.

Method	n	$\hat{\sigma}_{11}$	$\hat{\sigma}_{22}$	$\hat{\sigma}_{33}$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{13}$	$\hat{\sigma}_{23}$	$d(\hat{\Sigma}, \Sigma)$
LSE	100	0.117550	0.121344	0.121273	-0.057846	-0.059679	-0.061917	-
	1,000	0.081122	0.076944	0.076627	-0.040666	-0.040421	-0.036237	-
	10,000	0.072255	0.069788	0.073763	-0.034152	-0.038079	-0.035670	-
	∞	-	-	-	-	-	-	
SSE	100	0.113493	0.124740	0.110063	-0.062954	-0.051701	-0.058559	0.053908
	1,000	0.062691	0.062724	0.061552	-0.031817	-0.030782	-0.030871	0.017500
	10,000	0.047678	0.050129	0.048718	-0.024534	-0.023131	-0.025597	0.012451
	∞	0.041667	0.041667	0.041667	-0.020833	-0.020833	-0.020833	
MRCE	100	0.252922	0.249878	0.245900	-0.127500	-0.123730	-0.120269	0.256494
	1,000	0.176862	0.177429	0.178333	-0.087992	-0.089192	-0.089015	0.147942
	10,000	0.149451	0.153579	0.150804	-0.076088	-0.073316	-0.077506	0.119332
	∞	0.128981	0.128981	0.128981	-0.064491	-0.064491	-0.064491	

The Simple Score Estimator can be improved towards an Efficient Score Estimator (ESE)

Step 1: Differentiate the Sum of Squares:

Recall that the partial derivatives of the sum of squares w.r.t. $\beta_1, ...\beta_{d-1}$ are given by (for $j = 1, \dots, d-1$)

$$\frac{2}{n}\sum_{i=1}^{n}\left(\frac{\partial \mathbb{S}(\boldsymbol{\beta})}{\partial \beta_j}\right)^T \boldsymbol{X}_i \psi'(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i)\left\{Y_i - \psi(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i)\right\}$$

$\Rightarrow$ Estimate the derivative and construct an efficient score equation.

Step 2: Score equations:

$$\sum_{i=1}^{n}\left(\frac{\partial \mathbb{S}(\boldsymbol{\beta})}{\partial \beta_j}\right)^T \tilde{\psi}'_{n,\boldsymbol{\alpha}}(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i)\boldsymbol{X}_i\left\{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\mathbb{S}(\boldsymbol{\beta})^T \boldsymbol{X}_i)\right\} = 0.$$

The LSE $\hat{\psi}_{n,\alpha}$ is still used in the construction of the derivative estimate $\tilde{\psi}'_{nh,\alpha}$

Consider the smoothed LSE of ψ_0:

$$\tilde{\psi}_{nh,\boldsymbol{\alpha}}(u) = \int \mathbb{K}\left(\frac{u-x}{h}\right)\, d\hat{\psi}_{n,\boldsymbol{\alpha}}(x),$$

where $\mathbb{K}(x) = \int_{-\infty}^{x} K(x)dx$ and K is a classical kernel smoothing function.

Take its derivate:

$$\tilde{\psi}'_{nh,\boldsymbol{\alpha}}(u) = \frac{1}{h}\int K\left(\frac{u-x}{h}\right)\, d\hat{\psi}_{n,\boldsymbol{\alpha}}(x),$$

The Efficient Score Estimator (ESE) $\tilde{\boldsymbol{\alpha}}_n$ is given by $\tilde{\boldsymbol{\alpha}}_n \stackrel{def}{=} \mathbb{S}(\tilde{\boldsymbol{\beta}}_n)$, where $\tilde{\boldsymbol{\beta}}_n$ is a zero-crossing of

$$\sum_{i=1}^{n}\left(\frac{\partial \mathbb{S}\left(\boldsymbol{\beta}\right)}{\partial \beta_j}\right)^T \tilde{\psi}'_{nh,\boldsymbol{\alpha}}\big(\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{X}_i\big)\boldsymbol{X}_i\left\{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}\big(\mathbb{S}(\boldsymbol{\beta})^T\boldsymbol{X}_i\big)\right\}.$$

The ESE has better asymptotic performance

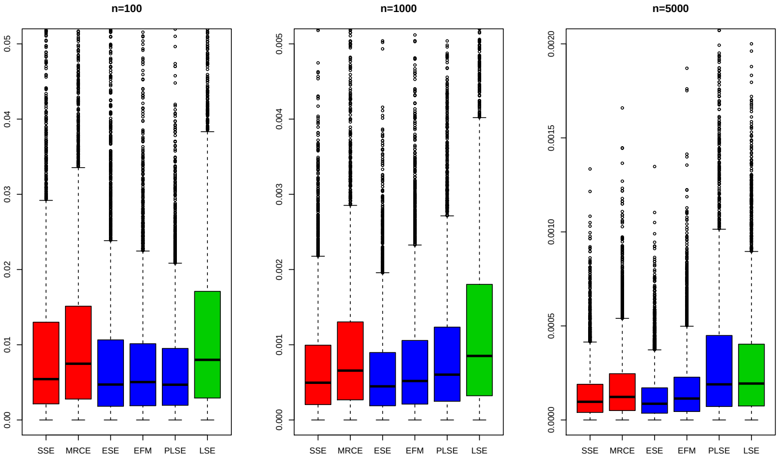
$$\tilde{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0 = O_p(n^{-1/2})$$
$$\sqrt{n}(\tilde{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) \to_d N_d\left(\mathbf{0}, \tilde{A}^{-}\tilde{\Sigma}\tilde{A}^{-}\right)$$

$\tilde{\alpha}_n$ attains a smaller asymptotic variance at the cost of extra smoothness conditions on ψ_0.

Asymptotic variances

	$X_i \sim U(0,1)$		$X_i \sim N(0,1)$	
Method	$n \cdot var(\hat{\boldsymbol{\alpha}}_{in})$	$n \cdot cov(\hat{\boldsymbol{\alpha}}_{in}, \hat{\boldsymbol{\alpha}}_{jn})$	$n \cdot var(\hat{\boldsymbol{\alpha}}_{in})$	$n \cdot var(\hat{\boldsymbol{\alpha}}_{in}, \hat{\boldsymbol{\alpha}}_{jn})$
ESE	0.617798	-0.308937	0.019608	-0.009804
SSE	0.692042	-0.346021	0.041667	-0.020833
MRCE	0.789576	-0.394788	0.128981	-0.064491

Efficient estimates have better asymptotic properties, at the cost of more complex estimation algorithms.



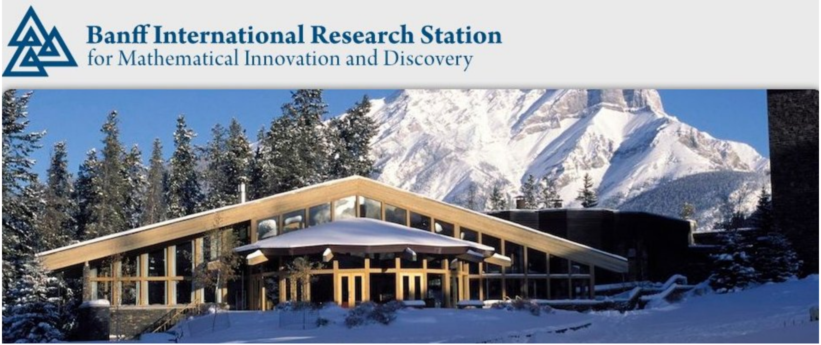
Boxplots of $\sum_{j=0}^{3}(\hat{\alpha}_j - \alpha_{0j})^2/3$ for $\sqrt{n}$-consistent but ineffcient methods (SSE and MRCE); $\sqrt{n}-$consistent and effcient methods (EFM (Cui et al., 2011) and PLSE (Kuchibhotla et al. 2017)) and a method with unknown limiting distribution (LSE).

A new method arises via score estimators based on the isotonic regression estimator in the monotone SIM.

The Simple Score Estimator

- depends on the non-smooth isotonic regression estimator
- has good asymptotic properties
  ($\sqrt{n}$ consistent, asymptotically normal)
- is computationally attractive due to its simplicity and lack of the need for smoothing methods
- can be improved via smoothing techniques resulting in efficient estimates

Thanks for your attendance!



kim.hendrickx@uhasselt.be

Balabdaoui, F., Groeneboom, P. and Hendrickx, K. (2017). Score estimation in the monotone single index model. *arXiv:1712.05593*.

Groeneboom, P. and Hendrickx, K. (2017). Estimation in monotone single index models. *Submitted for publication*.