Linear regression with interval data: the LIR approach

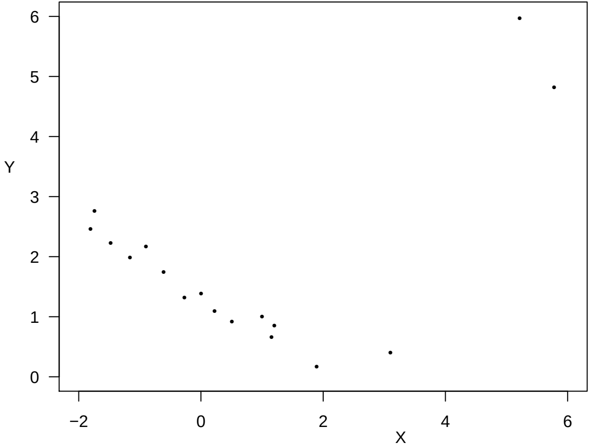
Andrea Wiencierz and Marco E. G. V. Cattaneo
Department of Statistics, LMU Munich

Statistische Woche, Vienna, Austria
September 20, 2012

Likelihood-based Imprecise Regression (LIR)

Likelihood-based Imprecise Regression (LIR)

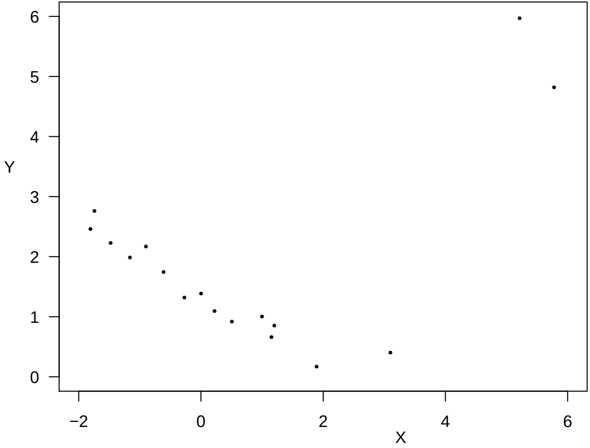
- $(X_1, Y_1), \ldots, (X_n, Y_n)$
  with $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} P$



Likelihood-based Imprecise Regression (LIR)

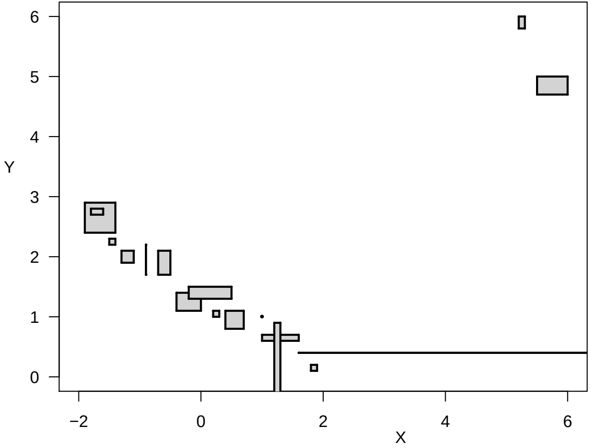
- $(X_1, Y_1), \ldots, (X_n, Y_n)$ with $(X_i, Y_i) \overset{\text{i.i.d.}}{\sim} P$
- simple linear regression:

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



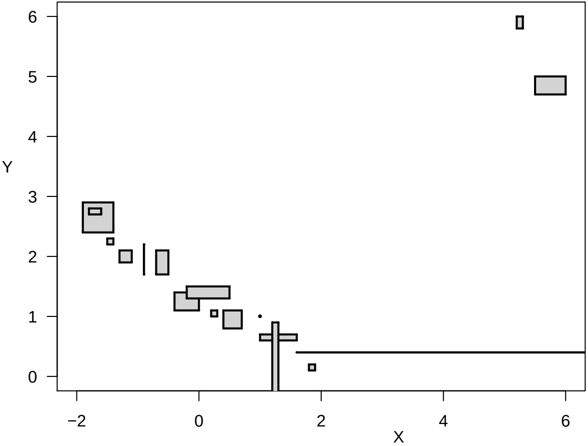
(Simple) linear LIR with interval data

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



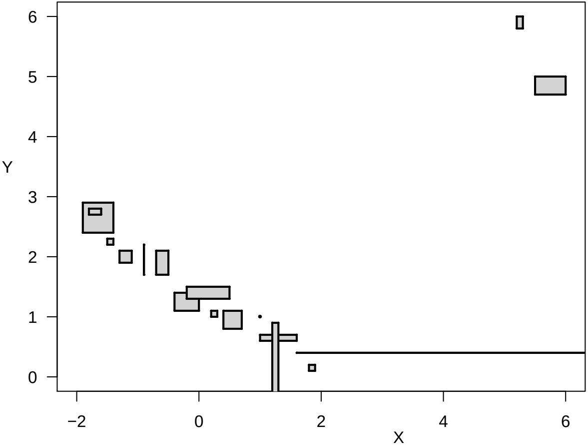
(Simple) linear LIR with interval data

- $(X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*)$
  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 $\varepsilon \in [0, 1]$
  $P((X_i, Y_i) \notin V_i^*) \leq \varepsilon$



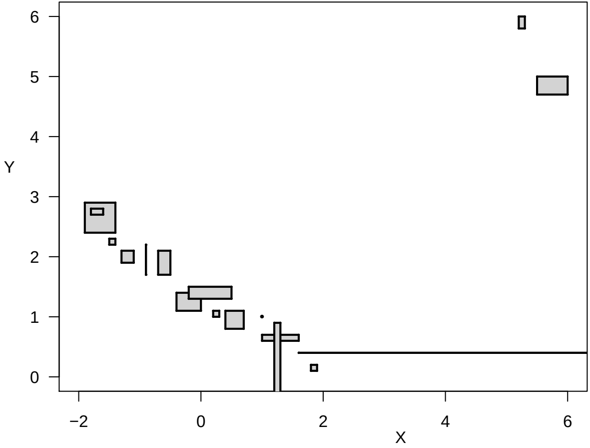
(Simple) linear LIR with interval data

- $(X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*)$
  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:
  $Y = f(X) = a + b\,X$



(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
  $R_{f,i} = |Y_i - f(X_i)|$

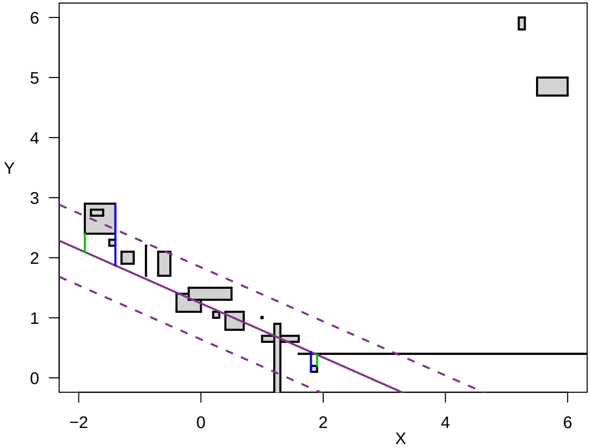


(Simple) linear LIR with interval data

- 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

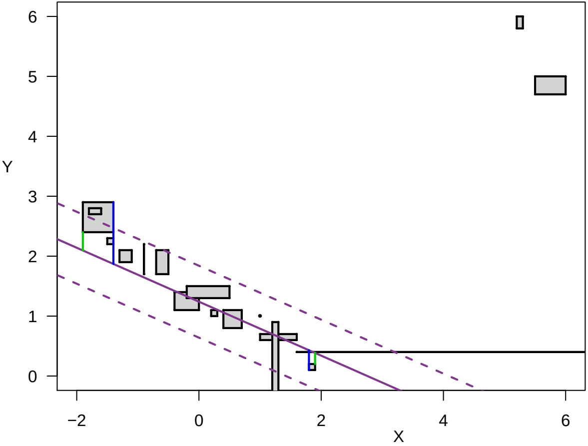
- imprecise residuals:

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

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

- uncertainty about f:

  data imprecision and statistical uncertainty



(Simple) linear LIR with interval data

- imprecise residuals:

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

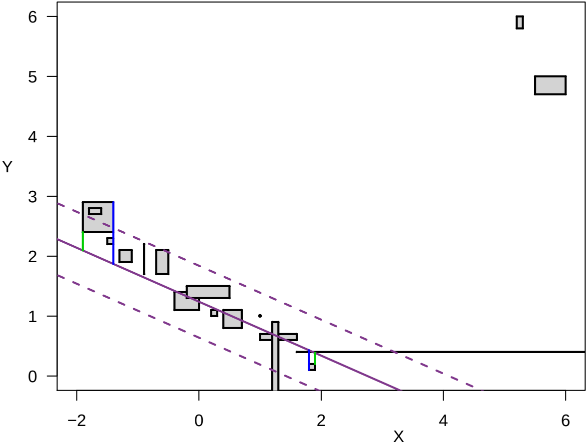
$$\overline{r}_{f,i} = \sup_{(x,y)\in v_i^*} |y - f(x)|$$

- 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:

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

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

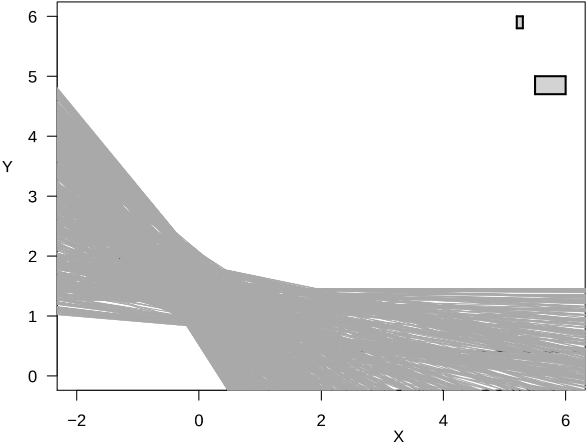
- uncertainty about f:

  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

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

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 $v_1^*, \ldots, v_n^*$ induce (normalized) profile likelihood function $lik_{Q_{R_f}}$ of the p-quantile of the distribution of R_f for each $f \in \mathcal{F}$

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:

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

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 LRM criterion and called f_{LRM}; LRM means Likelihood-based Region Minimax

Recapitulation: (simple) linear LIR with interval data

- $((X_i, Y_i), V_i^*) \overset{\text{i.i.d.}}{\sim} P\,,\ P \in \mathcal{P}_\varepsilon = \{P : P((X_i, Y_i) \notin V_i^*) \leq \varepsilon\}\ ,\ \varepsilon \in [0, 1]$
- $Y_i = f(X_i)\,,\ f \in \mathcal{F} = \left\{ f_{a,b} : \begin{array}{ccc} \mathbb{R} & \rightarrow & \mathbb{R} \\ X & \mapsto & a + b\,X \end{array}\ ,\ a, b \in \mathbb{R} \right\}$
- 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 LRM criterion and called f_{LRM}; LRM means Likelihood-based Region Minimax
- further details in: M. Cattaneo, A. Wiencierz (2012). *Likelihood-based Imprecise Regression*. Int. J. Approx. Reasoning 53. 1137-1154.

Statistical properties of the LIR method

Statistical properties of the LIR method

- the presented method for linear LIR generalizes Least Quantile of Squares (LQS) regression to imprecise data and to accounting directly for statistical uncertainty

Statistical properties of the LIR method

- the presented method for linear LIR generalizes Least Quantile of Squares (LQS) regression to imprecise data and to accounting directly for statistical uncertainty
- very robust due to nonparametric probability model and quantiles,

Statistical properties of the LIR method

- the presented method for linear LIR generalizes Least Quantile of Squares (LQS) regression to imprecise data and to accounting directly for statistical uncertainty
- very robust due to nonparametric probability model and quantiles,

  breakdown-point $\epsilon^* = \frac{\min\{\underline{k}, n-\overline{k}\}}{n} \overset{n\to\infty}{\longrightarrow} \min\{p, 1-p\} - \varepsilon$

Statistical properties of the LIR method

- the presented method for linear LIR generalizes Least Quantile of Squares (LQS) regression to imprecise data and to accounting directly for statistical uncertainty
- very robust due to nonparametric probability model and quantiles,

  breakdown-point $\epsilon^* = \frac{\min\{\underline{k}, n-\overline{k}\}}{n} \stackrel{n\to\infty}{\longrightarrow} \min\{p, 1-p\} - \varepsilon$
- exact confidence level of $\mathcal{C}_f$:

Statistical properties of the LIR method

- the presented method for linear LIR generalizes Least Quantile of Squares (LQS) regression to imprecise data and to accounting directly for statistical uncertainty
- very robust due to nonparametric probability model and quantiles,

  breakdown-point $\epsilon^* = \frac{\min\{\underline{k}, n-\overline{k}\}}{n} \overset{n\to\infty}{\longrightarrow} \min\{p, 1-p\} - \varepsilon$
- exact confidence level of $\mathcal{C}_f$:

$$\inf_{P \in \mathcal{P}_\varepsilon} P(\mathcal{C}_f \ni Q_{R_f}) = \begin{cases} \sum_{k=\underline{k}+1}^{\overline{k}} \binom{n}{k} p^k (1-p)^{n-k} & \varepsilon = 0 \\ \sum_{k=\underline{k}+1}^{\overline{k}} \binom{n}{k} (p+\varepsilon)^k (1-(p+\varepsilon))^{n-k} & \varepsilon > 0, p \leq 0.5 \\ \sum_{k=\underline{k}+1}^{\overline{k}} \binom{n}{k} (p-\varepsilon)^k (1-(p-\varepsilon))^{n-k} & \varepsilon > 0, p > 0.5 \end{cases}$$

Implementation: Exact algorithm for simple linear LIR

Implementation: Exact algorithm for simple linear LIR

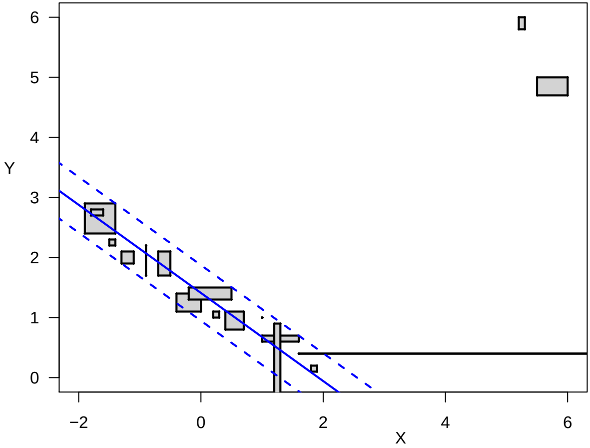
- aim: determine the set of undominated functions $\mathcal{U} = \{f \in \mathcal{F} : \underline{r}_{f,(\underline{k}+1)} \leq \overline{q}_{LRM}\}$

Implementation: Exact algorithm for simple linear LIR

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

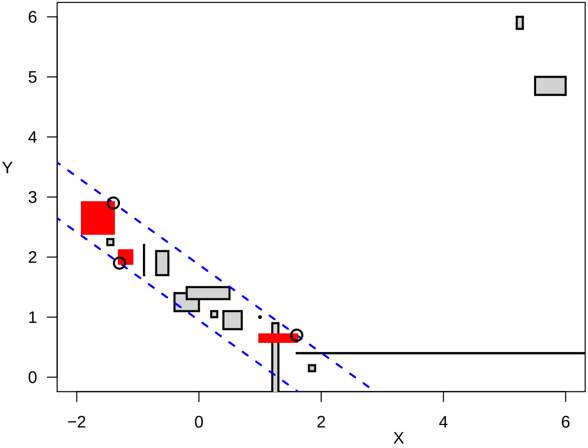
Implementation: Exact algorithm for simple linear LIR

- aim: determine the set of 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 $\overline{k} = 12$



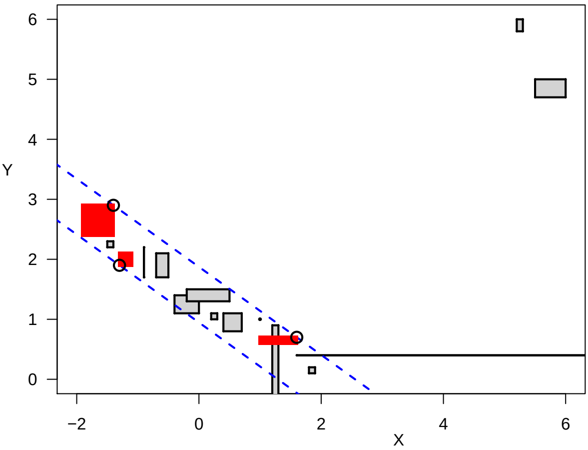
Implementation: Exact algorithm - Part 1

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



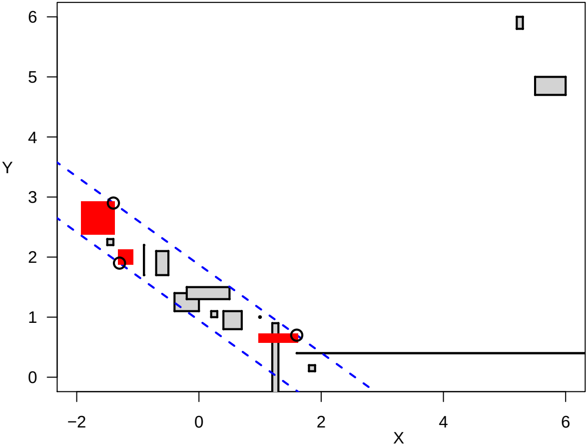
Implementation: Exact algorithm - Part 1

- 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



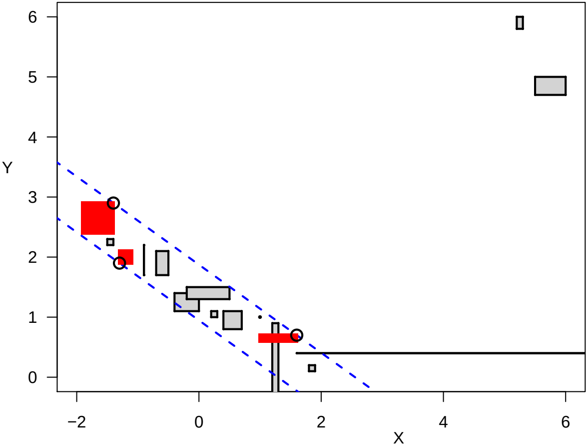
Implementation: Exact algorithm - Part 1

- 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}



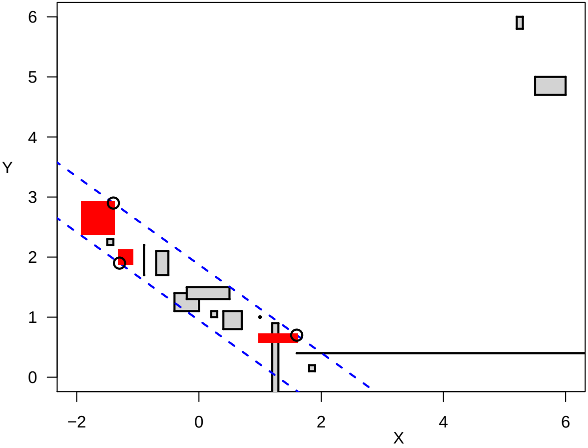
Implementation: Exact algorithm - Part 1

- 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



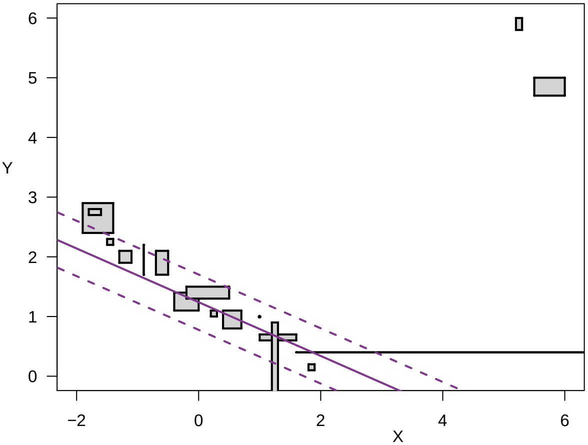
Implementation: Exact algorithm - Part 1

- 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
- $\overline{q}_{LRM} = \min_{b \in \mathcal{B}} \overline{r}_{f_{a_b,b}}$



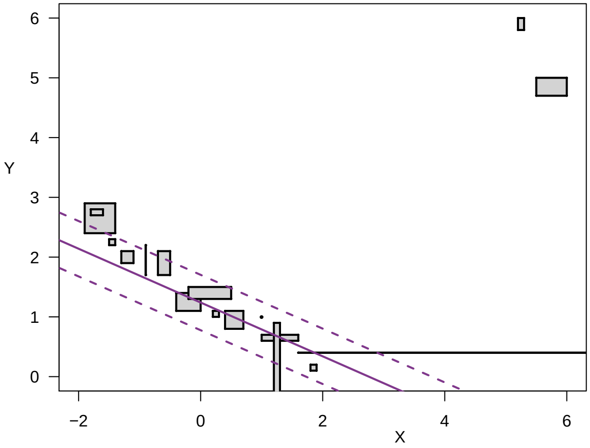
Implementation: Exact algorithm - Part 2

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



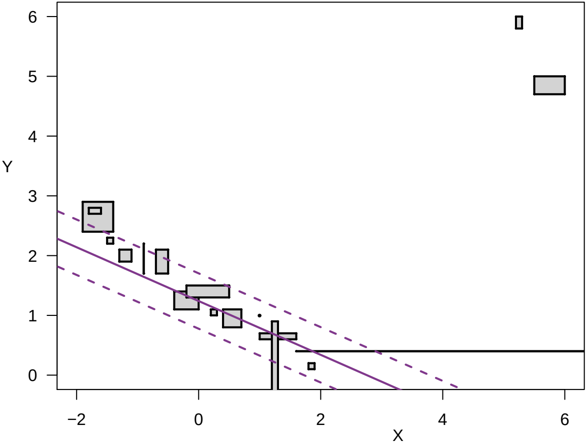
Implementation: Exact algorithm - Part 2

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



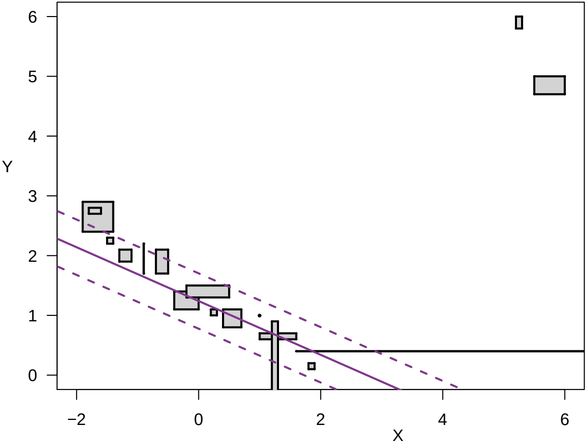
Implementation: Exact algorithm - Part 2

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

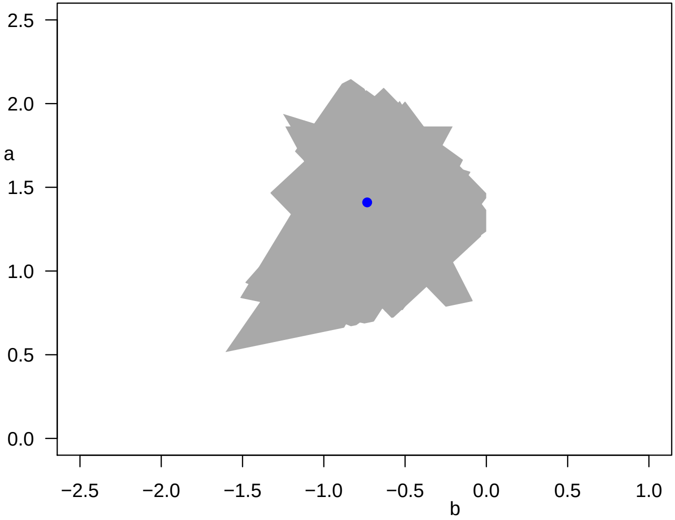


Implementation: Exact algorithm - Part 2

- step 2: determine $\mathcal{U}$
- if $f \in \mathcal{U}$, then $\overline{B}_{f,\overline{q}_{LRM}}$ intersects at least $\underline{k}+1$ imprecise data
- here $\underline{k} = 8$
- for each $b \in \mathbb{R}$ find the of intercept values $a \in \mathbb{R}$, for which $\underline{r}_{f_{a,b},(\underline{k}+1)} \leq \overline{q}_{LRM}$
- $\mathcal{U}$ can also be represented by the corresponding subset of the parameter space



Set of undominated parameters



Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- the 1st part of the algorithm generalizes the exact algorithm for LQS regression

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- the 1st part of the algorithm generalizes the exact algorithm for LQS 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

- the 1st part of the algorithm generalizes the exact algorithm for LQS 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

- the 1st part of the algorithm generalizes the exact algorithm for LQS 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

- the 1st part of the algorithm generalizes the exact algorithm for LQS 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

- the 1st part of the algorithm generalizes the exact algorithm for LQS 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

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

- the 1st part of the algorithm generalizes the exact algorithm for LQS 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
- `s.linlir` function implements the exact algorithm

Implementation of the algorithm in R

Recapitulation: Exact algorithm for simple linear LIR with interval data

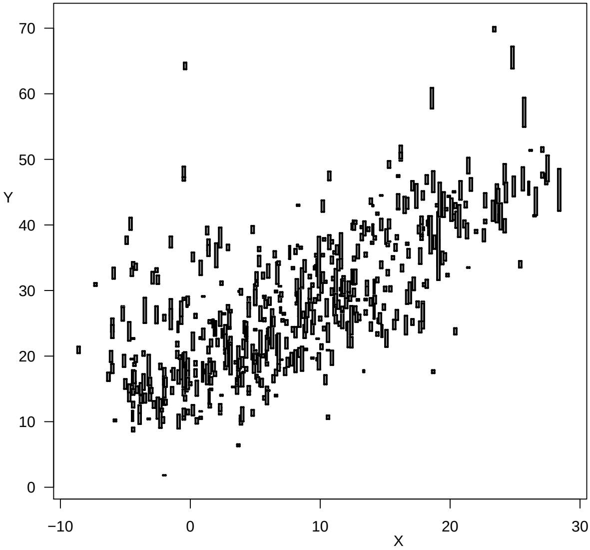
- the 1st part of the algorithm generalizes the exact algorithm for LQS regression
- the presented algorithm has computational complexity $O(n^3 \log n)$
- further details in: M. Cattaneo, A. Wiencierz (2012). *On the implementation of LIR: the case of simple linear regression with interval data*. Technical 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
- `s.linlir` function implements the exact algorithm
- further tools to summarize and visualize results

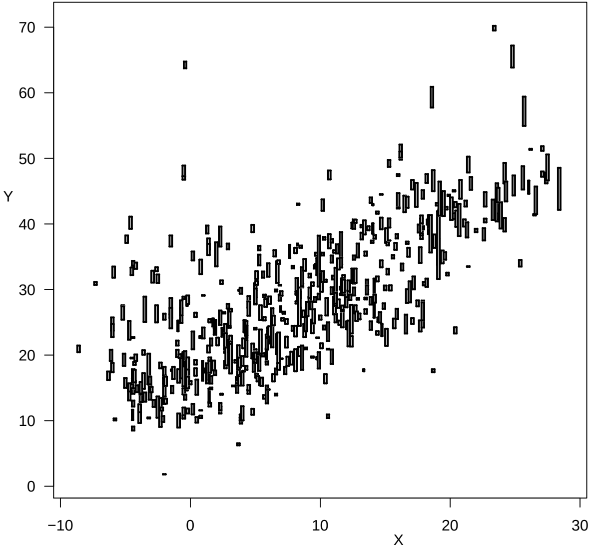
Example

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



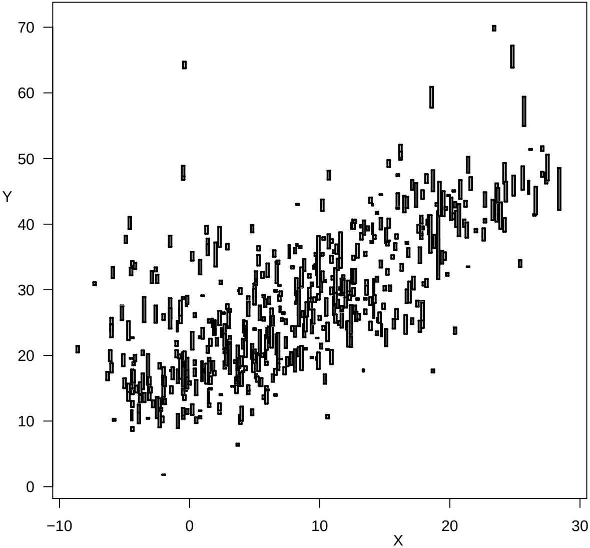
Example

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



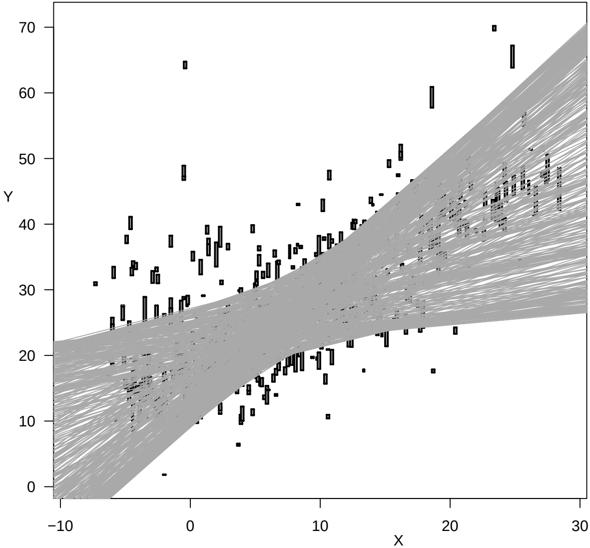
Example

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



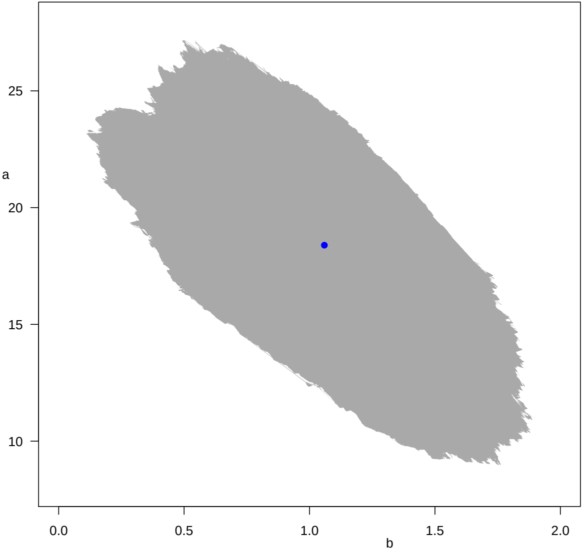
Example

- 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 undominated functions



Example

- 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 undominated functions
- obtained set of parameters



Summary and Outlook

Summary and Outlook

- the LIR approach provides a very robust regression method for imprecisely observed variables

Summary and Outlook

- the LIR approach provides a very robust regression method for imprecisely observed variables
- 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 LIR approach provides a very robust regression method for imprecisely observed variables
- 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 LIR approach provides a very robust regression method for imprecisely observed variables
- 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 exact algorithm is implemented in R as part of the `linLIR` package

Summary and Outlook

- the LIR approach provides a very robust regression method for imprecisely observed variables
- 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 exact algorithm is implemented in R as part of the `linLIR` package
- current / future work:

Summary and Outlook

- the LIR approach provides a very robust regression method for imprecisely observed variables
- 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 exact algorithm is implemented in R as part of the `linLIR` package
- current / future work:
  - further investigate statistical properties, in particular the confidence level of $\mathcal{U}$

Summary and Outlook

- the LIR approach provides a very robust regression method for imprecisely observed variables
- 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 exact algorithm is implemented in R as part of the `linLIR` package
- current / future work:
  - further investigate statistical properties, in particular the confidence level of $\mathcal{U}$
  - generalize algorithm to multiple linear regression