Utilizing Support Functions and Monotone Location Estimators for the Estimation of Partially Identified Regression Models

Motivation

Let $Y = \beta_0 + \beta_1 \cdot X + \varepsilon$ be the classical simple linear model and let $x = (x_1, \ldots, x_n)'$ and $y = (y_1, \ldots y_n)'$ be *i.i.id*. samples from the model and $z = (1, x)$ the corresponding design matrix.
The least squares estimator is given by:

$$\hat{\beta}_{ls} = (z'z)^{-1}z'y$$

and is linear in y but not linear in x.

Partially Identified Models

- If either X or Y or both X and Y are only observed in intervals, the model becomes generally only partially identified.
- One possible approach to cope with interval valued data is to simply collect the obtained estimates from a classical procedure for all precise data compatible with the observed intervals.
- If only Y is interval-valued, because of the linearity of the least squares estimator, for the application of least squares, this collection is easy enough to calculate
- If also X is interval-valued, the calculation of this collection is very hard.
- For other more sophisticated estimators the problem is getting worse.

Another estimator

Theil-Sen estimator (simplest form, only slope):

$$\hat{\beta}_1 = \text{median}_{i \neq j} \quad \beta_1^{i,j}$$

with

$$\beta_1^{i,j} = \frac{y_j - y_i}{x_j - x_i}.$$

For $i \neq j$ it is simple to calculate the upper bound $\beta_{1u}^{i,j}$ and the lower bound $\beta_{1l}^{i,j}$ of $\beta_1^{i,j}$ as the precise data x_i, x_j and y_i, y_j varies in between the observed intervals. Because the median is a monotone function of the data one can simply calculate

$$\begin{aligned} \hat{\beta}_{1u} &= \text{median}_{i \neq j} \quad \beta_{1u}^{i,j} \\ \hat{\beta}_{1l} &= \text{median}_{i \neq j} \quad \beta_{1l}^{i,j} \end{aligned}$$

as (non sharp) bounds for the maximal and minimal values for the Theil-Sen estimator.

Problem:

These bounds are not sharp because one data point (x_i, y_i) has impact on many different $\beta_1^{i,j}$ at the same time, but the maximization/minimization of the $\beta_1^{i,j}$ was done independently from each other for every $i \neq j$.

Idea: Choose not all pairs (i,j) with $i \neq j$ but a set M of pairs (i,j) such that every i and j occurs only exactly one time to obtain

$$\hat{\beta}_{1u}^M = \text{median}_{(i,j)\in M} \quad \beta_{1u}^{i,j}$$
$$\hat{\beta}_{1l}^M = \text{median}_{(i,j)\in M} \quad \beta_{1l}^{i,j}.$$

In fact, $\hat{\beta}_{1u}^M$ $\hat{\beta}_{1l}^M$ actually correspond to specific data points compatible with the interval data for which this „freely“ maximized/minimized values are actually obtained, so the bounds are sharp for the modified estimator

$$\hat{\beta}_1^M := \text{median}_{(i,j)\in M} \quad \beta_1^{i,j}$$

(but the estimator $\hat{\beta}_1^M$ is often less efficient than $\hat{\beta}_1$).

Further modifications:

- use not only the median but other monotone location estimators and
- weight the $\beta_1^{i,j}$ such that the variability of the obtained estimator is minimal.

Example: weighted mean, precise case

Let $x_1, \ldots, x_n$ be already in increasing order. For maximal efficiency of β_1 take

$$M = \{(1, N), (2, N-1), \ldots (N/2, N/2+1)\}$$

and the weight for $\beta_1^{i,j}$ proportional to $(x_j - x_i)^2$.
For the intercept take

$$\beta_0^{i,j} = y_i - \beta_1^{i,j} \cdot x_i \quad (= y_j - \beta_1^{i,j} \cdot x_j).$$

(And for an arbitrary linear combination $\langle d, \beta \rangle = d_0\beta_0 + d_1\beta_1$ take $\beta_d^{i,j} = d_0\beta_0^{i,j} + d_1\beta_1^{i,j}$.)
Then choose weights that minimize the variability of the corresponding estimator of β_0 (or β_d).
$\Longrightarrow$ The obtained estimator is then a linear form in y.

Example: $x = (1, 2, \ldots, 10)$

Estimation-matrix of least squares estimator:

$$\begin{pmatrix} 0.40 & 0.33 & 0.27 & 0.20 & 0.13 & 0.07 & 0.00 & -0.07 & -0.13 & -0.20 \\ -0.05 & -0.04 & -0.03 & -0.02 & -0.01 & 0.01 & 0.02 & 0.03 & 0.04 & 0.05 \end{pmatrix}$$

Variability under homoscedastic errors:

$$\beta_0 : \quad \frac{7}{15}\sigma^2 \approx 0.467\sigma^2$$
$$\beta_1 : \quad \frac{12}{990}\sigma^2 \approx 0.012\sigma^2$$

Estimation matrix of free weighted mean estimator:

$$\begin{pmatrix} 0.48 & 0.40 & 0.29 & 0.17 & 0.05 & -0.04 & -0.10 & -0.11 & -0.09 & -0.05 \\ -0.05 & -0.04 & -0.03 & -0.02 & -0.01 & 0.01 & 0.02 & 0.03 & 0.04 & 0.05 \end{pmatrix}$$

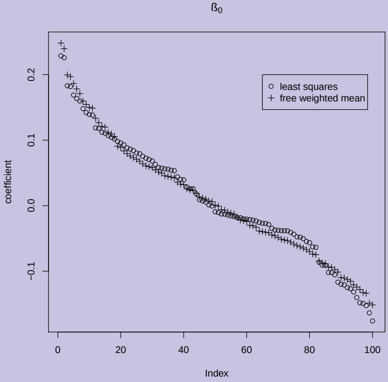
Variability under homoscedastic errors:

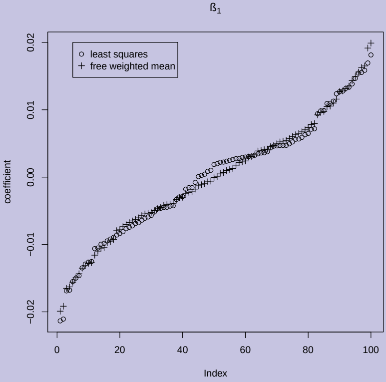
$$\beta_0 : \quad \frac{7}{15}\sigma^2 \approx 0.533\sigma^2$$
$$\beta_1 : \quad \approx 0.012\sigma^2$$

$\Longrightarrow$ *Efficiency of free weighted mean estimator:*

$$\beta_0 : \quad \approx 0.88$$
$$\beta_1 : \quad 1$$

Example: $X_1, \ldots, X_{100} \sim \mathcal{N}(10, 1)$, *Entries of the Estimation matrix (index corresponds to the ordered covariate values):*

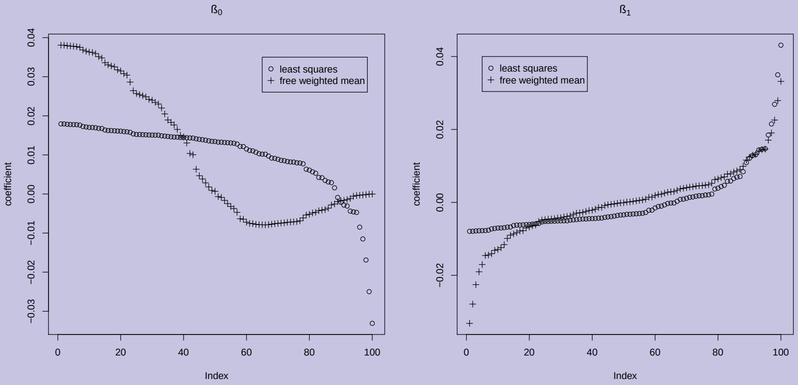




expected relative efficiency:

≈ 0.98 ≈ 0.99

Example: $X_1, \ldots, X_{100} \sim Exp(1)$*:*



expected relative efficiency:

≈ 0.56 ≈ 0.83

Relative efficiency for different settings and estimators (precise case)

Different settings ($N = 1000, X_1, \dots, X_n \sim \mathcal{N}(0,1)$)*:*

1. *standard setting*
2. *outliers in dependent variable („one wild": 10% of data randomly chosen and values multiplied by* 10*)*
3. *outliers in independent variable*
4. *error term t-distributed with* 3 *degrees of freedom*
5. *error term standard cauchy distributed*

Different Estimators:

1. *least squares*
2. *robust M-estimator rlm (psi = psi.huber)*
3. *MM-type estimator with bi-square redescending score function (with 50% breakdown point and 95% asymptotic efficiency for normal errors)*
4. *least quantile of squares (lqs, q=0.5)*
5. *different „free" estimators based on :*
    1. *median*
    2. *weighted median*
    3. *trimmed weighted Hodges-Lehmann estimator with winsorized weights (wwthl)*

estimated relative efficiencies based on $nrep = 10000$ samples:

setting	lm	weighted median	wwthl	median	lqs	rlm	lmrob
1	1.00	0.53	0.61	0.40	0.08	0.95	0.95
2	0.00	0.08	0.27	0.09	0.12	0.11	1.00
3	0.00	0.00	0.06	0.01	0.11	0.00	1.00
4	0.55	0.57	0.59	0.42	0.21	1.00	1.00
5	0.00	0.48	0.41	0.36	0.68	0.79	1.00

Imprecise case

- For maximal/minimal $\hat{\beta}_0^M, \hat{\beta}_1^M$ take bounds as described above.
- If one is interested in the whole identification region IR and not only in projections one can work with support functions and estimate for every $d \in \mathbb{R}^2$ the value of $\sup_{\beta \in IR} \langle d, \beta \rangle$ as

$$\beta_{du} = \sup_{x \in [\underline{x}, \overline{x}], y \in [\underline{y}, \overline{y}]} \beta_d$$

with

$$\beta_d = l\left(\beta_d^{1,N}, \beta_d^{2,N-1} \ldots, \beta_d^{\frac{N}{2}, \frac{N}{2}+1}\right)$$

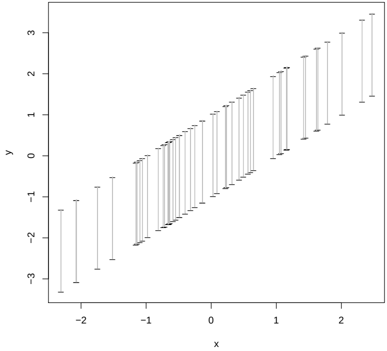
where l is an appropriate monotone location estimator (with weights $w(d)$ minimizing variability).

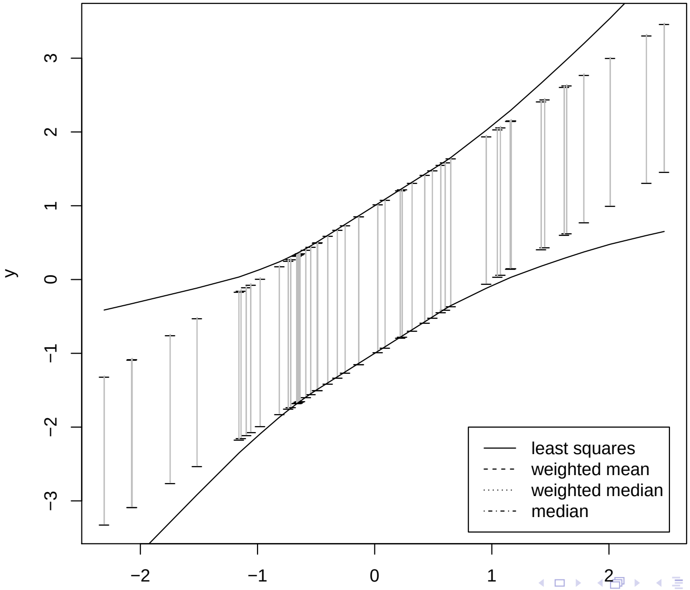
- The obtained estimate of the support function of the identified set is then generally no longer a support function of some set.

$\Longrightarrow$ Project the estimated function onto the space of support functions in a certain way.

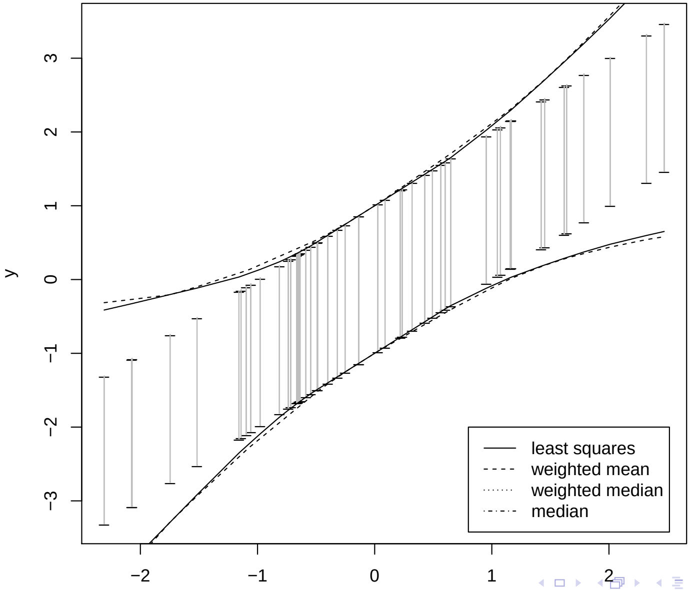
Example

Identification regions for different estimators illustrated as the predicted boundaries $\inf_{\beta \in IR} \beta_0 + \beta_1 x$ and $\sup_{\beta \in IR} \beta_0 + \beta_1 x$ for different covariate values x, where $X_1, \ldots, X_{50} \sim \mathcal{N}(0,1)$, $\underline{Y} = X - 1, \overline{Y} = X + 1$.

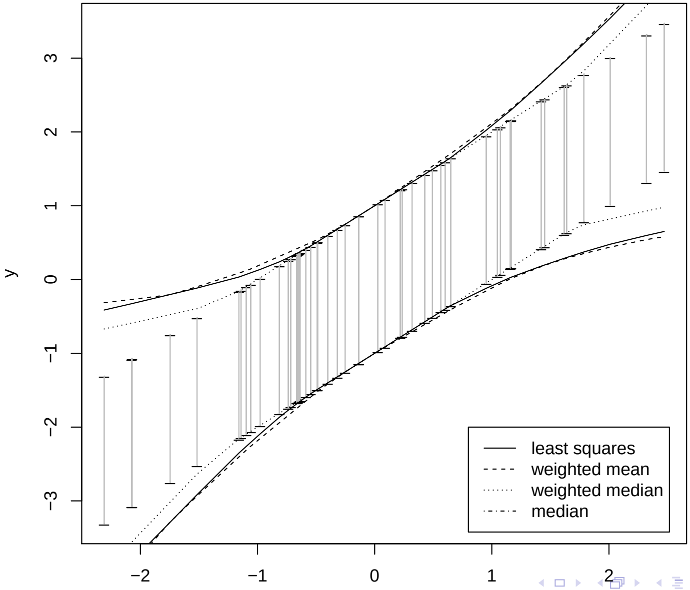




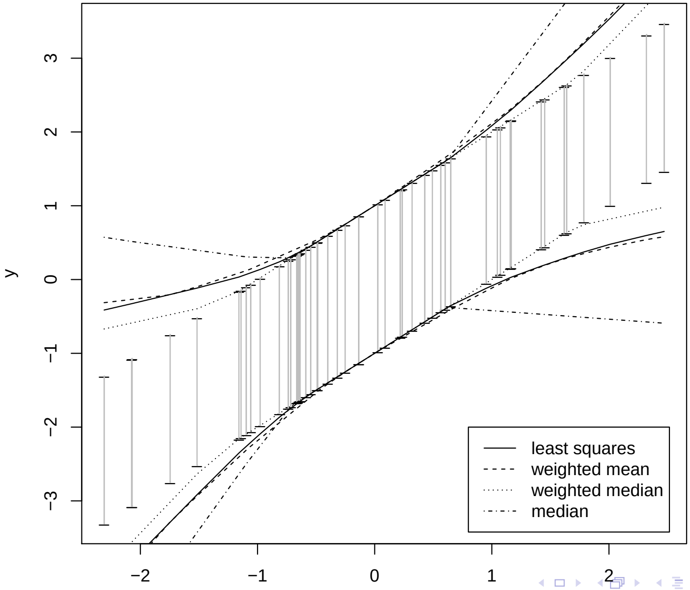
y
−3
−2
−1
0
1
2
3
−2
−1
0
1
2
least squares
weighted mean
weighted median
median



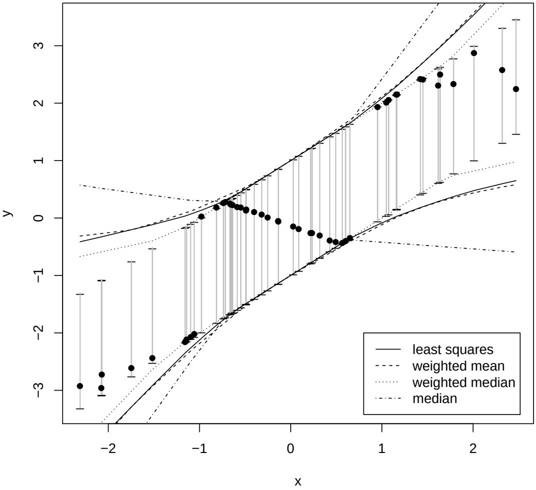
y
−3
−2
−1
0
1
2
3
−2
−1
0
1
2
least squares
weighted mean
weighted median
median



y
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
least squares
weighted mean
weighted median
median


y
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
least squares
weighted mean
weighted median
median

Minimal slope for *lmrob*:



Minimal slope for *lmrob*: