An exact algorithm for Likelihood-based Imprecise Regression in the case of simple linear regression with interval data

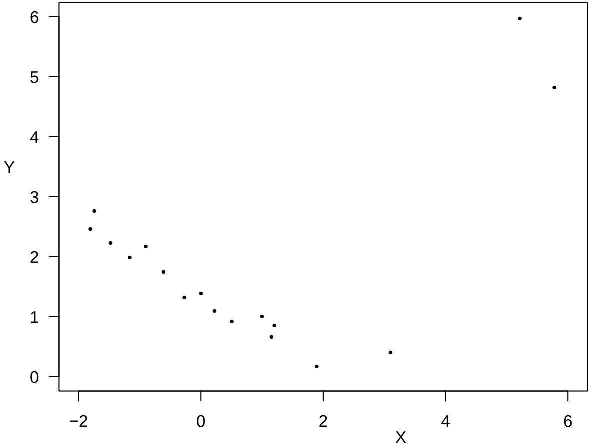
Andrea Wiencierz and Marco E. G. V. Cattaneo
Department of Statistics, LMU Munich

SMPS 6, Konstanz, Germany
October 4, 2012

Likelihood-based Imprecise Regression (LIR)

Likelihood-based Imprecise Regression (LIR)

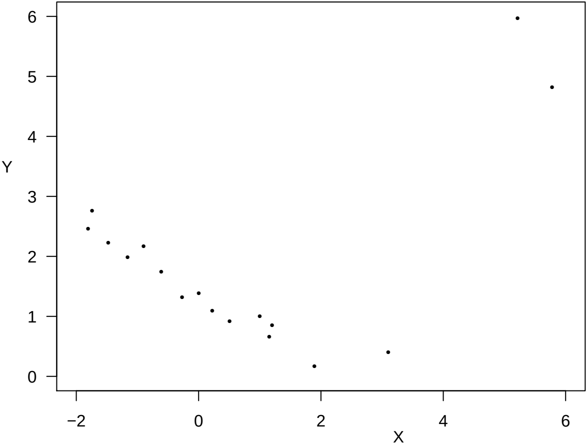
- $(X_1, Y_1), \ldots, (X_n, Y_n)$
  with $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} P$



Likelihood-based Imprecise Regression (LIR)

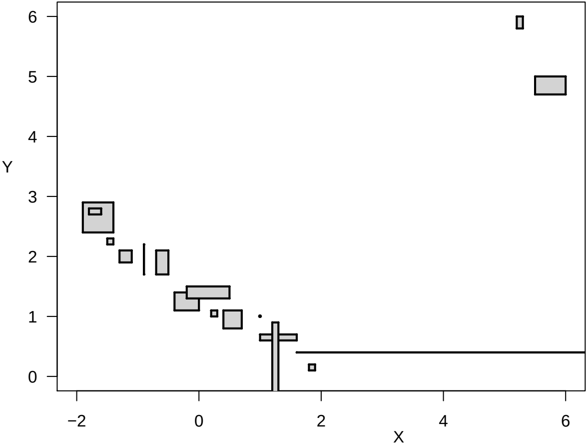
- $(X_1, Y_1), \ldots, (X_n, Y_n)$ with $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} P$
- simple linear regression:

  $Y = f(X) = a + b\,X$



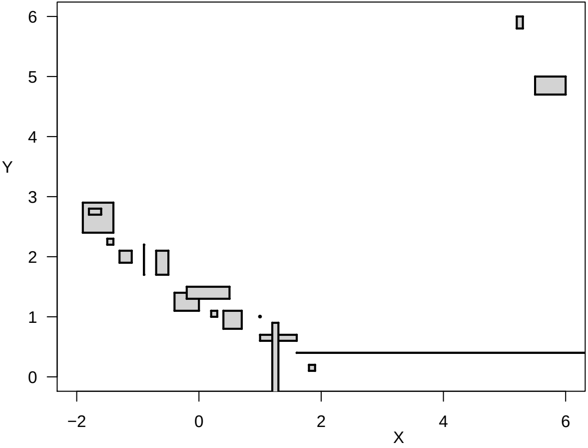
(Simple) linear LIR with interval data

- $(X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*)$
  where $X_i^* = [\underline{X}_i, \overline{X}_i]$
  and $Y_i^* = [\underline{Y}_i, \overline{Y}_i]$



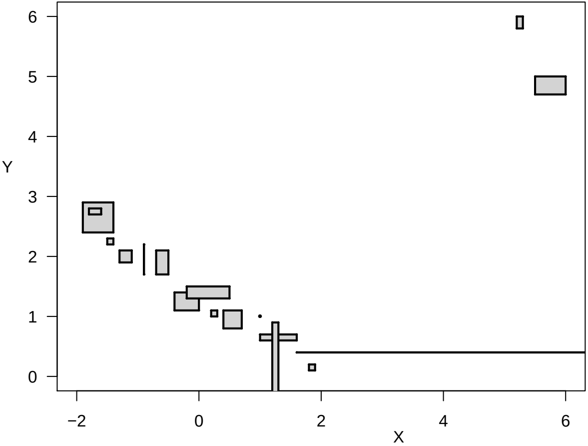
(Simple) linear LIR with interval data

- $(X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*)$
  where $X_i^* = [\underline{X}_i, \overline{X}_i]$
  and $Y_i^* = [\underline{Y}_i, \overline{Y}_i]$
- with $V_i^* = X_i^* \times Y_i^*$
  $((X_i, Y_i), V_i^*) \overset{\text{i.i.d.}}{\sim} P$
  such that for $\varepsilon \in [0, 1]$
  $P((X_i, Y_i) \notin V_i^*) \leq \varepsilon$



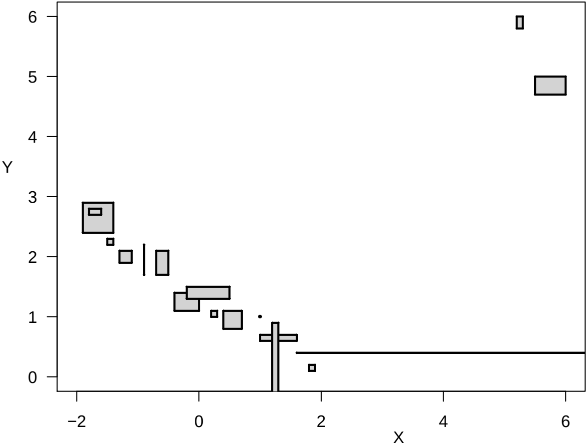
(Simple) linear LIR with interval data

- $(X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*)$
  where $X_i^* = [\underline{X}_i, \overline{X}_i]$
  and $Y_i^* = [\underline{Y}_i, \overline{Y}_i]$
- with $V_i^* = X_i^* \times Y_i^*$
  $((X_i, Y_i), V_i^*) \overset{\text{i.i.d.}}{\sim} P$
  such that for $\varepsilon \in [0, 1]$
  $P((X_i, Y_i) \notin V_i^*) \leq \varepsilon$
- simple linear regression:
  $Y = f(X) = a + b\,X$



(Simple) linear LIR with interval data

- $(X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*)$ where $X_i^* = [\underline{X}_i, \overline{X}_i]$ and $Y_i^* = [\underline{Y}_i, \overline{Y}_i]$
- with $V_i^* = X_i^* \times Y_i^*$ $((X_i, Y_i), V_i^*) \overset{\text{i.i.d.}}{\sim} P$ such that for $\varepsilon \in [0, 1]$ $P((X_i, Y_i) \notin V_i^*) \leq \varepsilon$
- simple linear regression: $Y = f(X) = a + b\,X$
- p-quantile $Q_{R_f,p}$, with $p \in (0, 1)$, of the distribution of the residuals $R_{f,i} = |Y_i - f(X_i)|$

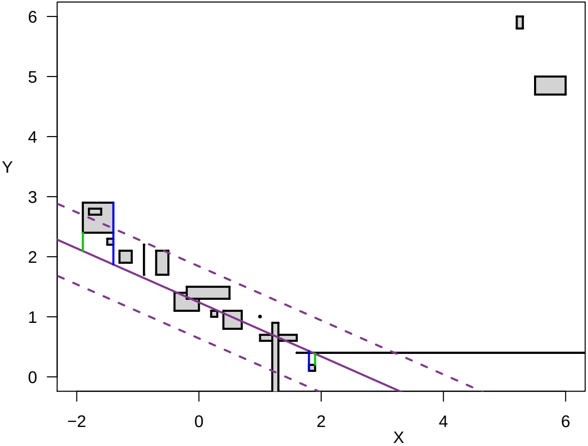


(Simple) linear LIR with interval data

- imprecise residuals:

$$\underline{r}_{f,i} = \min_{(x,y) \in v_i^*} |y - f(x)|$$

$$\overline{r}_{f,i} = \sup_{(x,y) \in v_i^*} |y - f(x)|$$



(Simple) linear LIR with interval data

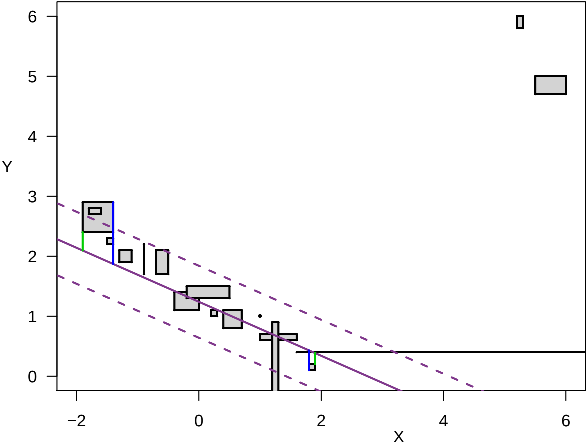
- imprecise residuals:

$$\underline{r}_{f,i} = \min_{(x,y)\in v_i^*} |y - f(x)|$$

$$\overline{r}_{f,i} = \sup_{(x,y)\in v_i^*} |y - f(x)|$$

- uncertainty about f:

  data imprecision and statistical uncertainty



(Simple) linear LIR with interval data

- imprecise residuals:

$$\underline{r}_{f,i} = \min_{(x,y) \in v_i^*} |y - f(x)|$$

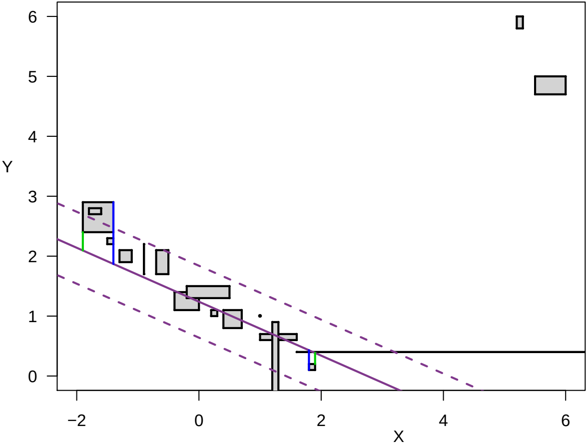
$$\overline{r}_{f,i} = \sup_{(x,y) \in v_i^*} |y - f(x)|$$

- uncertainty about f:

  data imprecision and statistical uncertainty

- consider $\mathcal{C}_{f,p,\beta,\varepsilon}$:

  likelihood-based confidence region for $Q_{R_f,p}$ with cutoff point $\beta \in (0,1)$



(Simple) linear LIR with interval data

- imprecise residuals:

$$\underline{r}_{f,i} = \min_{(x,y)\in v_i^*} |y - f(x)|$$

$$\overline{r}_{f,i} = \sup_{(x,y)\in v_i^*} |y - f(x)|$$

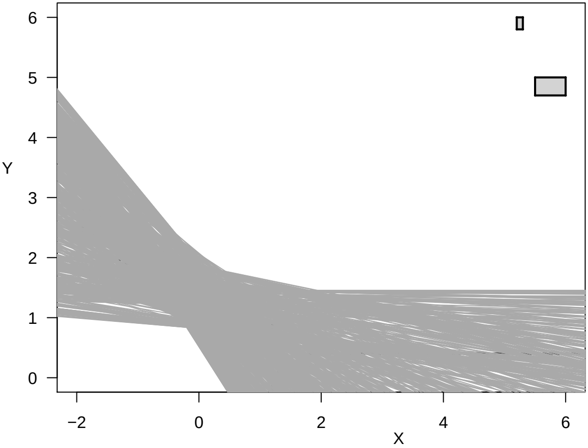
- uncertainty about f:

  data imprecision and statistical uncertainty

- consider $\mathcal{C}_{f,p,\beta,\varepsilon}$:

  likelihood-based confidence region for $Q_{R_f,p}$ with cutoff point $\beta \in (0,1)$

- result $\mathcal{U}$: set of all plausible functions



Recapitulation: (simple) linear LIR with interval data

Recapitulation: (simple) linear LIR with interval data

- $((X_i, Y_i), V_i^*) \stackrel{\text{i.i.d.}}{\sim} P\,,\ P \in \mathcal{P}_\varepsilon = \{P : P((X_i, Y_i) \notin V_i^*) \leq \varepsilon\}\ ,\ \varepsilon \in [0, 1]$

Recapitulation: (simple) linear LIR with interval data

- $((X_i, Y_i), V_i^*) \overset{\text{i.i.d.}}{\sim} P\,,\ P \in \mathcal{P}_\varepsilon = \{P : P((X_i, Y_i) \notin V_i^*) \leq \varepsilon\}\ ,\ \varepsilon \in [0,1]$
- $Y_i = f(X_i)\,,\ f \in \mathcal{F} = \left\{ f_{a,b} : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ X & \mapsto & a + b\,X \end{array}\ ,\ a, b \in \mathbb{R} \right\}$

Recapitulation: (simple) linear LIR with interval data

- $((X_i, Y_i), V_i^*) \overset{\text{i.i.d.}}{\sim} P\,,\ P \in \mathcal{P}_\varepsilon = \{P : P((X_i, Y_i) \notin V_i^*) \leq \varepsilon\}\ ,\ \varepsilon \in [0, 1]$
- $Y_i = f(X_i)\,,\ f \in \mathcal{F} = \left\{ f_{a,b} : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ X & \mapsto & a + b\,X \end{array}\ ,\ a, b \in \mathbb{R} \right\}$
- observations $v_1^*, \dots, v_n^*$ induce (normalized) profile likelihood function $lik_{Q_{R_f}}$ of the p-quantile of the distribution of R_f for each $f \in \mathcal{F}$

Recapitulation: (simple) linear LIR with interval data

- $((X_i, Y_i), V_i^*) \stackrel{\text{i.i.d.}}{\sim} P\,,\ P \in \mathcal{P}_\varepsilon = \{P : P((X_i, Y_i) \notin V_i^*) \leq \varepsilon\}\ ,\ \varepsilon \in [0,1]$
- $Y_i = f(X_i)\,,\ f \in \mathcal{F} = \left\{ f_{a,b} : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ X & \mapsto & a + b\,X \end{array}\ ,\ a, b \in \mathbb{R} \right\}$
- observations $v_1^*, \ldots, v_n^*$ induce (normalized) profile likelihood function $lik_{Q_{R_f}}$ of the p-quantile of the distribution of R_f for each $f \in \mathcal{F}$
- $lik_{Q_{R_f}}$ is a stepwise constant function with points of discontinuity at:

  $0 = \underline{r}_{f,(0)}, \ldots, \underline{r}_{f,(\lceil n(p-\varepsilon)\rceil)}, \overline{r}_{f,(\lfloor n(p+\varepsilon)\rfloor+1)}, \ldots, \overline{r}_{f,(n+1)} = +\infty$

Recapitulation: (simple) linear LIR with interval data

- $((X_i, Y_i), V_i^*) \overset{\text{i.i.d.}}{\sim} P\,, \; P \in \mathcal{P}_\varepsilon = \{P : P((X_i, Y_i) \notin V_i^*) \leq \varepsilon\}\,, \; \varepsilon \in [0, 1]$
- $Y_i = f(X_i)\,, \; f \in \mathcal{F} = \left\{ f_{a,b} : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ X & \mapsto & a + b\,X \end{array}, \; a, b \in \mathbb{R} \right\}$
- observations $v_1^*, \ldots, v_n^*$ induce (normalized) profile likelihood function $lik_{Q_{R_f}}$ of the p-quantile of the distribution of R_f for each $f \in \mathcal{F}$
- $lik_{Q_{R_f}}$ is a stepwise constant function with points of discontinuity at:

  $0 = \underline{r}_{f,(0)}, \ldots, \underline{r}_{f,(\lceil n(p-\varepsilon) \rceil)}, \overline{r}_{f,(\lfloor n(p+\varepsilon) \rfloor + 1)}, \ldots, \overline{r}_{f,(n+1)} = +\infty$
- $\mathcal{C}_f = [\underline{r}_{f,(\underline{k}+1)}, \overline{r}_{f,(\overline{k})}]\,,$ values of $\underline{k}, \overline{k} \in \mathbb{N} \cup \{0\}$ depend on n, p, β, ε

Recapitulation: (simple) linear LIR with interval data

- $((X_i, Y_i), V_i^*) \overset{\text{i.i.d.}}{\sim} P\,,\; P \in \mathcal{P}_\varepsilon = \{P : P((X_i, Y_i) \notin V_i^*) \leq \varepsilon\}\,,\; \varepsilon \in [0, 1]$
- $Y_i = f(X_i)\,,\; f \in \mathcal{F} = \left\{ f_{a,b} : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ X & \mapsto & a + b\,X \end{array}\,,\; a, b \in \mathbb{R} \right\}$
- observations $v_1^*, \ldots, v_n^*$ induce (normalized) profile likelihood function $lik_{Q_{R_f}}$ of the p-quantile of the distribution of R_f for each $f \in \mathcal{F}$
- $lik_{Q_{R_f}}$ is a stepwise constant function with points of discontinuity at:

  $0 = \underline{r}_{f,(0)}, \ldots, \underline{r}_{f,(\lceil n(p-\varepsilon) \rceil)}, \overline{r}_{f,(\lfloor n(p+\varepsilon) \rfloor + 1)}, \ldots, \overline{r}_{f,(n+1)} = +\infty$
- $\mathcal{C}_f = [\underline{r}_{f,(\underline{k}+1)}, \overline{r}_{f,(\overline{k})}]\,,$ values of $\underline{k}, \overline{k} \in \mathbb{N} \cup \{0\}$ depend on n, p, β, ε
- LIR result $\mathcal{U} = \{f \in \mathcal{F} : \underline{r}_{f,(\underline{k}+1)} \leq \overline{q}_{LRM}\}\,,$ where $\overline{q}_{LRM} = \inf_{f \in \mathcal{F}} \overline{r}_{f,(\overline{k})}$

Recapitulation: (simple) linear LIR with interval data

- $((X_i, Y_i), V_i^*) \overset{\text{i.i.d.}}{\sim} P\,,\ P \in \mathcal{P}_\varepsilon = \{P : P((X_i, Y_i) \notin V_i^*) \leq \varepsilon\}\ ,\ \varepsilon \in [0, 1]$
- $Y_i = f(X_i)\,,\ f \in \mathcal{F} = \left\{ f_{a,b} : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ X & \mapsto & a + b\,X \end{array}\ ,\ a, b \in \mathbb{R} \right\}$
- observations $v_1^*, \ldots, v_n^*$ induce (normalized) profile likelihood function $lik_{Q_{R_f}}$ of the p-quantile of the distribution of R_f for each $f \in \mathcal{F}$
- $lik_{Q_{R_f}}$ is a stepwise constant function with points of discontinuity at:

  $0 = \underline{r}_{f,(0)}, \ldots, \underline{r}_{f,(\lceil n(p-\varepsilon)\rceil)}, \overline{r}_{f,(\lfloor n(p+\varepsilon)\rfloor+1)}, \ldots, \overline{r}_{f,(n+1)} = +\infty$
- $\mathcal{C}_f = [\underline{r}_{f,(\underline{k}+1)}, \overline{r}_{f,(\overline{k})}]\,,$ values of $\underline{k}, \overline{k} \in \mathbb{N} \cup \{0\}$ depend on n, p, β, ε
- LIR result $\mathcal{U} = \{f \in \mathcal{F} : \underline{r}_{f,(\underline{k}+1)} \leq \overline{q}_{LRM}\}\,,$ where $\overline{q}_{LRM} = \inf_{f \in \mathcal{F}} \overline{r}_{f,(\overline{k})}$
- if there is a unique f with $\overline{r}_{f,(\overline{k})} = \overline{q}_{LRM}$, it is optimal according to the Likelihood-based Region Minimax (LRM) criterion and called f_{LRM}

Recapitulation: (simple) linear LIR with interval data

- $((X_i, Y_i), V_i^*) \overset{\text{i.i.d.}}{\sim} P\,,\; P \in \mathcal{P}_\varepsilon = \{P : P((X_i, Y_i) \notin V_i^*) \leq \varepsilon\}\,,\; \varepsilon \in [0, 1]$
- $Y_i = f(X_i)\,,\; f \in \mathcal{F} = \left\{ f_{a,b} : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ X & \mapsto & a + b\,X \end{array}\,,\; a, b \in \mathbb{R} \right\}$
- observations $v_1^*, \ldots, v_n^*$ induce (normalized) profile likelihood function $lik_{Q_{R_f}}$ of the p-quantile of the distribution of R_f for each $f \in \mathcal{F}$
- $lik_{Q_{R_f}}$ is a stepwise constant function with points of discontinuity at:

  $0 = \underline{r}_{f,(0)}, \ldots, \underline{r}_{f,(\lceil n(p-\varepsilon) \rceil)}, \overline{r}_{f,(\lfloor n(p+\varepsilon) \rfloor + 1)}, \ldots, \overline{r}_{f,(n+1)} = +\infty$
- $\mathcal{C}_f = [\underline{r}_{f,(\underline{k}+1)}, \overline{r}_{f,(\overline{k})}]\,,$ values of $\underline{k}, \overline{k} \in \mathbb{N} \cup \{0\}$ depend on n, p, β, ε
- LIR result $\mathcal{U} = \{f \in \mathcal{F} : \underline{r}_{f,(\underline{k}+1)} \leq \overline{q}_{LRM}\}\,,$ where $\overline{q}_{LRM} = \inf_{f \in \mathcal{F}} \overline{r}_{f,(\overline{k})}$
- if there is a unique f with $\overline{r}_{f,(\overline{k})} = \overline{q}_{LRM}$, it is optimal according to the Likelihood-based Region Minimax (LRM) criterion and called f_{LRM}
- further details in: M. Cattaneo, A. Wiencierz (2012). *Likelihood-based Imprecise Regression*. Int. J. Approx. Reasoning 53. 1137-1154.

Implementation: Exact algorithm for simple linear LIR

Implementation: Exact algorithm for simple linear LIR

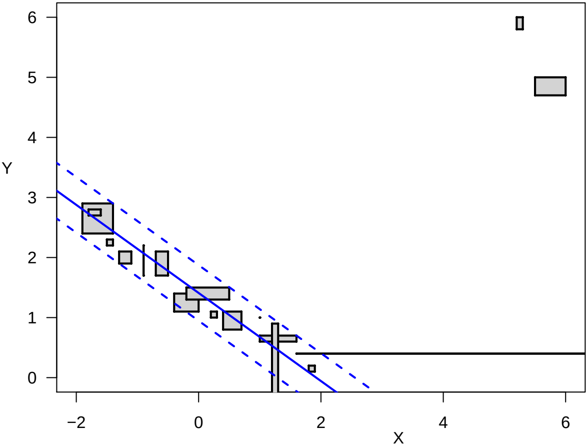
- aim: determine the set of all undominated functions $\mathcal{U} = \{f \in \mathcal{F} : \underline{r}_{f,(\underline{k}+1)} \leq \overline{q}_{LRM}\}$

Implementation: Exact algorithm for simple linear LIR

- aim: determine the set of all undominated functions $\mathcal{U} = \{f \in \mathcal{F} : \underline{r}_{f,(\underline{k}+1)} \leq \overline{q}_{LRM}\}$
- 1st step: find $\overline{q}_{LRM}$

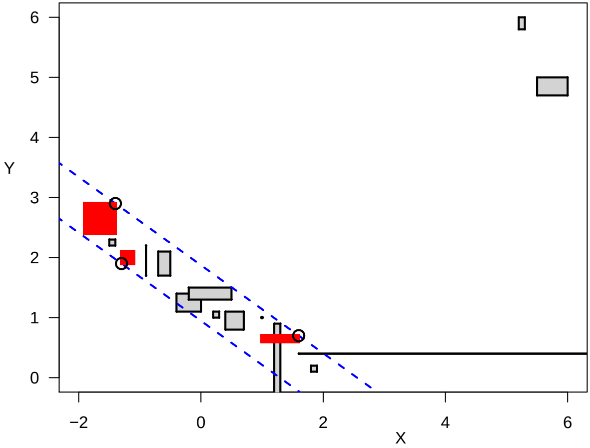
Implementation: Exact algorithm for simple linear LIR

- aim: determine the set of all undominated functions $\mathcal{U} = \{f \in \mathcal{F} : \underline{r}_{f,(\underline{k}+1)} \leq \overline{q}_{LRM}\}$
- 1st step: find $\overline{q}_{LRM}$
- $\overline{B}_{f_{LRM},\overline{q}_{LRM}}$ (blue dashed lines) is the thinnest band containing at least $\overline{k}$ imprecise data
- here $\beta = 0.8, p = 0.6,$ $n = 17\,,$ and $\overline{k} = 12$



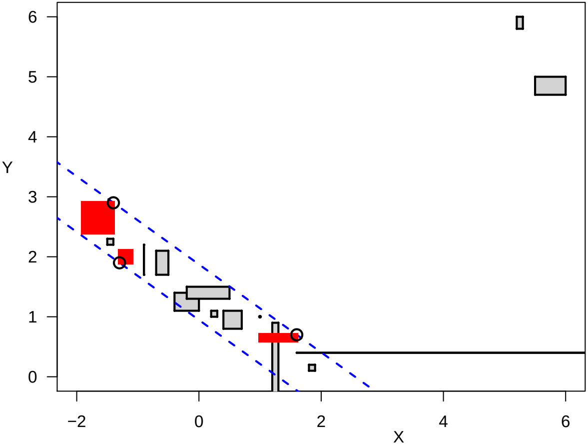
Implementation: Exact algorithm - Part 1

- some of the included $\overline{k}$ imprecise data touch the border of $\overline{B}_{f_{LRM},\overline{q}_{LRM}}$ in 3 different points



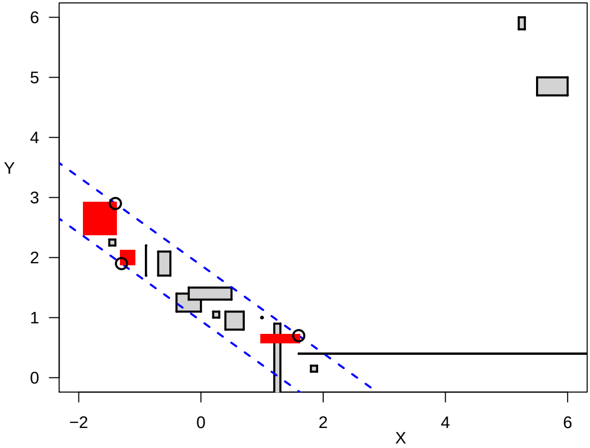
Implementation: Exact algorithm - Part 1

- some of the included $\overline{k}$ imprecise data touch the border of $\overline{B}_{f_{LRM},\overline{q}_{LRM}}$ in 3 different points
- b_{LRM} can be any slope determined by the corresponding corner points of 2 imprecise data or 0



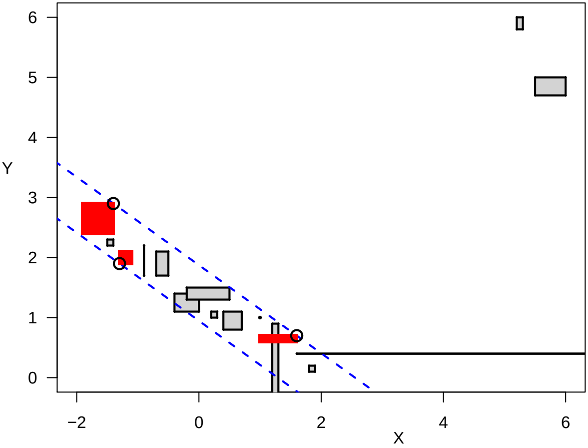
Implementation: Exact algorithm - Part 1

- some of the included $\overline{k}$ imprecise data touch the border of $\overline{B}_{f_{LRM},\overline{q}_{LRM}}$ in 3 different points
- b_{LRM} can be any slope determined by the corresponding corner points of 2 imprecise data or 0
- $\mathcal{B}$: set of all $4\binom{n}{2}+1$ possible values for b_{LRM}



Implementation: Exact algorithm - Part 1

- some of the included $\overline{k}$ imprecise data touch the border of $\overline{B}_{f_{LRM},\overline{q}_{LRM}}$ in 3 different points
- b_{LRM} can be any slope determined by the corresponding corner points of 2 imprecise data or 0
- $\mathcal{B}$: set of all $4\binom{n}{2}+1$ possible values for b_{LRM}
- for each $b \in \mathcal{B}$ find $a_b \in \mathbb{R}$ for which $\overline{r}_{f_{a_b,b},(\overline{k})}$ is minimal



Implementation: Exact algorithm - Part 1

for each $b \in \mathcal{B}$

- consider transformed data $z^*_{b,i} = [\underline{z}_{b,i}, \overline{z}_{b,i}]$ with

$$\underline{z}_{b,i} = \begin{cases} \underline{y}_i - b\,\overline{x}_i\,, & b > 0 \\ \underline{y}_i - b\,\underline{x}_i\,, & b \leq 0 \end{cases} \quad \text{and} \quad \overline{z}_{b,i} = \begin{cases} \overline{y}_i - b\,\underline{x}_i\,, & b > 0 \\ \overline{y}_i - b\,\overline{x}_i\,, & b \leq 0 \end{cases}$$

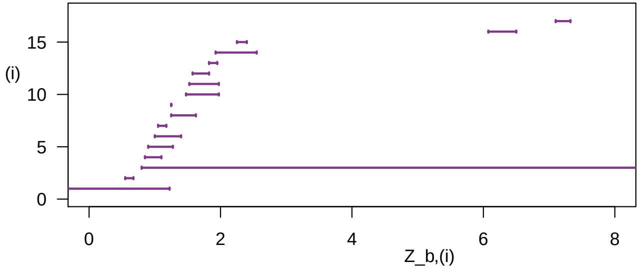
Implementation: Exact algorithm - Part 1

for each $b \in \mathcal{B}$

- consider transformed data $z^*_{b,i} = [\underline{z}_{b,i}, \overline{z}_{b,i}]$ with

$$\underline{z}_{b,i} = \begin{cases} \underline{y}_i - b\,\overline{x}_i\,, & b > 0 \\ \underline{y}_i - b\,\underline{x}_i\,, & b \leq 0 \end{cases} \quad \text{and} \quad \overline{z}_{b,i} = \begin{cases} \overline{y}_i - b\,\underline{x}_i\,, & b > 0 \\ \overline{y}_i - b\,\overline{x}_i\,, & b \leq 0 \end{cases}$$

- for example, $z^*_{b,i}$ for $b = -0.25$, ordered by lower endpoint



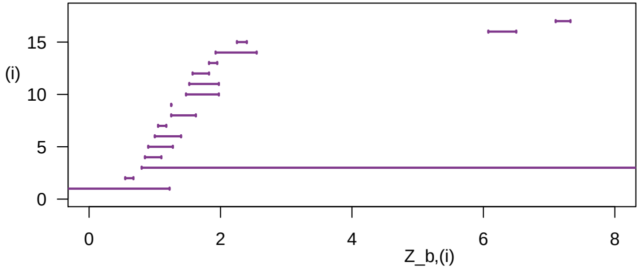
Implementation: Exact algorithm - Part 1

for each $b \in \mathcal{B}$

- consider transformed data $z^*_{b,i} = [\underline{z}_{b,i}, \overline{z}_{b,i}]$ with

$$\underline{z}_{b,i} = \begin{cases} \underline{y}_i - b\,\overline{x}_i\,, & b > 0 \\ \underline{y}_i - b\,\underline{x}_i\,, & b \leq 0 \end{cases} \quad \text{and} \quad \overline{z}_{b,i} = \begin{cases} \overline{y}_i - b\,\underline{x}_i\,, & b > 0 \\ \overline{y}_i - b\,\overline{x}_i\,, & b \leq 0 \end{cases}$$

- for example, $z^*_{b,i}$ for $b = -0.25$, ordered by lower endpoint



- find the shortest interval $\mathcal{I}_b$ containing at least $\overline{k}$ of the transformed data $z^*_{b,i}$, i.e., determine the shortest of the $n - \overline{k} + 1$ intervals of the form $[\underline{z}_{b,(j)}, \overline{z}_{b,[j]}]$, where $\overline{z}_{b,[j]}$ is the $\overline{k}$th smallest value among the $\overline{z}_{b,i}$ such that $\underline{z}_{b,i} \geq \underline{z}_{b,(j)}$

Implementation: Exact algorithm - Part 1

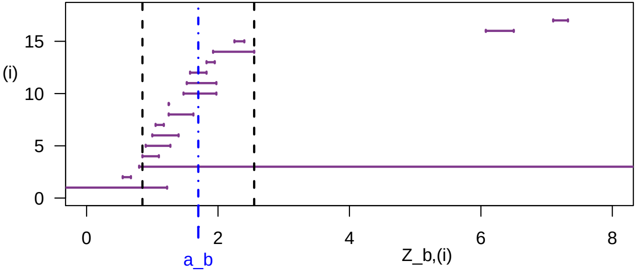
for each $b \in \mathcal{B}$

- the length of $\mathcal{I}_b$ corresponds to the width of the closed band around the function $f_{a_b,b}$ containing at least $\overline{k}$ imprecise data

Implementation: Exact algorithm - Part 1

for each $b \in \mathcal{B}$

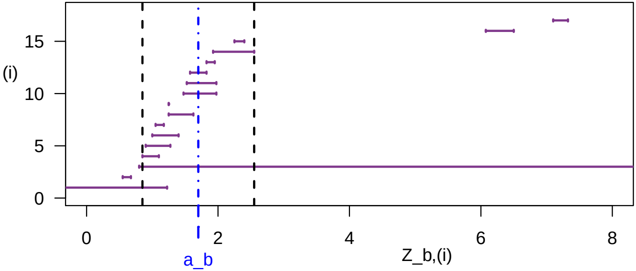
- the length of $\mathcal{I}_b$ corresponds to the width of the closed band around the function $f_{a_b,b}$ containing at least $\overline{k}$ imprecise data
- the corresponding intercept a_b is given by the midpoint of the interval $\mathcal{I}_b$
- for example, $b = -0.25$



Implementation: Exact algorithm - Part 1

for each $b \in \mathcal{B}$

- the length of $\mathcal{I}_b$ corresponds to the width of the closed band around the function $f_{a_b,b}$ containing at least $\overline{k}$ imprecise data
- the corresponding intercept a_b is given by the midpoint of the interval $\mathcal{I}_b$
- for example, $b = -0.25$

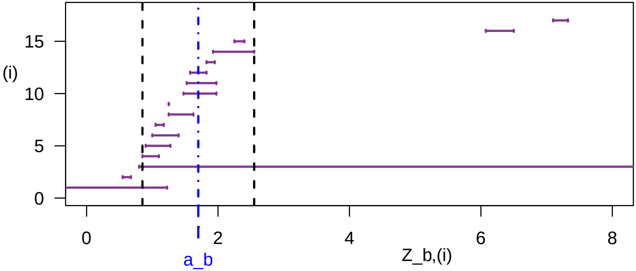


- the associated $\overline{r}_{f_{a_b,b},(\overline{k})}$ corresponds to half of the length of $\mathcal{I}_b$

Implementation: Exact algorithm - Part 1

for each $b \in \mathcal{B}$

- the length of $\mathcal{I}_b$ corresponds to the width of the closed band around the function $f_{a_b,b}$ containing at least $\overline{k}$ imprecise data
- the corresponding intercept a_b is given by the midpoint of the interval $\mathcal{I}_b$
- for example, $b = -0.25$



- the associated $\overline{r}_{f_{a_b,b},(\overline{k})}$ corresponds to half of the length of $\mathcal{I}_b$

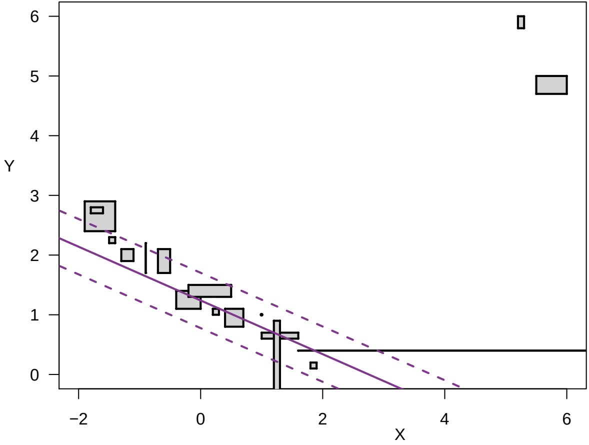
then, we obtain $\overline{q}_{LRM}$ by $\quad \overline{q}_{LRM} = \dfrac{1}{2} \min_{(b,j)\in\mathcal{B}\times\{1,\ldots,n-\overline{k}+1\}} \left(\overline{z}_{b,[j]} - \underline{z}_{b,(j)}\right)$

Implementation: Exact algorithm - Part 2

- step 2: determine $\mathcal{U}$

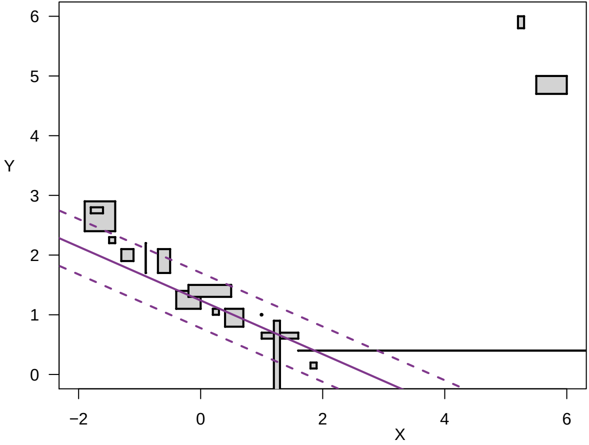
Implementation: Exact algorithm - Part 2

- step 2: determine $\mathcal{U}$
- for $f \in \mathcal{U}$, the band $\overline{B}_{f,\overline{q}_{LRM}}$ intersects at least $\underline{k} + 1$ data
- here $\underline{k} = 8$



Implementation: Exact algorithm - Part 2

- step 2: determine $\mathcal{U}$
- for $f \in \mathcal{U}$, the band $\overline{B}_{f,\overline{q}_{LRM}}$ intersects at least $\underline{k} + 1$ data
- here $\underline{k} = 8$
- for $b \in \mathbb{R}$ find all intercept values $a \in \mathbb{R}$, for which $\underline{r}_{f_{a,b},(\underline{k}+1)} \leq \overline{q}_{LRM}$



Implementation: Exact algorithm - Part 2

for a given $b \in \mathbb{R}$

- consider again the transformed data $z^*_{b,i}$

Implementation: Exact algorithm - Part 2

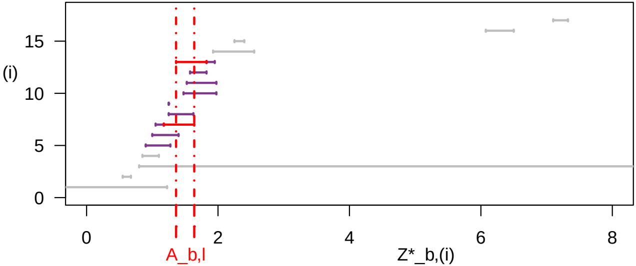
for a given $b \in \mathbb{R}$

- consider again the transformed data $z^*_{b,i}$
- $A_{b,l}$ is the set of interval midpoints $a \in \mathbb{R}$, for which the interval $[a - \overline{q}_{LRM}, a + \overline{q}_{LRM}]$ intersects all $z^*_{b,i}$ of the lth subset of size $\underline{k} + 1$

Implementation: Exact algorithm - Part 2

for a given $b \in \mathbb{R}$

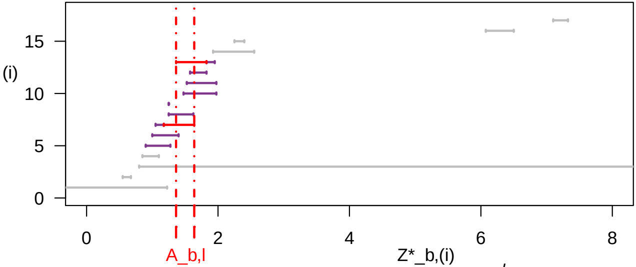
- consider again the transformed data $z^*_{b,i}$
- $A_{b,l}$ is the set of interval midpoints $a \in \mathbb{R}$, for which the interval $[a - \overline{q}_{LRM}, a + \overline{q}_{LRM}]$ intersects all $z^*_{b,i}$ of the lth subset of size $\underline{k} + 1$
- for example, $b = -0.25$



Implementation: Exact algorithm - Part 2

for a given $b \in \mathbb{R}$

- consider again the transformed data $z^*_{b,i}$
- $A_{b,l}$ is the set of interval midpoints $a \in \mathbb{R}$, for which the interval $[a - \overline{q}_{LRM}, a + \overline{q}_{LRM}]$ intersects all $z^*_{b,i}$ of the lth subset of size $\underline{k} + 1$
- for example, $b = -0.25$

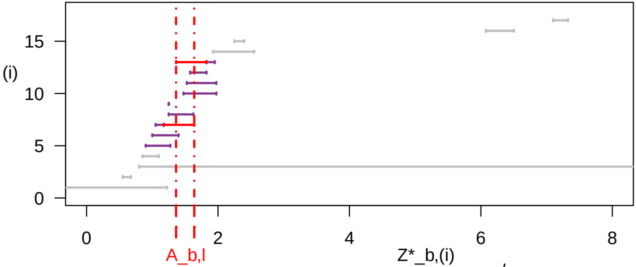


- the union of all $A_{b,l}$ is equivalent to $\mathcal{A}_b = \bigcup_{j=1}^{n-\underline{k}} [\underline{z}_{b,(\underline{k}+j)} - \overline{q}_{LRM}, \overline{z}_{b,(j)} + \overline{q}_{LRM}]$

Implementation: Exact algorithm - Part 2

for a given $b \in \mathbb{R}$

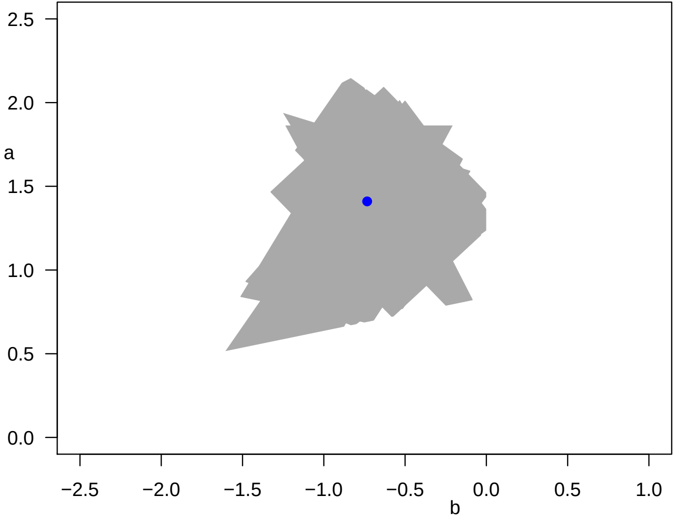
- consider again the transformed data $z^*_{b,i}$
- $A_{b,l}$ is the set of interval midpoints $a \in \mathbb{R}$, for which the interval $[a - \overline{q}_{LRM}, a + \overline{q}_{LRM}]$ intersects all $z^*_{b,i}$ of the lth subset of size $\underline{k} + 1$
- for example, $b = -0.25$



- the union of all $A_{b,l}$ is equivalent to $\mathcal{A}_b = \bigcup_{j=1}^{n-\underline{k}} [\underline{z}_{b,(\underline{k}+j)} - \overline{q}_{LRM}, \overline{z}_{b,(j)} + \overline{q}_{LRM}]$

finally, we obtain $\mathcal{U}$ as the set $\{f_{a,b} : b \in \mathbb{R} \text{ and } a \in \mathcal{A}_b\}$

Resulting set of undominated parameters



Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- aim: determine the LIR result $\mathcal{U}$, i.e., the set of all functions that are plausible relations of X and Y in the light of the imprecise observations

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- aim: determine the LIR result $\mathcal{U}$, i.e., the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- aim: determine the LIR result $\mathcal{U}$, i.e., the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression
- the presented algorithm has computational complexity $O(n^3 \log n)$

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- aim: determine the LIR result $\mathcal{U}$, i.e., the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression
- the presented algorithm has computational complexity $O(n^3 \log n)$
- further details in: M. Cattaneo, A. Wiencierz (2012). *On the implementation of LIR: the case of simple linear regression with interval data*. Technical Report 127. Department of Statistics. LMU Munich.

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- aim: determine the LIR result $\mathcal{U}$, i.e., the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression
- the presented algorithm has computational complexity $O(n^3 \log n)$
- further details in: M. Cattaneo, A. Wiencierz (2012). *On the implementation of LIR: the case of simple linear regression with interval data*. Technical Report 127. Department of Statistics. LMU Munich.

`linLIR` package

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- aim: determine the LIR result $\mathcal{U}$, i.e., the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression
- the presented algorithm has computational complexity $O(n^3 \log n)$
- further details in: M. Cattaneo, A. Wiencierz (2012). *On the implementation of LIR: the case of simple linear regression with interval data*. Technical Report 127. Department of Statistics. LMU Munich.

`linLIR` package

- `linLIR`: linear Likelihood-based Imprecise Regression, available at CRAN: `http://cran.r-project.org/`

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- aim: determine the LIR result $\mathcal{U}$, i.e., the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression
- the presented algorithm has computational complexity $O(n^3 \log n)$
- further details in: M. Cattaneo, A. Wiencierz (2012). *On the implementation of LIR: the case of simple linear regression with interval data*. Technical Report 127. Department of Statistics. LMU Munich.

`linLIR` package

- `linLIR`: linear Likelihood-based Imprecise Regression, available at CRAN: `http://cran.r-project.org/`
- function to plot 2-dimensional interval data set, two example data sets

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- aim: determine the LIR result $\mathcal{U}$, i.e., the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression
- the presented algorithm has computational complexity $O(n^3 \log n)$
- further details in: M. Cattaneo, A. Wiencierz (2012). *On the implementation of LIR: the case of simple linear regression with interval data*. Technical Report 127. Department of Statistics. LMU Munich.

`linLIR` package

- `linLIR`: linear Likelihood-based Imprecise Regression, available at CRAN: `http://cran.r-project.org/`
- function to plot 2-dimensional interval data set, two example data sets
- `s.linlir` function implements the exact algorithm

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

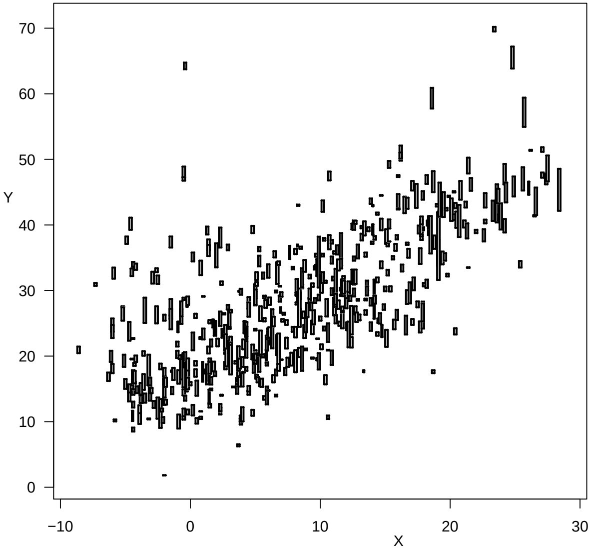
- aim: determine the LIR result $\mathcal{U}$, i.e., the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression
- the presented algorithm has computational complexity $O(n^3 \log n)$
- further details in: M. Cattaneo, A. Wiencierz (2012). *On the implementation of LIR: the case of simple linear regression with interval data*. Technical Report 127. Department of Statistics. LMU Munich.

`linLIR` package

- `linLIR`: linear Likelihood-based Imprecise Regression, available at CRAN: `http://cran.r-project.org/`
- function to plot 2-dimensional interval data set, two example data sets
- `s.linlir` function implements the exact algorithm
- further tools to summarize and visualize results

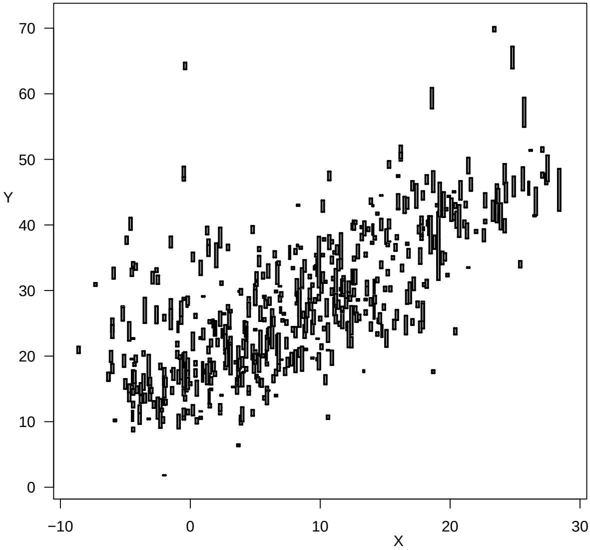
Example - Data set

- 2-dimensional interval data set of $n = 514$ observations



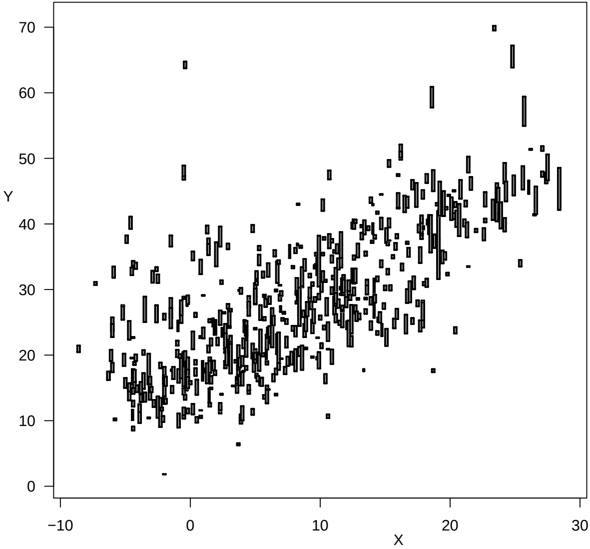
Example - Data set

- 2-dimensional interval data set of $n = 514$ observations
- LIR analysis with $p = 0.5, \beta = 0.26, \varepsilon = 0$



Example - Data set

- 2-dimensional interval data set of $n = 514$ observations
- LIR analysis with $p = 0.5, \beta = 0.26, \varepsilon = 0$
- $\underline{k} = 238\,,\ \overline{k} = 276$



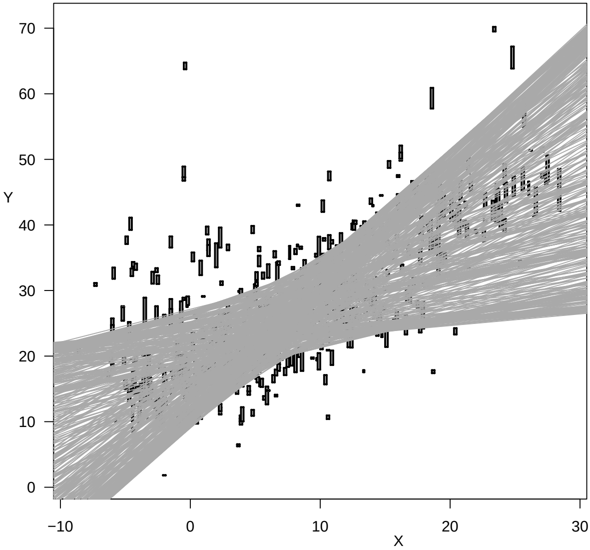
Example - R code

```
library(linLIR)
data(pm10)
pm.idf <- idf.create(pm10, var.labels=c("X","Y"))
pm.lir <- s.linlir(pm.idf, p = 0.5, bet=0.26, epsilon = 0)
summary(pm.lir)
```

```
Ranges of parameter values of the undominated functions:
intercept of f in [8.977766,27.18173]
slope of f in [0.11,1.898]

Bandwidth:  10.79207

Estimated parameters of the function f.lrm:
intercept of f.lrm:  18.39075
slope of f.lrm:  1.059185

Number of observations:  514

LIR settings:
p:  0.5 beta:  0.26 epsilon:  0 k.l:  238 k.u:  276
confidence level of each confidence interval:  90.61 %
```

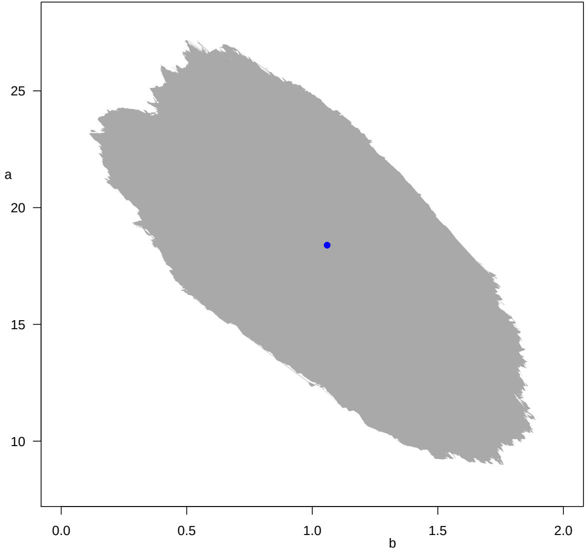
Example - Undominated regression functions

- 2-dimensional interval data set of $n = 514$ observations
- LIR analysis with $p = 0.5, \beta = 0.26, \varepsilon = 0$
- $\underline{k} = 238\,,\ \overline{k} = 276$
- obtained set of all undominated regression functions



Example - Undominated parameters

- 2-dimensional interval data set of $n = 514$ observations
- LIR analysis with $p = 0.5, \beta = 0.26, \varepsilon = 0$
- $\underline{k} = 238$, $\overline{k} = 276$
- obtained set of all undominated regression functions
- obtained set of parameters



Summary and Outlook

Summary and Outlook

- the imprecise result of the LIR analysis is the set of all functions that are plausible relations of X and Y in the light of the imprecise observations

Summary and Outlook

- the imprecise result of the LIR analysis is the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- for the special case of simple linear regression with interval data, we developed an exact algorithm to determine the set of all undominated functions

Summary and Outlook

- the imprecise result of the LIR analysis is the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- for the special case of simple linear regression with interval data, we developed an exact algorithm to determine the set of all undominated functions
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression

Summary and Outlook

- the imprecise result of the LIR analysis is the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- for the special case of simple linear regression with interval data, we developed an exact algorithm to determine the set of all undominated functions
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression
- the presented algorithm has computational complexity $O(n^3 \log n)$

Summary and Outlook

- the imprecise result of the LIR analysis is the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- for the special case of simple linear regression with interval data, we developed an exact algorithm to determine the set of all undominated functions
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression
- the presented algorithm has computational complexity $O(n^3 \log n)$
- the exact algorithm is implemented in R as part of the `linLIR` package

Summary and Outlook

- the imprecise result of the LIR analysis is the set of all functions that are plausible relations of X and Y in the light of the imprecise observations
- for the special case of simple linear regression with interval data, we developed an exact algorithm to determine the set of all undominated functions
- the first part of the algorithm generalizes the exact algorithm for Least Quantile of Squares regression
- the presented algorithm has computational complexity $O(n^3 \log n)$
- the exact algorithm is implemented in R as part of the `linLIR` package
- future work: generalize algorithm to multiple linear regression