Linear models and partial identification: Imprecise linear regression with interval data

The Problem

We have to estimate the parameter $\beta \in \mathbb{R}^{p+1}$

The Problem

We have to estimate the parameter $\beta \in \mathbb{R}^{p+1}$ of a linear model

The Problem

We have to estimate the parameter $\beta \in \mathbb{R}^{p+1}$ of a linear model

$$Y^* = X\beta + \varepsilon$$

The Problem

We have to estimate the parameter $\beta \in \mathbb{R}^{p+1}$ of a linear model

$$Y^* \quad = \quad X\beta + \varepsilon$$

with a fixed design-matrix

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ & & & \vdots & \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix},$$

The Problem

We have to estimate the parameter $\beta \in \mathbb{R}^{p+1}$ of a linear model

$$Y^* \quad = \quad X\beta + \varepsilon$$

with a fixed design-matrix

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ & & & \vdots & \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix},$$

a multivariat normal i.i.d. error ε

The Problem

We have to estimate the parameter $\beta \in \mathbb{R}^{p+1}$ of a linear model

$$Y^* \quad = \quad X\beta + \varepsilon$$

with a fixed design-matrix

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ & & & \vdots & \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix},$$

a multivariat normal i.i.d. error ε and a dependend n dimensional random variable $Y^* = (Y_1^*, \dots, Y_n^*)$,

The Problem

We have to estimate the parameter $\beta \in \mathbb{R}^{p+1}$ of a linear model

$$Y^* \quad = \quad X\beta + \varepsilon$$

with a fixed design-matrix

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ & & & \vdots & \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix},$$

a multivariat normal i.i.d. error ε and a dependend n dimensional random variable $Y^* = (Y_1^*, \dots, Y_n^*)$, that is only known to lie in the intervall $[\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$ of the known random variables $\underline{\mathrm{Y}}$ and $\overline{\mathrm{Y}}$.

A basic paradigm for solving the problem

One simple approach:

A basic paradigm for solving the problem

One simple approach:
look at all possible candidates Y for the unknown Y^* compatible with the restriction $Y \in [\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$

A basic paradigm for solving the problem

One simple approach:
look at all possible candidates Y for the unknown Y^* compatible with the restriction $Y \in [\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$ and estimate for all such candidates Y the corresponding estimates $\hat{\beta}$ to get the set-valued estimator

A basic paradigm for solving the problem

One simple approach:
look at all possible candidates Y for the unknown Y^* compatible with the restriction $Y \in [\underline{Y}, \overline{Y}]$ and estimate for all such candidates Y the corresponding estimates $\hat{\beta}$ to get the set-valued estimator

$$\hat{S} \quad := \quad \{\hat{\beta}(y) | y \in [\underline{y}, \overline{y}]\}$$

A basic paradigm for solving the problem

One simple approach:
look at all possible candidates Y for the unknown Y^* compatible with the restriction $Y \in [\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$ and estimate for all such candidates Y the corresponding estimates $\hat{\beta}$ to get the set-valued estimator

$$\hat{S} \quad := \quad \{\hat{\beta}(y) | y \in [\underline{\mathrm{y}}, \overline{\mathrm{y}}]\}$$

where $\underline{\mathrm{y}}$ and $\overline{\mathrm{y}}$ are n -dimensional samples from $\underline{\mathrm{Y}}$ and $\overline{\mathrm{Y}}$ and y is a n-dimensional vector satisfiying

$$y_i \in [\underline{\mathrm{Y}}_i, \ \overline{\mathrm{Y}}_i], i = 1, \ldots, n,$$

A basic paradigm for solving the problem

One simple approach:
look at all possible candidates Y for the unknown Y^* compatible with the restriction $Y \in [\underline{\mathrm{Y}}, \overline{\mathrm{Y}}]$ and estimate for all such candidates Y the corresponding estimates $\hat{\beta}$ to get the set-valued estimator

$$\hat{S} \quad := \quad \{\hat{\beta}(y) | y \in [\underline{\mathrm{y}}, \overline{\mathrm{y}}]\}$$

where $\underline{\mathrm{y}}$ and $\overline{\mathrm{y}}$ are n -dimensional samples from $\underline{\mathrm{Y}}$ and $\overline{\mathrm{Y}}$ and y is a n-dimensional vector satisfiying

$$y_i \in [\underline{\mathrm{Y}}_i, \ \overline{\mathrm{Y}}_i], i = 1, \ldots, n,$$

which stands for a possible sample of Y compatible with the interval-valued observed data.

One possibility:

One possibility:

Choose the classical linear estimator

One possibility:

Choose the classical linear estimator

$$\hat{\beta}(y) = (X'X)^{-1}X'y.$$

One possibility:

Choose the classical linear estimator

$$\hat{\beta}(y) = (X'X)^{-1}X'y.$$

to get an estimate $\hat{\beta}(y)$ for all y.

Linear Regression with one covariate:

Linear Regression with one covariate:

$$\hat{\beta}(y) \quad =$$

Linear Regression with one covariate:

$$\hat{\beta}(y) = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Linear Regression with one covariate:

$$\hat{\beta}(y) = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix}^{-1} \begin{pmatrix} \bar{y} \\ \overline{x \cdot y} \end{pmatrix}$$

Linear Regression with one covariate:

$$
\begin{aligned}
\hat{\beta}(y) &= \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix}^{-1} \begin{pmatrix} \bar{y} \\ \overline{x \cdot y} \end{pmatrix} \\
&= \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} y_i \\ \frac{1}{n} \sum_{i=1}^{n} x_i \cdot y_i \end{pmatrix}
\end{aligned}
$$

Linear Regression with one covariate:

$$\hat{\beta}(y) = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix}^{-1} \begin{pmatrix} \bar{y} \\ \overline{x \cdot y} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} y_i \\ \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i \end{pmatrix}$$

$$=: P \cdot \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} y_i \\ \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i \end{pmatrix}$$

Definition

The Minkowski Sum of two sets A, B *in* $\mathbb{R}^d$ *is defined as:*

$$A \oplus B := \{a + b | a \in A, b \in B\}.$$

Definition

The Minkowski Sum of two sets A, B *in* $\mathbb{R}^d$ *is defined as:*

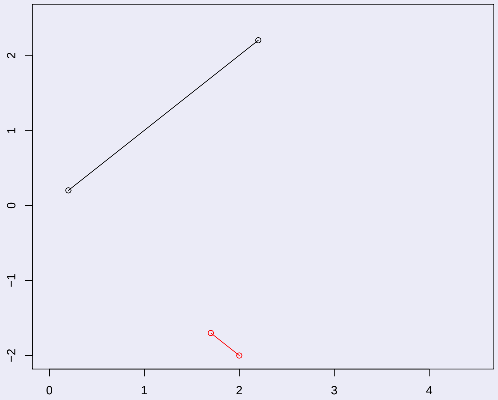
$$A \oplus B := \{a + b | a \in A, b \in B\}.$$

The Minkowski Mean of n *pointsets* $A_1, \ldots, A_n$ *is defined as:*

$$\frac{1}{n} \bigoplus_{i=1}^{n} A_i := \left\{ \frac{1}{n} \sum_{i=1}^{n} a_i \middle| a_i \in A_i, i = 1, \ldots, n \right\}$$

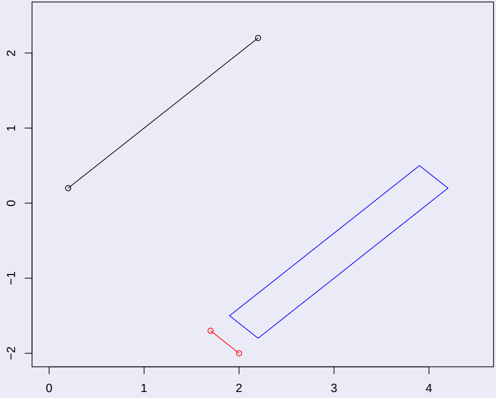
Example

The Minkowski Sum of two line segments in $\mathbb{R}^2$:



Example

The Minkowski Sum of two line segments in $\mathbb{R}^2$:



The calculation of $\hat{\beta}(y)$ for all $y \in [\underline{y}, \overline{y}]$ is nothing else than the computation of the linear image of the 2 dimensional minkowski mean of the n line segments p_i formed by the points $(\underline{y}_i, x_i \cdot \underline{y}_i)$ and $(\overline{y}_i, x_i \cdot \overline{y}_i)$ under the mapping induced by the matrix P:

$$\hat{S} = P \cdot \left(\frac{1}{n} \bigoplus_{i=1}^{n} p_i \right)$$

The set-valued estimate $\hat{S}$ has the following properties:

The set-valued estimate $\hat{S}$ has the following properties:

a) it is

The set-valued estimate $\hat{S}$ has the following properties:

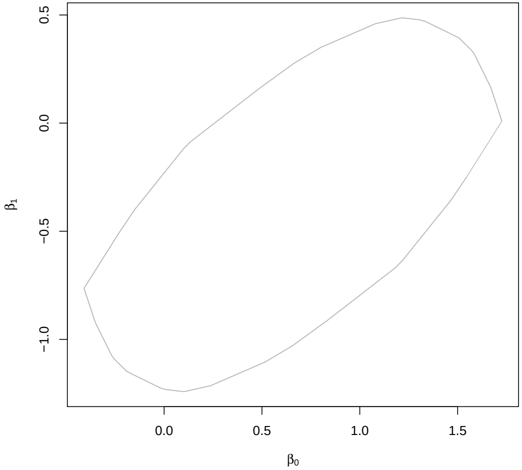
a) it is (, as the linear image of a convex, bounded set)

The set-valued estimate $\hat{S}$ has the following properties:

a) it is (, as the linear image of a convex, bounded set) convex and bounded.

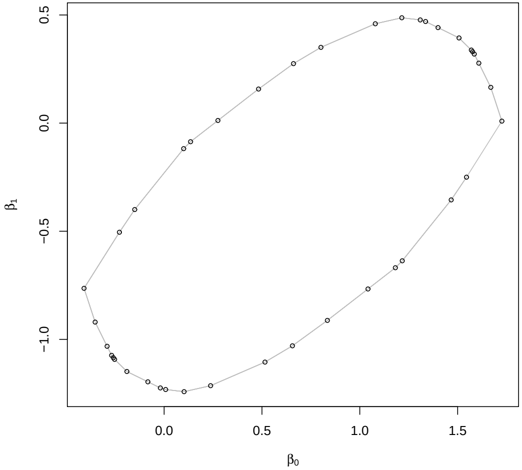
The set-valued estimate $\hat{S}$ has the following properties:

a) it is (, as the linear image of a convex, bounded set) convex and bounded.



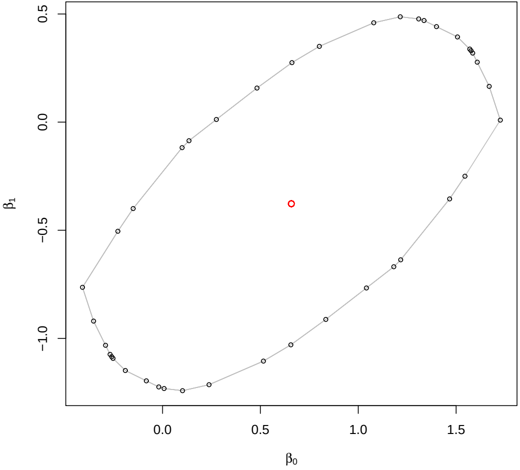
The set-valued estimate $\hat{S}$ has the following properties:

a) it has finite many extremepoints.



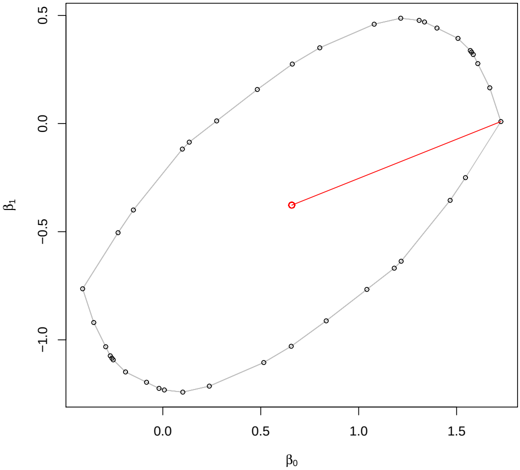
The set-valued estimate $\hat{S}$ has the following properties:

b) it is central symmetric with the center $\hat{\beta}(\frac{\underline{y}+\overline{y}}{2})$.



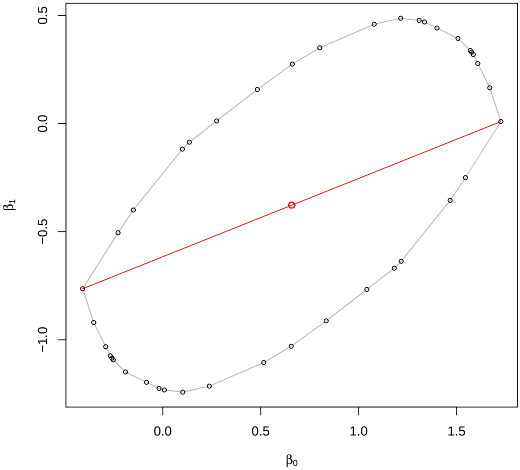
The set-valued estimate $\hat{S}$ has the following properties:

c) it is central symmetric with the center $\hat{\beta}(\frac{\underline{y}+\overline{y}}{2})$.



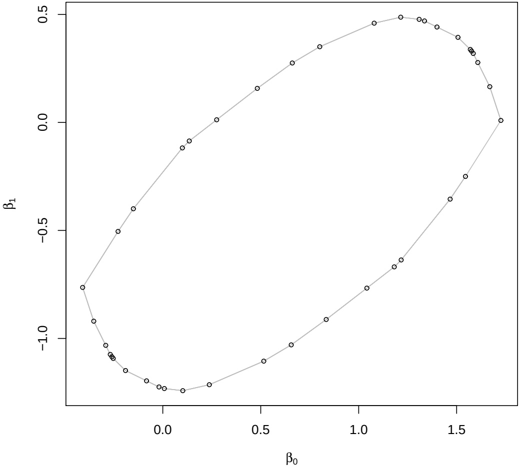
The set-valued estimate $\hat{S}$ has the following properties:

c) it is central symmetric with the center $\hat{\beta}(\frac{\underline{y}+\overline{y}}{2})$.



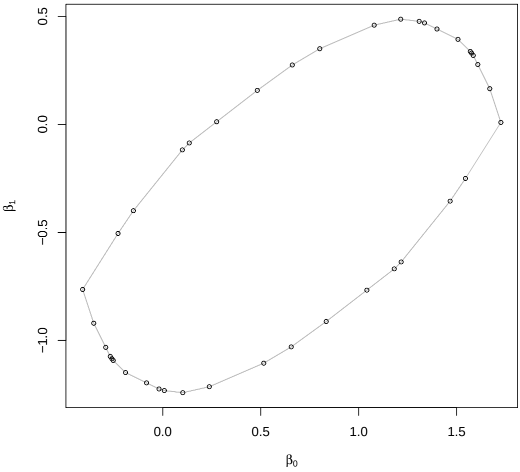
The set-valued estimate $\hat{S}$ has the following properties:

c) its facets are central symmetric, too.



The set-valued estimate $\hat{S}$ has the following properties:

c) in geometry it is, as the Minkowski Sum of n line segments, called a zonotope.



Another way to look at the estimate $\hat{S}$ (or a zonotope in general):

Another way to look at the estimate $\hat{S}$ (or a zonotope in general):

$$\hat{S} = \{A \cdot y | y \in [\underline{y}, \overline{y}]\}$$

Another way to look at the estimate $\hat{S}$ (or a zonotope in general):

$$\hat{S} = \{A \cdot y | y \in [\underline{y}, \overline{y}]\}$$

is the linear image of a n-dimensional cuboid

Another way to look at the estimate $\hat{S}$ (or a zonotope in general):

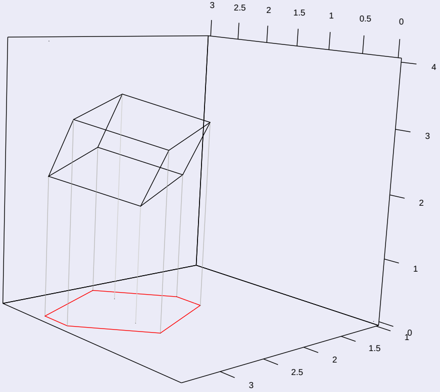
$$\hat{S} = \{A \cdot y | y \in [\underline{y}, \overline{y}]\}$$

is the linear image of a *n*-dimensional cuboid or equivalently the affine linear image of the *n*-dimensional unit-cube:

Another way to look at the estimate $\hat{S}$ (or a zonotope in general):

$$\hat{S} = \{A \cdot y | y \in [\underline{y}, \overline{y}]\}$$

is the linear image of a *n*-dimensional cuboid or equivalently the affine linear image of the *n*-dimensional unit-cube:



One can show:

One can show:

a) *The inverse images* y *of the extremepoints of* $\hat{S}$ *are extremepoints of the cuboid* $[\underline{y}, \overline{y}]$ *and have the following structure:*

One can show:

a) *The inverse images y of the extremepoints of $\hat{S}$ are extremepoints of the cuboid $[\underline{y}, \overline{y}]$ and have the following structure:*

$$y = y^u_{\geq c} = \begin{cases} \overline{y}_i & \text{if } x_i \geq c \\ \underline{y}_i & \text{else} \end{cases} \quad (1)$$

One can show:

a) *The inverse images y of the extremepoints of $\hat{S}$ are extremepoints of the cuboid $[\underline{y}, \overline{y}]$ and have the following structure:*

$$y = y^{u}_{\geq c} = \begin{cases} \overline{y}_i & \text{if } x_i \geq c \\ \underline{y}_i & \text{else} \end{cases} \quad (1)$$

or

$$y = y^{l}_{\geq c} = \begin{cases} \underline{y}_i & \text{if } x_i \geq c \\ \overline{y}_i & \text{else} \end{cases}, \quad (2)$$

for some $c \in \mathbb{R}$.

One can show:

a) *The inverse images y of the extremepoints of $\hat{S}$ are extremepoints of the cuboid $[\underline{\mathrm{y}}, \overline{\mathrm{y}}]$ and have the following structure:*

$$y = y^u_{\geq c} = \begin{cases} \overline{\mathrm{y}}_i & \text{if } x_i \geq c \\ \underline{\mathrm{y}}_i & \text{else} \end{cases} \tag{1}$$

or

$$y = y^l_{\geq c} = \begin{cases} \underline{\mathrm{y}}_i & \text{if } x_i \geq c \\ \overline{\mathrm{y}}_i & \text{else ,} \end{cases} \tag{2}$$

for some $c \in \mathbb{R}$. Here we call these y pseudodata.

One can show:

a) *The inverse images* y *of the extremepoints of* $\hat{S}$ *are extremepoints of the cuboid* $[\underline{y}, \overline{y}]$ *and have the following structure:*

$$y = y^u_{\geq c} = \begin{cases} \overline{y}_i & \text{if } x_i \geq c \\ \underline{y}_i & \text{else} \end{cases} \tag{1}$$

or

$$y = y^l_{\geq c} = \begin{cases} \underline{y}_i & \text{if } x_i \geq c \\ \overline{y}_i & \text{else}\ , \end{cases} \tag{2}$$

for some $c \in \mathbb{R}$. *Here we call these* y *pseudodata.*
It suffices to take only $c = x_i, i = 1, \ldots, n$.

b) *all pseudodata are mapped to the boundary of* $\hat{S}$.

b) *all pseudodata are mapped to the boundary of* $\hat{S}$*.*

c) *if there are no ties in* x*, then all pseudodata are actually mapped to extremepoints of* $\hat{S}$*.*

b) *all pseudodata are mapped to the boundary of* $\hat{S}$.

c) *if there are no ties in* x, *then all pseudodata are actually mapped to extremepoints of* $\hat{S}$.

$\Longrightarrow$ *it suffices to look at all pseudodata instead of the whole cuboid to observe* $\hat{S}$:

b) *all pseudodata are mapped to the boundary of* $\hat{S}$.

c) *if there are no ties in* x*, then all pseudodata are actually mapped to extremepoints of* $\hat{S}$.

$\Longrightarrow$ *it suffices to look at all pseudodata instead of the whole cuboid to observe* $\hat{S}$:

$$\hat{S} = \text{co}\left\{A \cdot y | y \text{ is a pseudodata } \right\}$$

We have two perspectives on $\hat{S}$

(1) $\hat{S}$ as the linear image of the minkowski mean of line segments, which could be also seen as the linear image of the minkowski mean of the set-valued data point $(p_1, \ldots, p_n)$:

We have two perspectives on $\hat{S}$

(1) $\hat{S}$ as the linear image of the minkowski mean of line segments, which could be also seen as the linear image of the minkowski mean of the set-valued data point $(p_1, \ldots, p_n)$:

$$\hat{S} = P \cdot \left(\frac{1}{n} \bigoplus_{i=1}^{n} p_i \right)$$

We have two perspectives on $\hat{S}$

(1) $\hat{S}$ as the linear image of the minkowski mean of line segments, which could be also seen as the linear image of the minkowski mean of the set-valued data point $(p_1, \ldots, p_n)$:

$$\hat{S} = P \cdot \left(\frac{1}{n} \bigoplus_{i=1}^{n} p_i \right)$$

Since $$\hat{S} = \left\{ P \cdot \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} y_i \\ \frac{1}{n} \sum_{i=1}^{n} x_i \cdot y_i \end{pmatrix} \middle| y \in [\underline{y}, \overline{y}] \right\}$$

We have two perspectives on $\hat{S}$

(1) $\hat{S}$ as the linear image of the minkowski mean of line segments, which could be also seen as the linear image of the minkowski mean of the set-valued data point $(p_1, \ldots, p_n)$:

$$\hat{S} = P \cdot \left(\frac{1}{n} \bigoplus_{i=1}^{n} p_i \right)$$

Since $$\hat{S} = \left\{ P \cdot \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} y_i \\ \frac{1}{n} \sum_{i=1}^{n} x_i \cdot y_i \end{pmatrix} \middle| \, y \in [\underline{y}, \overline{y}] \right\}$$

we could understand $\hat{S}$ as a point-estimator, which estimates the linear image of the (set-valued) so called Aumann Expectation

We have two perspectives on $\hat{S}$

(1) $\hat{S}$ as the linear image of the minkowski mean of line segments, which could be also seen as the linear image of the minkowski mean of the set-valued data point $(p_1, \ldots, p_n)$:

$$\hat{S} = P \cdot \left(\frac{1}{n} \bigoplus_{i=1}^{n} p_i \right)$$

Since $\hat{S} = \left\{ P \cdot \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} y_i \\ \frac{1}{n} \sum_{i=1}^{n} x_i \cdot y_i \end{pmatrix} \middle| \, y \in [\underline{y}, \overline{y}] \right\}$

we could understand $\hat{S}$ as a point-estimator, which estimates the linear image of the (set-valued) so called Aumann Expectation

$\left\{ \begin{pmatrix} \mathbb{E}(Y) \\ \mathbb{E}(X \cdot Y) \end{pmatrix} \middle| \, Y \in [\underline{Y}, \overline{Y}] \right\}$ under P

We have two perspectives on $\hat{S}$

(1) $\hat{S}$ as the linear image of the minkowski mean of line segments, which could be also seen as the linear image of the minkowski mean of the set-valued data point $(p_1, \ldots, p_n)$:

$$\hat{S} = P \cdot \left(\frac{1}{n} \bigoplus_{i=1}^{n} p_i \right)$$

Since $\hat{S} = \left\{ P \cdot \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} y_i \\ \frac{1}{n} \sum_{i=1}^{n} x_i \cdot y_i \end{pmatrix} \middle| y \in [\underline{y}, \overline{y}] \right\}$

we could understand $\hat{S}$ as a point-estimator, which estimates the linear image of the (set-valued) so called Aumann Expectation

$\left\{ \begin{pmatrix} \mathbb{E}(Y) \\ \mathbb{E}(X \cdot Y) \end{pmatrix} \middle| Y \in [\underline{Y}, \overline{Y}] \right\}$ under P (often called the sharp identification region).

So $\hat{S}$ could at first hand be seen as a (set-valued) pointestimator for a (set-valued) parameter (the Aumann Expectation under P). Here we can use random set theory to analyze the estimator.

(2) $\hat{S}$ as the collection of all precise pointestimators obtained by all possible data-completions $y \in [\underline{\mathrm{y}}, \overline{\mathrm{y}}]$.

This two views induces (at least) two ideas to construct confidence regions:

This two views induces (at least) two ideas to construct confidence regions:

(1) *analyze or estimate the distribution of the pointestimator* $\hat{S}$ *to construct a confidenceregion.*

This two views induces (at least) two ideas to construct confidence regions:

(1) *analyze or estimate the distribution of the pointestimator $\hat{S}$ to construct a confidenceregion. Since $\hat{S}$ is set-valued, we need a propper metric for the space of sets in $\mathbb{R}^d$:*

This two views induces (at least) two ideas to construct confidence regions:

(1) *analyze or estimate the distribution of the pointestimator $\hat{S}$ to construct a confidenceregion. Since $\hat{S}$ is set-valued, we need a propper metric for the space of sets in $\mathbb{R}^d$:*
one suggestion often quoted as natural:

This two views induces (at least) two ideas to construct confidence regions:

(1) *analyze or estimate the distribution of the pointestimator $\hat{S}$ to construct a confidenceregion. Since $\hat{S}$ is set-valued, we need a propper metric for the space of sets in $\mathbb{R}^d$:*
one suggestion often quoted as natural: the Hausdorff Distance:

This two views induces (at least) two ideas to construct confidence regions:

(1) *analyze or estimate the distribution of the pointestimator $\hat{S}$ to construct a confidenceregion. Since $\hat{S}$ is set-valued, we need a propper metric for the space of sets in $\mathbb{R}^d$:*
one suggestion often quoted as natural: the Hausdorff Distance:

$$H(A, B) := max\{dH(A, B), dH(B, A)\}$$

This two views induces (at least) two ideas to construct confidence regions:

(1) *analyze or estimate the distribution of the pointestimator $\hat{S}$ to construct a confidenceregion. Since $\hat{S}$ is set-valued, we need a propper metric for the space of sets in $\mathbb{R}^d$:*
one suggestion often quoted as natural: the Hausdorff Distance:

$$H(A, B) := max\{dH(A, B), dH(B, A)\}$$

with the directed Hausdorff Distance

This two views induces (at least) two ideas to construct confidence regions:

(1) *analyze or estimate the distribution of the pointestimator $\hat{S}$ to construct a confidenceregion. Since $\hat{S}$ is set-valued, we need a propper metric for the space of sets in $\mathbb{R}^d$:*
one suggestion often quoted as natural: the Hausdorff Distance:

$$H(A,B) := max\{dH(A,B), dH(B,A)\}$$

with the directed Hausdorff Distance

$$dH(A,B) := \sup_{a\in A} \inf_{b\in B} d(a,b)$$

This two views induces (at least) two ideas to construct confidence regions:

(1) *analyze or estimate the distribution of the pointestimator $\hat{S}$ to construct a confidenceregion. Since $\hat{S}$ is set-valued, we need a propper metric for the space of sets in $\mathbb{R}^d$:*
one suggestion often quoted as natural: the Hausdorff Distance:

$$H(A,B) := max\{dH(A,B), dH(B,A)\}$$

with the directed Hausdorff Distance

$$dH(A,B) := \sup_{a \in A} \inf_{b \in B} d(a,b)$$

and a metric d in $\mathbb{R}^d$ (e.g. the euclidean metric).

This approach is develepoed in Beresteanu, Molinari 2008*:*

There the authors estimate $\hat{S}$ *and draw bootstrap-samples from the data to estimate further* $\hat{S}^{\star}$ *and look on the distribution of* $dH(\hat{S}^{\star}, \hat{S})$.

There the authors estimate $\hat{S}$ and draw bootstrap-samples from the data to estimate further $\hat{S}^{\star}$ and look on the distribution of $dH(\hat{S}^{\star}, \hat{S})$. From this distribution they obtain a critical value c_{α} and construct the confidence collection

There the authors estimate $\hat{S}$ and draw bootstrap-samples from the data to estimate further $\hat{S}^\star$ and look on the distribution of $dH(\hat{S}^\star, \hat{S})$. From this distribution they obtain a critical value c_α and construct the confidence collection

$$HCR = \bigcup_{\substack{S \subset \mathbb{R}^d \\ dH(S,\hat{S}) \leq c_\alpha}} S.$$

There the authors estimate $\hat{S}$ and draw bootstrap-samples from the data to estimate further $\hat{S}^\star$ and look on the distribution of $dH(\hat{S}^\star, \hat{S})$. From this distribution they obtain a critical value c_α and construct the confidence collection

$$HCR = \bigcup_{\substack{S \subset \mathbb{R}^d \\ dH(S,\hat{S}) \leq c_\alpha}} S.$$

This confidenceregion asymptotically covers <u>the whole</u> sharp identification region with probability at least $1-\alpha$

There the authors estimate $\hat{S}$ and draw bootstrap-samples from the data to estimate further $\hat{S}^\star$ and look on the distribution of $dH(\hat{S}^\star, \hat{S})$. From this distribution they obtain a critical value c_α and construct the confidence collection

$$HCR = \bigcup_{\substack{S \subset \mathbb{R}^d \\ dH(S,\hat{S}) \leq c_\alpha}} S.$$

This confidenceregion asymptotically covers the whole sharp identification region with probability at least $1 - \alpha$ (under some regularity assumptions).

There the authors estimate $\hat{S}$ and draw bootstrap-samples from the data to estimate further $\hat{S}^\star$ and look on the distribution of $dH(\hat{S}^\star, \hat{S})$. From this distribution they obtain a critical value c_α and construct the confidence collection

$$HCR = \bigcup_{\substack{S \subset \mathbb{R}^d \\ dH(S,\hat{S}) \leq c_\alpha}} S.$$

This confidenceregion asymptotically covers the whole sharp identification region with probability at least $1-\alpha$ (under some regularity assumptions).

If one is in the situation, that there is a precise parameter β behind the scenes, it would be sufficient, that a confidenceregion covers not necessarily the whole sharp identification region but only the true parameter β with at least probability $1-\alpha$, which is a weaker demand. So in this situation HCR is a (conservative) confidenceregion for the true parameter β.

(2) collect the classical confidence-elipsoides for the classical least-squares-estimator and all possible data $y \in [\underline{y}, \overline{y}]$:

(2) collect the classical confidence-elipsoides for the classical least-squares-estimator and all possible data $y \in [\underline{y}, \overline{y}]$:

$$SCR := \bigcup_{y \in [\underline{y}, \overline{y}]} CE(y)$$

(2) collect the classical confidence-elipsoides for the classical least-squares-estimator and all possible data $y \in [\underline{y}, \overline{y}]$:

$$SCR := \bigcup_{y \in [\underline{y}, \overline{y}]} CE(y)$$

with the classical confidence-ellipsoides

(2) collect the classical confidence-elipsoides for the classical least-squares-estimator and all possible data $y \in [\underline{y}, \overline{y}]$:

$$SCR := \bigcup_{y \in [\underline{y}, \overline{y}]} CE(y)$$

with the classical confidence-ellipsoides

$$CE(y) := \left\{\beta \middle| (\beta - \hat{\beta}(y))'(X'X)(\beta - \hat{\beta}(y)) \leq (p+1) \cdot \hat{\sigma}^2(y) \cdot F_{1-\alpha}(p+1, n-p+1)\right\}.$$

But how do we compute this confidence region?

But how do we compute this confidence region?

Lemma

Let a partially identified linear model $y = \beta_0 + \beta_1 \cdot x + \varepsilon$ *be given.*

But how do we compute this confidence region?

Lemma

Let a partially identified linear model $y = \beta_0 + \beta_1 \cdot x + \varepsilon$ be given.
Under some not too strong conditions the simple confidenceregion SCR is a subset of the ellipsoid-type-confidenceregion

$$ECR := \text{co}\left(\bigcup_{c \in \{x_1, \ldots, x_n\}} CE(y^u_{\geq c}) \cup CE(y^l_{\geq c}) \right)$$

with arbitrary high probability $p < 1$, if $n = n(p)$ is large enough.

One „real-world-example“:

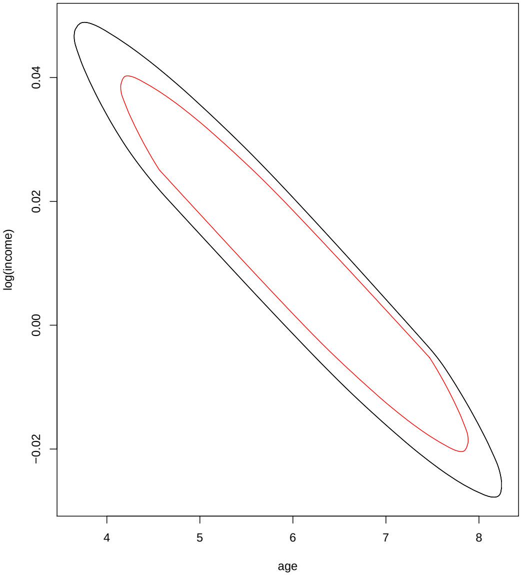
One „real-world-example“: Allbus data:

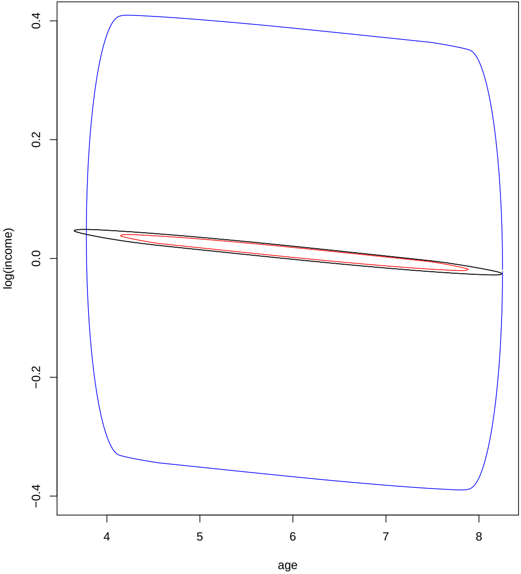
One „real-world-example“: Allbus data:

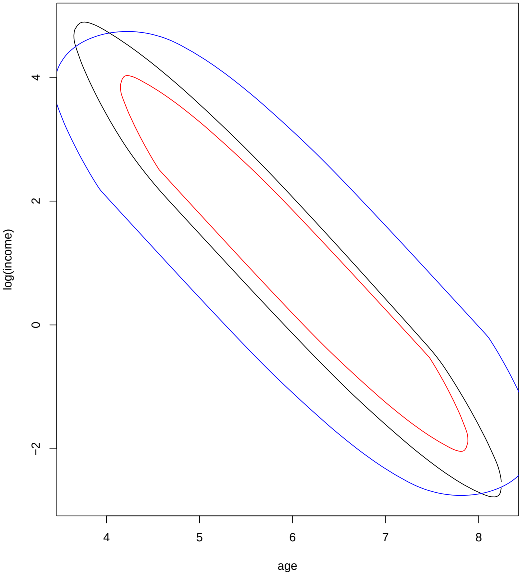
- sample from East Germany ($n = 1077$)

One „real-world-example“: Allbus data:

- sample from East Germany ($n = 1077$)
- age (x, precise) and logarithm of income (y, interval-valued)


0.04
0.02
0.00
−0.02
log(income)
4
5
6
7
8
age


0.4
0.2
0.0
−0.2
−0.4
log(income)
4
5
6
7
8
age


log(income)
4
2
0
−2
4
5
6
7
8
age

- Beresteanu, A., Molinari, F. (2008) Asymptotic Properties for a Class of Partially Identified Models, Econometrica, vol. 76, issue 4, pages 763-814.
- Schön, S., Kutterer, H. (2004) Using Zonotopes for Overestimation-Free Interval Least-Squares - Some Geodetic Applications. Reliable Computing, vol. 11, pages 137-155.
- Cerny, M., Rada M. (2011). On the Possibilistic Approach to Linear Regression with Rounded or Interval-Censored Data. Measurement Science Review, vol. 11 No. 2.

short simulationstudy:

short simulationstudy:

for 4 coarsening-processes:

short simulationstudy:

for 4 coarsening-processes:

- coarsening 1: $y = 10 \cdot x + 10 + \varepsilon, \quad \varepsilon \sim N(0,1)$
  $\underline{y} = y - \exp(\varepsilon_2), \quad \overline{y} = y + \exp(\varepsilon_3), \quad \varepsilon, \varepsilon_2, \varepsilon_3 : i.i.d., \sim N(0,1),$

short simulationstudy:

for 4 coarsening-processes:

- coarsening 1: $y = 10 \cdot x + 10 + \varepsilon, \quad \varepsilon \sim N(0,1)$
  $\underline{y} = y - \exp(\varepsilon_2), \quad \overline{y} = y + \exp(\varepsilon_3), \quad \varepsilon, \varepsilon_2, \varepsilon_3 : i.i.d., \sim N(0,1),$
- coarsening 2: $\underline{y} = min\{y, y_2\}, \quad \overline{y} = max\{y, y_2\}, \quad y_2 = 13 \cdot x + 9 + \varepsilon_2$

short simulationstudy:

for 4 coarsening-processes:

- coarsening 1: $y = 10 \cdot x + 10 + \varepsilon, \quad \varepsilon \sim N(0,1)$
  $\underline{y} = y - \exp(\varepsilon_2), \quad \bar{y} = y + \exp(\varepsilon_3), \quad \varepsilon, \varepsilon_2, \varepsilon_3 : i.i.d., \sim N(0,1),$
- coarsening 2: $\underline{y} = min\{y, y_2\}, \quad \bar{y} = max\{y, y_2\}, \quad y_2 = 13 \cdot x + 9 + \varepsilon_2$
- coarsening 3: $\underline{y} = y - \varepsilon_2^2 \cdot 10^{-5}, \quad \bar{y} = y + \varepsilon_3^2 \cdot 10^{-5} \cdot p, \quad p \sim B(n, 0.05)$

short simulationstudy:

for 4 coarsening-processes:

- coarsening 1: $y = 10 \cdot x + 10 + \varepsilon, \quad \varepsilon \sim N(0,1)$
  $\underline{y} = y - \exp(\varepsilon_2), \quad \overline{y} = y + \exp(\varepsilon_3), \quad \varepsilon, \varepsilon_2, \varepsilon_3 : i.i.d., \sim N(0,1),$
- coarsening 2: $\underline{y} = min\{y, y_2\}, \quad \overline{y} = max\{y, y_2\}, \quad y_2 = 13 \cdot x + 9 + \varepsilon_2$
- coarsening 3: $\underline{y} = y - \varepsilon_2^2 \cdot 10^{-5}, \quad \overline{y} = y + \varepsilon_3^2 \cdot 10^{-5} \cdot p, \quad p \sim B(n, 0.05)$
- coarsening 4: $\underline{y} = p \cdot y + (1-p) \cdot min\{-200, y\}$
  $\overline{y} = y + \varepsilon^2 \cdot q$
  $p \sim B(n, u_1), \quad u_1 \sim u[0,1]$
  $q \sim B(n, u_2^2), \quad u_2 \sim u[0,1]$

Covering Probabilities:

Covering Probabilities:

coarsening	N	SIR	HCR	ECR
1	10	0.96	1	1
1	100	1	1	1
1	1000	1	1	1
2	10	0.43	1	0.99
2	100	0.59	0.99	0.99
2	1000	0.80	1	1
3	10	0	0.93	1
3	100	0	0.92	0.95
3	1000	0	0.96	0.95
4	10	0.22	1	1
4	100	0.54	1	1
4	1000	0.82	1	1

Areas:

Areas:

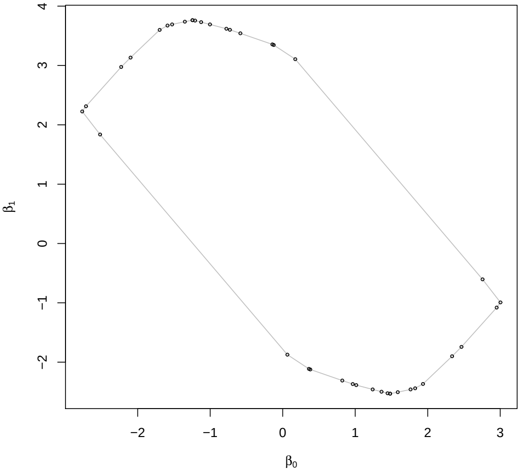
coarsening	N	SIR	HCR	ECR
1	10	7.18	102.33	55.40
1	100	6.22	14.31	13.07
1	1000	6.14	8.62	8.08
2	10	5.33	25.81	22.90
2	100	5.60	8.79	8.67
2	1000	5.62	6.57	6.51
3	10	$7 \cdot 10^{-11}$	3.97	3.37
3	100	$6.29 \cdot 10^{-11}$	0.19	0.19
3	1000	$6.39 \cdot 10^{-11}$	0.02	0.02
4	10	9.90	15848.89	10485.69
4	100	1.22	142.84	87.30
4	1000	0.31	1.48	1.25

An Idea of robustification

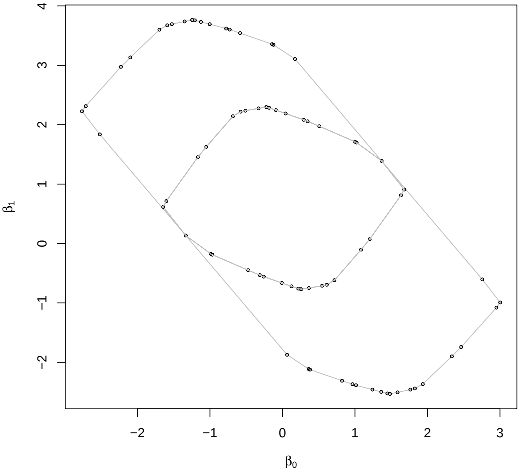
1) a bad idea:

1) a bad idea:

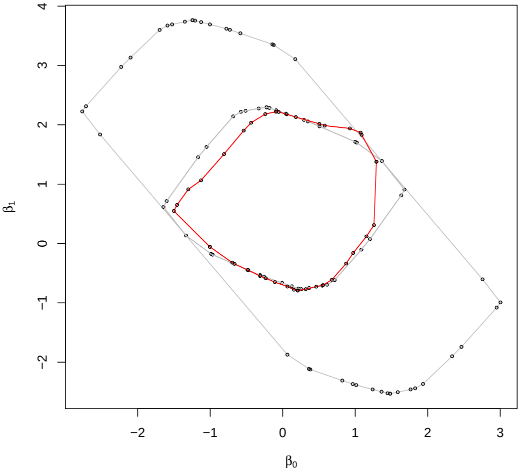
apply a robust method to all pseudodata.



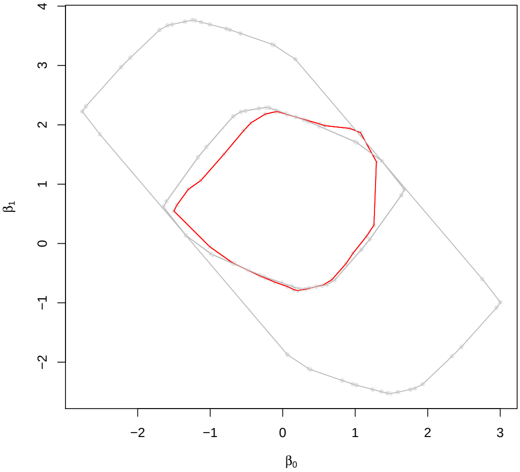
4
3
2
1
0
−1
−2
β_1
−2
−1
0
1
2
3
β_0



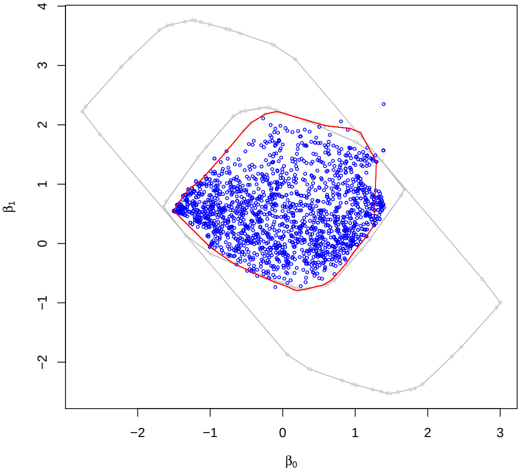
4
3
2
1
0
−1
−2
β_1
−2
−1
0
1
2
3
β_0



4
3
2
1
0
−1
−2
β_1
−2
−1
0
1
2
3
β_0



4
3
2
1
0
−1
−2
β_1
−2
−1
0
1
2
3
β_0



4
3
2
1
0
−1
−2
β_1
−2
−1
0
1
2
3
β_0

2) maybe a better idea:

2) maybe a better idea:
find for all pseudodata an appropriate (pseudo-)weightvector p.

2) maybe a better idea:
find for all pseudodata an appropriate (pseudo-)weightvector p.

- Calculate a global weight-vector g, that is acceptable in relation to all pseudoweightvectors

2) maybe a better idea:
find for all pseudodata an appropriate (pseudo-)weightvector p.

- Calculate a global weight-vector g, that is acceptable in relation to all pseudoweightvectors in the sense, that g lies between $\mathbb{1}$ and p for every pseudoweightvector p.

2) maybe a better idea:
find for all pseudodata an appropriate (pseudo-)weightvector p.

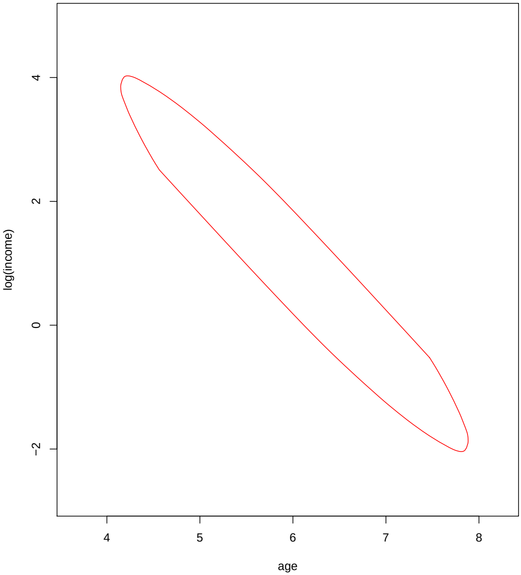
- Calculate a global weight-vector g, that is acceptable in relation to all pseudoweightvectors in the sense, that g lies between $\mathbb{1}$ and p for every pseudoweightvector p.
- Now use the weighted least-suqres-estimator with this weights.

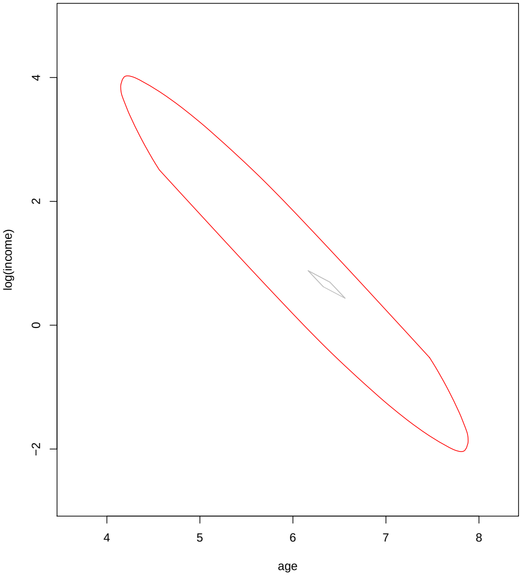
2) maybe a better idea:
find for all pseudodata an appropriate (pseudo-)weightvector p.

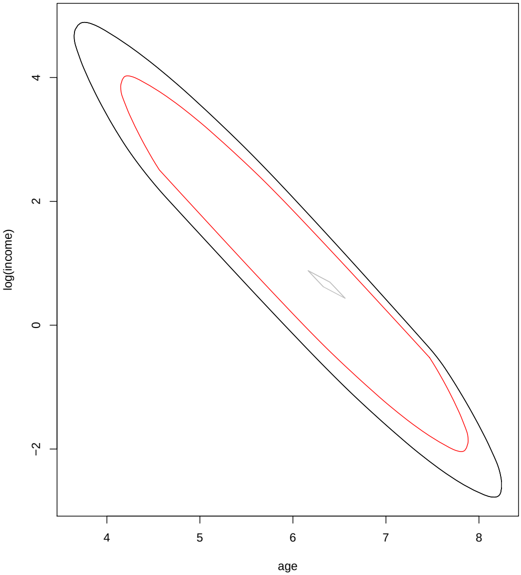
- Calculate a global weight-vector g, that is acceptable in relation to all pseudoweightvectors in the sense, that g lies between $\mathbb{1}$ and p for every pseudoweightvector p.
- Now use the weighted least-suqres-estimator with this weights.
All properties of the unweighted zonotope-estimator are kept.

2) maybe a better idea:
find for all pseudodata an appropriate (pseudo-)weightvector p.

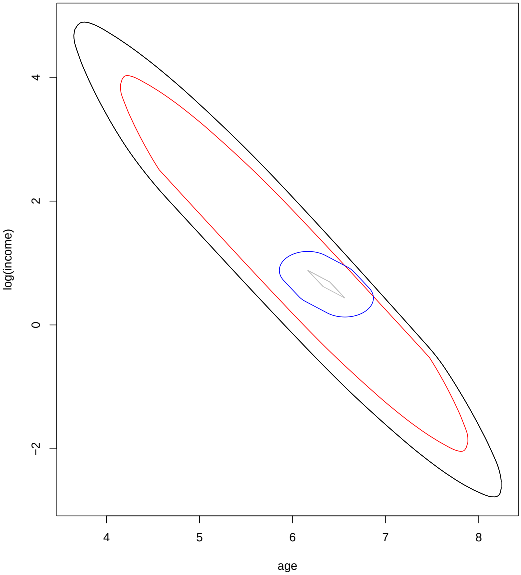
- Calculate a global weight-vector g, that is acceptable in relation to all pseudoweightvectors in the sense, that g lies between $\mathbb{1}$ and p for every pseudoweightvector p.
- Now use the weighted least-suqres-estimator with this weights.
  All properties of the unweighted zonotope-estimator are kept.
- For confidenceregions use the Hausdorff-based approach of Beresteanu and Molinari, but maybe with another d in the definition of the Hausdorff-distance.

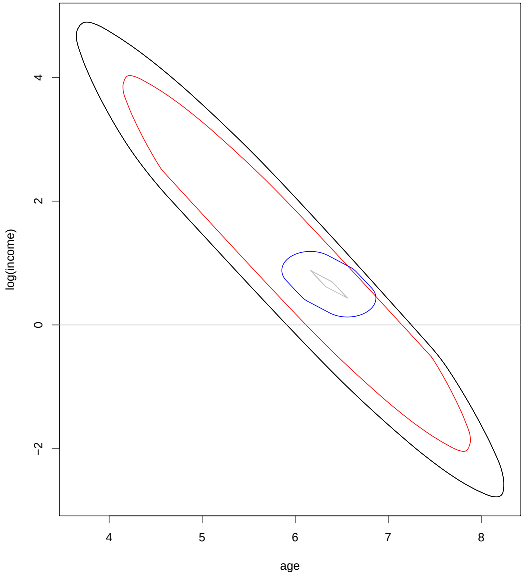

log(income)
4
2
0
−2
4
5
6
7
8
age


log(income)
4
2
0
−2
4
5
6
7
8
age



log(income)
4
2
0
−2
4
5
6
7
8
age


log(income)
4
2
0
−2
4
5
6
7
8
age



log(income)
4
2
0
−2
4
5
6
7
8
age

Covering Probabilities:

Covering Probabilities:

coarsening	N	SIR	HCR	ECR	GRHCR
1	10	0.96	1	1	1
1	100	1	1	1	1
1	1000	1	1	1	1
2	10	0.43	1	0.99	1
2	100	0.59	0.99	0.99	1
2	1000	0.80	1	1	1
3	10	0	0.93	1	1
3	100	0	0.92	0.95	0.95
3	1000	0	0.96	0.95	0.96
4	10	0.22	1	1	
4	100	0.54	1	1	1
4	1000	0.82	1	1	1

Areas:

Areas:

coarsening	N	SIR	HCR	ECR	GRHCR
1	10	7.18	119.55	32.522	120.38
1	100	6.22	14.31	13.07	13.59
1	1000	6.14	8.62	8.08	7.90
2	10	5.33	25.81	22.90	24.57
2	100	5.60	8.79	8.67	8.56
2	1000	5.62	6.57	6.51	6.24
3	10	$7 \cdot 10^{-11}$	3.97	3.37	3.99
3	100	$6.29 \cdot 10^{-11}$	0.19	0.19	0.2
3	1000	$6.39 \cdot 10^{-11}$	0.02	0.02	0.02
4	10	9.90	15848.89	10485.69	15994.15
4	100	1.22	142.84	87.30	110.02
4	1000	0.31	1.48	1.25	

Beresteanu, A., Molinari, F. (2008) Asymptotic Properties for a Class of Partially Identified Models, Econometrica, vol. 76, issue 4, pages 763-814.

Schön, S., Kutterer, H. (2004) Using Zonotopes for Overestimation-Free Interval Least-Squares - Some Geodetic Applications. Reliable Computing, vol. 11, pages 137-155.

Cerny, M., Rada M. (2011). On the Possibilistic Approach to Linear Regression with Rounded or Interval-Censored Data. Measurement Science Review, vol. 11 No. 2.