ARE THE BAD LEVERAGE POINTS THE MOST DIFFICULT PROBLEM FOR ESTIMATING THE UNDERLYING REGRESSION MODEL?

RESEARCH SUPPORTED BY THE GRANT OF THE CZECH SCIENCE FOUNDATION 13-01930S PANEL P402
ROBUST METHODS FOR NONSTANDARD SITUATIONS, THEIR DIAGNOSTICS AND IMPLEMENTATIONS

ROBUST 2016, 11 - 16 SEPTEMBER, 2016
Sporthotel Kurzovní, Jeseníky

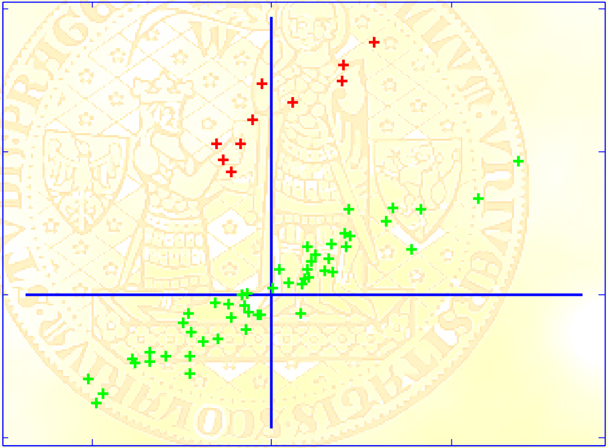
JAN ÁMOS VÍŠEK
CHARLES UNIVERSITY IN PRAGUE

I unlike JV @JW ...

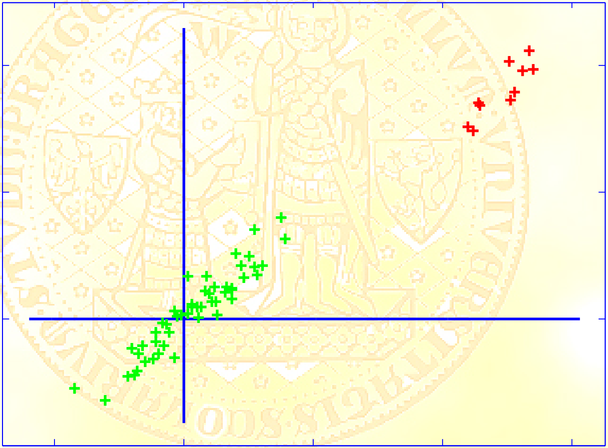
Dovolte bych suše předpokládal,...

(Allow me to just simply assume, ...)

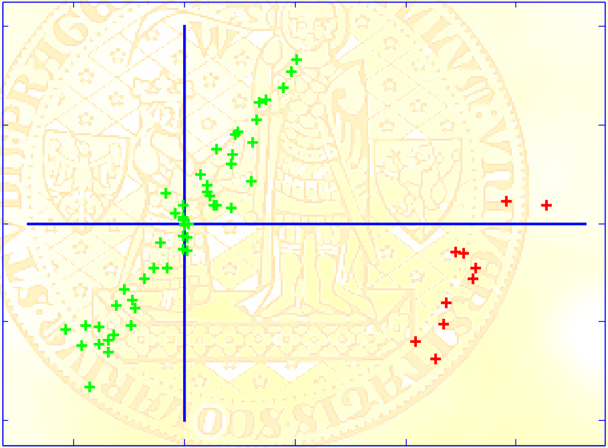
OUTLIERS



GOOD LEVERAGE POINTS



BAD LEVERAGE POINTS



ARE THE BAD LEVERAGE POINTS

THE MOST DIFFICULT PROBLEM

FOR ESTIMATING

THE UNDERLYING REGRESSION MODEL?

ARE THE BAD LEVERAGE POINTS

THE MOST DIFFICULT PROBLEM

FOR ESTIMATING

THE UNDERLYING REGRESSION MODEL?

Sometimes yes, sometimes no.

The main goal of talk (except of the issue given in title)

The main goal of talk (except of the issue given in title)

To attract Your attention to:

S-weighted estimators

The main goal of talk (except of the issue given in title)

To attract Your attention to:

S-weighted estimators

Hence I am going to speak about:

1. Inspiration - through the history of (highly) robust methods
   - the largest part of talk

The main goal of talk (except of the issue given in title)

To attract Your attention to:

S-weighted estimators

Hence I am going to speak about:

1. Inspiration - through the history of (highly) robust methods
   - the largest part of talk
2. Technicalities - consistency, etc. - *very briefly*

The main goal of talk (except of the issue given in title)

To attract Your attention to:

S-weighted estimators

Hence I am going to speak about:

1. Inspiration - through the history of (highly) robust methods
- the largest part of talk
2. Technicalities - consistency, etc. - *extremely briefly*

The main goal of talk (except of the issue given in title)

To attract Your attention to:

S-weighted estimators

Hence I am going to speak about:

1. Inspiration - through the history of (highly) robust methods
   - the largest part of talk
2. Technicalities - consistency, etc. - *extremely briefly*
3. How S-weighted estimators work (including also algorithm)
   - I believe, the most interesting part of talk

Content

Content

1. The history of high breakdown point estimation
   - Basic framework
   - The pursuit for high breakdown point estimator
   - Frustrations and rebirth - by the smooth depression of ...
2. S-weighted estimator
   - Combining S-estimator and the Least Weighted Squares
   - Asymptotics of SWE
   - SWE - how does it work ? The algorithm and patterns of results

Basic framework and notations

For today explanation,

let's assume the framework as simple as possible:

Regression model

$$Y_i = X_i'\beta^0 + \varepsilon_i, \qquad i = 1, 2, ..., n, \quad X_i \in R^p$$

Basic framework and notations

For today explanation,
let's assume the framework as simple as possible:

Regression model

$$Y_i = X_i'\beta^0 + \varepsilon_i, \qquad i = 1, 2, ..., n, \quad X_i \in R^p$$

Ordinary least squares

$$\hat{\beta}^{(OLS,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} (Y_i - X_i'\beta)^2$$

Basic framework and notations

For today explanation,
let's assume the framework as simple as possible:

Regression model

$$Y_i = X_i' \beta^0 + \varepsilon_i, \qquad i = 1, 2, ..., n, \quad X_i \in R^p$$

Ordinary least squares

$$\hat{\beta}^{(OLS,n)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} (Y_i - X_i' \beta)^2 = (X'X)^{-1} X'Y$$

Basic framework and notations

For today explanation,

let's assume the framework as simple as possible:

Regression model

$$Y_i = X_i' \beta^0 + \varepsilon_i, \qquad i = 1, 2, ..., n, \quad X_i \in R^p$$

Ordinary least squares

$$\hat{\beta}^{(OLS,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} (Y_i - X_i'\beta)^2 = (X'X)^{-1} X'Y$$

$$= \left(\frac{1}{n} X'X\right)^{-1} \frac{1}{n} X'Y = \widehat{cov}^{-1}(X, X)\, \widehat{cov}(X, Y)$$

Basic framework and notations

The best estimator of scale of error terms (of disturbances) is

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \left(Y_i - X_i' \hat{\beta}^{(OLS,n)} \right)^2 ,$$

Basic framework and notations

The best estimator of scale of error terms (of disturbances) is

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \left(Y_i - X_i' \hat{\beta}^{(OLS,n)} \right)^2 ,$$

so that we can change the definition of $\hat{\beta}^{(OLS,n)}$ to

$$\hat{\beta}^{(OLS,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} (Y_i - X_i' \beta)^2$$

Basic framework and notations

The best estimator of scale of error terms (of disturbances) is

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \left(Y_i - X_i' \hat{\beta}^{(OLS,n)} \right)^2,$$

so that we can change the definition of $\hat{\beta}^{(OLS,n)}$ to

$$\hat{\beta}^{(OLS,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} (Y_i - X_i'\beta)^2$$

$$= \arg\min_{\beta \in R^p} \left\{ \sigma : \frac{1}{n-p} \sum_{i=1}^{n} \left(\frac{Y_i - X_i'\beta}{\sigma} \right)^2 = 1 \right\}.$$

Basic framework and notations

All vectors will be assumed as column vector !

The orthogonality condition is assumed to be fulfilled (for simplicity).

Basic framework and notations

All vectors will be assumed as column vector !

The orthogonality condition is assumed to be fulfilled (for simplicity).

Residuals: for any $\beta \in R^p \quad \rightarrow \quad r_i(\beta) = Y_i - X_i'\beta$

Order statistics of squared residuals:

$$r^2_{(1)}(\beta) \leq r^2_{(2)}(\beta) \leq \ldots \leq r^2_{(n)}(\beta)$$

Basic framework and notations

All vectors will be assumed as column vector !

The orthogonality condition is assumed to be fulfilled (for simplicity).

Residuals: for any $\beta \in R^p \quad \rightarrow \quad r_i(\beta) = Y_i - X_i'\beta$

Order statistics of squared residuals:

$$r^2_{(1)}(\beta) \leq r^2_{(2)}(\beta) \leq ... \leq r^2_{(n)}(\beta)$$

(notice that with changing β, the order of squared residuals is also changing).

Content

Breakdown point - "finite" sample version

Data: $(x_1, x_2, \ldots, x_n) \quad \Rightarrow \quad T_n(x_1, x_2, \ldots, x_n)$

Breakdown point - "finite" sample version

Data: $(x_1, x_2, ..., x_n) \quad \Rightarrow \quad T_n(x_1, x_2, ..., x_n)$

Find maximal m_n such that for data:

$$(x_1, x_2, ..., x_{n-m_n}, y_1, y_2, ..., y_{m_n})$$
$$\Rightarrow \quad \|T_n(x_1, x_2, ..., x_{n-m_n}, y_1, y_2, ..., y_{m_n})\| < \infty$$

A pursuit for highly robust estimator of regression coefficients

When the *median* had been taken into account, a question appeared:

CAN WE CONSTRUCT AN ESTIMATOR OF REGRESSION COEFFICIENTS WITH $\varepsilon^* = \frac{1}{2}$?

A pursuit for highly robust estimator of regression coefficients

When the *median* had been taken into account, a question appeared:

CAN WE CONSTRUCT AN ESTIMATOR OF REGRESSION COEFFICIENTS WITH $\varepsilon^* = \frac{1}{2}$?

Andrews, D. F., P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, J. W. Tukey (1972): *Robust Estimates of Location: Survey and Advances.* Princeton University Press, Princeton, N. J.

or

Bickel, P. J. (1975): One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc. 70, 428–433.*

A pursuit for highly robust estimator of regression coefficients

Why did we crave
for high brakedown point estimators ?

A pursuit for highly robust estimator of regression coefficients

Why did we crave
for high brakedown point estimators ?

1. First of all, it was challenge !!

A pursuit for highly robust estimator of regression coefficients

Why did we crave
for high brakedown point estimators ?

1. First of all, it was challenge !!
2. A hope: At a cost of a loss of efficiency,
   a reliable information about the underlying model.

A pursuit for highly robust estimator of regression coefficients

Why did we crave
for high brakedown point estimators ?

1. First of all, it was challenge !!
2. A hope: At a cost of a loss of efficiency,
   a reliable information about the underlying model.

A pursuit for highly robust estimator of regression coefficients

Why did we crave
for high brakedown point estimators ?

1. First of all, it was challenge !!
2. A hope: At a cost of a loss of efficiency,
   a reliable information about the underlying model.

It became a nightmare
of statisticians and econometricians !

The first estimator with 50% breakdown point

Repeated medians

Siegel, A. F. (1982): Robust regression using repeated medians.
Biometrica, 69, 242 - 244.

$$\hat{\beta}^{(j)} = \operatorname*{med}_{i_1=1,2,\ldots,n} \left(\ldots \left(\operatorname*{med}_{i_{p-1}=1,2,\ldots,n} \left(\operatorname*{med}_{i_p=1,2,\ldots,n} \left(\hat{\beta}_j^{(OLS,n)} \Big((Y_{i_1}, X_{i_1}), (Y_{i_2}, X_{i_2}), \ldots, (Y_{i_p}, X_{i_p}) \Big) \right) \right) \right) \right).$$

(Regrettably, unfeasible, except for the simple regression.)

The first estimator with 50% breakdown point

Repeated medians

Siegel, A. F. (1982): Robust regression using repeated medians. *Biometrica, 69, 242 - 244.*

$$\hat{\beta}^{(j)} = \operatorname*{med}_{i_1=1,2,\ldots,n} \left(\ldots \left(\operatorname*{med}_{i_{p-1}=1,2,\ldots,n} \left(\operatorname*{med}_{i_p=1,2,\ldots,n} \left(\hat{\beta}_j^{(OLS,n)} \left((Y_{i_1}, X_{i_1}), (Y_{i_2}, X_{i_2}), \ldots, (Y_{i_p}, X_{i_p}) \right) \right) \right) \right) \right).$$

(Regrettably, unfeasible, except for the simple regression.)

The first feasible solution broke the mystery and implied a chain of others

the Least Median of Squares

Rousseeuw, P. J. (1983): Least median of square regression.
Journal of Amer. Statist. Association 79, pp. 871-880.

$$\hat{\beta}^{(LMS,n,h)} = \arg\min_{\beta \in R^p} r^2_{(h)}(\beta) \qquad \frac{n}{2} < h \leq n.$$

Many advantages - mainly

1. *breakdown point equal to* $([\frac{n-p}{2}] + 1)n^{-1}$ *if* $h = [\frac{n}{2}] + [\frac{p+1}{2}]$,
2. *scale- and regression equivariant*
(without any studentization of residuals - in contrast to M-estimators).

The first feasible solution broke the mystery and implied a chain of others

the Least Median of Squares

Rousseeuw, P. J. (1983): Least median of square regression.
Journal of Amer. Statist. Association 79, pp. 871-880.

$$\hat{\beta}^{(LMS,n,h)} = \arg\min_{\beta \in R^p} \; r^2_{(h)}(\beta) \qquad \frac{n}{2} < h \le n.$$

Many advantages - mainly

1. *breakdown point equal to* $([\frac{n-p}{2}] + 1)n^{-1}$ *if* $h = [\frac{n}{2}] + [\frac{p+1}{2}]$,
2. *scale- and regression equivariant*
   (without any studentization of residuals - in contrast to M-estimators).

Main disadvantage $\sqrt[3]{n}\left(\hat{\beta}^{(LMS,n,h)} - \beta^0\right) = \mathcal{O}_p(1).$

Let's remove the deficiency of LMS

the Least Trimmed Squares

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986): *Robust Statistics – The Approach Based on Influence Functions.* New York: J.Wiley & Son.

$$\hat{\beta}^{(LTS,n,h)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{h} r^2_{(i)}(\beta) \qquad \frac{n}{2} < h \le n,$$

(Notice the order of words, remember there is also the Trimmed Least Squares.)

Many advantages - e. g.

1. the breakdown point equal to $([\frac{n-p}{2}] + 1)n^{-1}$ if $h = [\frac{n}{2}] + [\frac{p+1}{2}]$,
2. *scale- and regression equivariant*,
3. $\sqrt{n}\left(\hat{\beta}^{(LTS,n,h)} - \beta^0\right) = \mathcal{O}_p(1)$.

An algorithm for LTS

A

Find the plane through $p+1$ randomly selected observations.

Evaluate squared residuals of all observations, order them and compute $S(\hat{\beta}_{present}) = \sum_{i=1}^{h} r_{(i)}^2(\beta)$.

Is $S(\hat{\beta}_{present})$ less than $S(\hat{\beta}_{past})$?

no

B

yes

Reorder the observations according to the order of squared residuals from the second step and establish *new* $\hat{\beta}_{present}$ just applying *OLS* on the h-tuple of the first h reordered observations.

Basic idea of algorithm for LTS

1. Consider any h-tuple of points from n observations, say $\boldsymbol{d}$, $\boldsymbol{h} \in (\frac{\boldsymbol{n}}{2}, \boldsymbol{n}]$.

Basic idea of algorithm for LTS

1. Consider any h-tuple of points from n observations, say $\boldsymbol{d}$, $\boldsymbol{h} \in (\frac{\boldsymbol{n}}{\boldsymbol{2}}, \boldsymbol{n}]$.
2. Apply *OLS* on them and compute the residual sum of squares $S(\boldsymbol{d})$.

Basic idea of algorithm for LTS

1. Consider any h-tuple of points from n observations, say $\boldsymbol{d}$, $\boldsymbol{h} \in (\frac{\boldsymbol{n}}{\boldsymbol{2}}, \boldsymbol{n}]$.
2. Apply *OLS* on them and compute the residual sum of squares $S(\boldsymbol{d})$.
3. $\exists$ an h-tuple $\boldsymbol{d}^*$ such that

$$\boldsymbol{S}(\boldsymbol{d}^*) \leq \boldsymbol{S}(\boldsymbol{d}).$$

An algorithm for LTS

B

Was ℓ-times found the same model with minimal value of $S(\beta)$?

no

Was already k-times repeated outer cycle ?

no

A

yes

yes

As $\hat{\beta}^{(LTS,n,w)}$ we will assume $\beta \in R^p$ for which the functional $S(\beta)$ attained - through just described iterations - minimal value.

ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994

Number of industries 91

X_ℓ	-	*export from i-th industry,*
US_ℓ	-	*number of university-passed employees in the i-th industry,*
HS_ℓ	-	*nuber of high school-passed employees in the i-th industry,*
VA_ℓ	-	*value added in the i-th industry,*
K_ℓ	-	*capital in the i-th industry,*
CR_ℓ	-	*percentage of market occupied by 3 largest producers,*
$TFPW_\ell$	-	*by wages normed productvity in the i-th industry,*
Bal_ℓ	-	*Balasa index in the i-th industry„*
DP_ℓ	-	*cost discontinuity in 1993 in the i-th industry*
	-	*etc., about 20 explanatory variables*

ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994

Number of industries 91

X_ℓ	-	*export from i-th industry,*
US_ℓ	-	*number of university-passed employees in the i-th industry,*
HS_ℓ	-	*nuber of high school-passed employees in the i-th industry,*
VA_ℓ	-	*value added in the i-th industry,*
K_ℓ	-	*capital in the i-th industry,*
CR_ℓ	-	*percentage of market occupied by 3 largest producers,*
$TFPW_\ell$	-	*by wages normed productvity in the i-th industry,*
Bal_ℓ	-	*Balasa index in the i-th industry„*
DP_ℓ	-	*cost discontinuity in 1993 in the i-th industry*
	-	*etc., about 20 explanatory variables*

NO REASONABLE MODEL BY OLS - COEFFICIENT OF DETERMINATION 0.28

ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *Least Trimmed Squares*

has found:

MAIN SUBGROUP

with number of industries 54 and the model

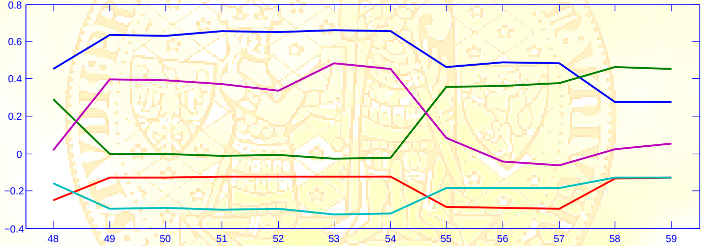
$$\frac{X_\ell}{S_\ell} = 4.64 - 0.032 \cdot \frac{US_\ell}{VA_\ell} - 0.022 \cdot \frac{HS_\ell}{VA_\ell} - 0.124 \cdot \frac{K_\ell}{VA_\ell} + 1.035 \cdot CR_\ell$$
$$-3.199 \cdot TFPW_\ell + 1.048 \cdot BAL_\ell + 0.452 \cdot DP_\ell + \varepsilon_\ell$$

X_ℓ	-	*export from i-th industry,*
US_ℓ	-	*number of university-passed employees in the i-th industry,*
HS_ℓ	-	*nuber of high school-passed employees in the i-th industry,*
VA_ℓ	-	*value added in the i-th industry,*
K_ℓ	-	*capital in the i-th industry,*
CR_ℓ	-	*percentage of market occupied by 3 largest producers,*
$TFPW_\ell$	-	*by wages normed productvity in the i-th industry,*
Bal_ℓ	-	*Balasa index in the i-th industry„*
DP_ℓ	-	*cost discontinuity in 1993 in the i-th industry*

with coefficient of determination 0.97 and stable submodels

Diagnostics by LTS

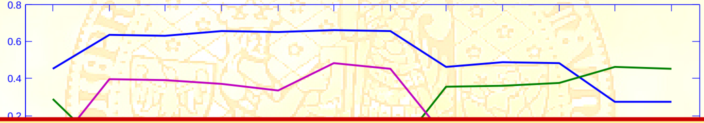
ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *least trimmed squares*.



The development of the estimates of regression coefficients. The blue line represents $\hat{\beta}_1^{(LTS,n,h)}$ (down-scaled by $\frac{1}{10}$), the purple is $\hat{\beta}_8^{(LTS,n,h)}$, the green is $\hat{\beta}_3^{(LTS,n,h)}$, the red is $\hat{\beta}_4^{(LTS,n,h)}$ and light blue (the lowest curve) is $\hat{\beta}_6^{(LTS,n,h)}$ (down-scaled again by $\frac{1}{10}$). There is an evident break at 54.

Diagnostics by LTS

ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *least trimmed squares*.



Atkinson, A. C., M. Riani, A. Cerioli (2004):
Exploring multivariate data with the forward search.
Springer, NY, Berlin, Heidelberg.

The development of the estimates of regression coefficients. The blue line represents $\hat{\beta}_1^{(LTS,n,h)}$ (down-scaled by $\frac{1}{10}$), the purple is $\hat{\beta}_8^{(LTS,n,h)}$, the green is $\hat{\beta}_3^{(LTS,n,h)}$, the red is $\hat{\beta}_4^{(LTS,n,h)}$ and light blue (the lowest curve) is $\hat{\beta}_6^{(LTS,n,h)}$ (down-scaled again by $\frac{1}{10}$). There is an evident break at 54.

Analysis of the export from the Czech Republic to EU in 1994 by means of the *Least Trimmed Squares*

has found:

Complementary subgroup

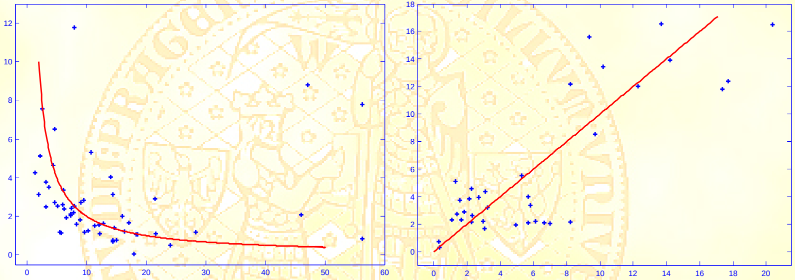
with number of industries 33 and the model

$$\frac{X_\ell}{S_\ell} = -0.634 + 0.089 \cdot \frac{US_\ell}{VA_\ell} + 0.235 \cdot \frac{HS_\ell}{VA_\ell} + 0.249 \cdot \frac{K_\ell}{VA_\ell} + 1.174 \cdot CR_\ell$$
$$+0.690 \cdot TFPW_\ell + 2.691 \cdot BAL_\ell - 0.051 \cdot DP_\ell + \varepsilon_\ell$$

X_ℓ	-	*export from i-th industry,*
US_ℓ	-	*number of university-passed employees in the i-th industry,*
HS_ℓ	-	*nuber of high school-passed employees in the i-th industry,*
VA_ℓ	-	*value added in the i-th industry,*
K_ℓ	-	*capital in the i-th industry,*
CR_ℓ	-	*percentage of market occupied by 3 largest producers,*
$TFPW_\ell$	-	*by wages normed productvity in the i-th industry,*
Bal_ℓ	-	*Balasa index in the i-th industry„*
DP_ℓ	-	*cost discontinuity in 1993 in the i-th industry*

with coefficient of determination 0.93 and stable submodels

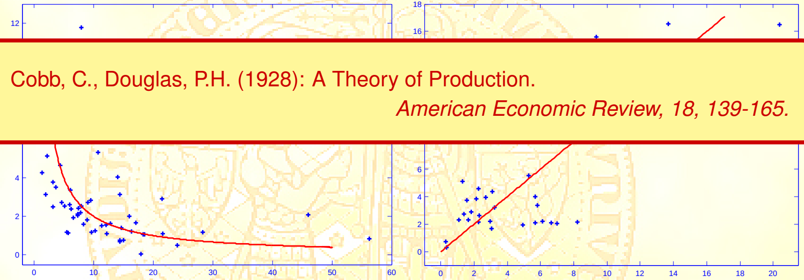
ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *Least Trimmed Squares*.



RELATION BETWEEN K/W AND L/S FOR THE *Main subpopulation* (LEFT PICTURE)

AND FOR THE *Complementary subpopulation* (RIGHT PICTURE).

ANALYSIS OF THE EXPORT FROM THE CZECH REPUBLIC TO EU IN 1994 BY MEANS OF THE *Least Trimmed Squares*.



Cobb, C., Douglas, P.H. (1928): A Theory of Production.
American Economic Review, 18, 139-165.

RELATION BETWEEN K/W AND L/S FOR THE *Main subpopulation* (LEFT PICTURE)

AND FOR THE *Complementary subpopulation* (RIGHT PICTURE).

A serious disadvantage of LTS

the Least Trimmed Squares

$$\hat{\beta}^{(LTS,n,h)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{h} r_{(i)}^2(\beta) \qquad \frac{n}{2} < h \leq n,$$

A serious disadvantage of LTS - order statistics in definition

the Least Trimmed Squares

$$\hat{\beta}^{(LTS,n,h)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{h} r^2_{(i)}(\beta) \qquad \frac{n}{2} < h \leq n,$$

A serious disadvantage of LTS - order statistics in definition

the Least Trimmed Squares

$$\hat{\beta}^{(LTS,n,h)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{h} r_{(i)}^2(\beta) \qquad \frac{n}{2} < h \leq n,$$

Moreover, probably lower efficiency than necessary !!

Let's increase the efficiency with simultaneously keeping high breakdown point

S-estimators

Rousseeuw, P. J., V. Yohai (1984):
Robust regression by means of *S*-estimators.
Lecture Notes in Statistics No. 26 Springer Verlag, New York, 256-272.

$$\hat{\beta}^{(S,n,\rho)} = \arg\min_{\beta \in R^p} \left\{ \sigma \in R^+ : \sum_{i=1}^{n} \rho\left(\frac{r_i(\beta)}{\sigma}\right) = b \right\}$$

where $b = \boldsymbol{E}\rho\left(\frac{e_i}{\sigma_0}\right)$ with $\sigma_0^2 = \boldsymbol{E}e_1^2$.

An alternative definition of OLS

The best estimator of scale of error terms (of disturbances) is

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \left(Y_i - X_i' \hat{\beta}^{(OLS,n)} \right)^2 ,$$

so that we can change the definition of $\hat{\beta}^{(OLS,n)}$ to

$$\hat{\beta}^{(OLS,n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} (Y_i - X_i'\beta)^2$$

$$= \arg\min_{\beta \in R^p} \left\{ \sigma : \frac{1}{n-p} \sum_{i=1}^{n} \left(\frac{Y_i - X_i'\beta}{\sigma} \right)^2 = 1 \right\} .$$

Let's increase the efficiency with simultaneously keeping high breakdown point

S-estimators

Rousseeuw, P. J., V. Yohai (1984):
Robust regression by means of S-estimators.
Lecture Notes in Statistics No. 26 Springer Verlag, New York, 256-272.

$$\hat{\beta}^{(S,n,\rho)} = \arg\min_{\beta \in R^p} \left\{ \sigma \in R^+ : \sum_{i=1}^{n} \rho\left(\frac{r_i(\beta)}{\sigma}\right) = b \right\}$$

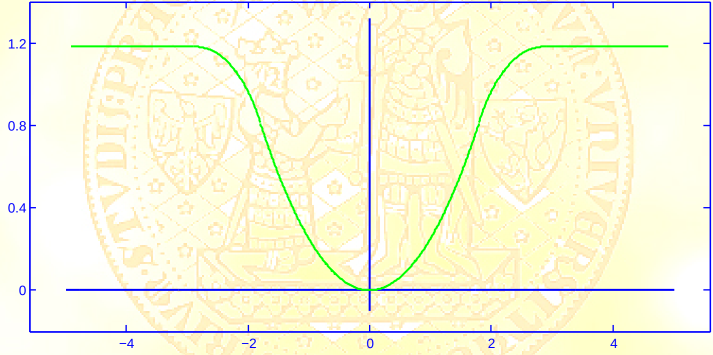
where $b = \boldsymbol{E}\rho\left(\frac{e_i}{\sigma_0}\right)$ with $\sigma_0^2 = \boldsymbol{E}e_1^2$ (for ρ see next slide).

Many advantages - e. g.

1. the breakdown point equal to 50% (if adjusted so),
2. *scale- and regression equivariant,*
3. $\sqrt{n}\left(\hat{\beta}^{(S,n,\rho)} - \beta^0\right) = \mathcal{O}_p(1)$,
4. much higher efficiency than LTS.

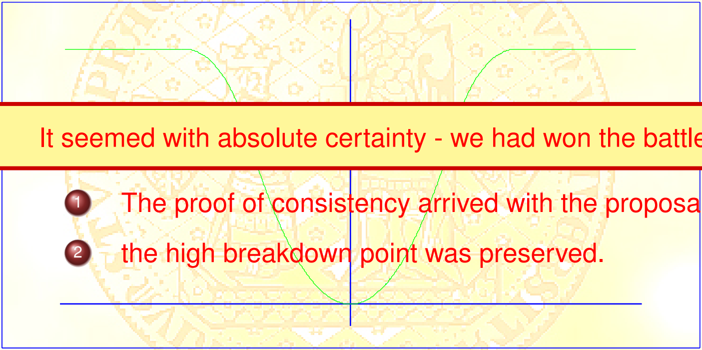
Tuckey's biweight function ρ

$$\rho : (-\infty, \infty) \to (0, \infty), \ \rho(x) = \rho(-x), \ \rho(0) = 0, \rho(x) = c \ \text{ for } \ x > d.$$



Tuckey's biweight function ρ

$$\rho : (-\infty, \infty) \to (0, \infty), \ \rho(x) = \rho(-x), \ \rho(0) = 0, \rho(x) = c \ \text{ for } \ x > d.$$



It seemed with absolute certainty - we had won the battle.

1. The proof of consistency arrived with the proposal,
2. the high breakdown point was preserved.

Computational simplicity

Moreover, nowadays we have at hand an efficient algorithm for it,

indicating that S-estimators are close to idea of the *minimal volume estimator*.

Computing *S*-estimator

Campbell, N. A., Lopuhaa, H. P., Rousseeuw, P. J. (1998):
On the calculation of a robust S-estimator of a covariance matrix.
Statist. Med. 17, 2685 - 2695.

$$\hat{\mu}_S = \frac{\sum_{i=1}^{n} \psi(d_i) \cdot d_i^{-1} \cdot X_i}{\sum_{i=1}^{n} \psi(d_i) \cdot d_i^{-1}}$$

with

$$d_i^2 = (X_i - \hat{\mu}_S)' \, \widehat{cov}_S^{-1} (X, X) \, (X_i - \hat{\mu}_S)$$

and

$$\widehat{cov}_S (X, X) = \frac{\sum_{i=1}^{n} \psi(d_i) \cdot d_i^{-1} \cdot (X_i - \hat{\mu}_S)(X_i - \hat{\mu}_S)'}{p^{-1} \sum_{i=1}^{n} \psi(d_i) \cdot d_i}$$

Computing *S*-estimator

Cohen-Freue, V. G., Hernan Ortiz-Molina, H., Zamar, R. H. (2013):
A natural robustification of the ordinary instrumental variables estimator.
Biometrics 69, 641 - 650.

$$\hat{\beta}^{(S,n,\rho)} = \widehat{cov}_S^{-1}(X, X)\,\widehat{cov}_S(X, Y),$$

Computing S-estimator

Cohen-Freue, V. G., Hernan Ortiz-Molina, H., Zamar, R. H. (2013):
A natural robustification of the ordinary instrumental variables estimator.
Biometrics 69, 641 - 650.

$$\hat{\beta}^{(S,n,\rho)} = \widehat{cov}_S^{-1}(X,X)\,\widehat{cov}_S(X,Y),$$

of course to obtain $\widehat{cov}_S(X,Y)$, we calculate

$$\widehat{cov}_S^{-1}((X,Y)(X,Y))$$

and take the last column without the last coordinate.

Content

A shock and frustration - Engine Knock Data

Hettmansperger, T. P., S. J. Sheather (1992):
A Cautionary Note on the Method of Least Median Squares.
The American Statistician 46, 79–83.

Engine Knock Data $(n = 16, p = 4, h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
2	13.3	14.1	30	697	84.1
3	13.4	15.2	32	700	88.4
4	12.7	13.8	31	669	84.2
⋮	⋮	⋮	⋮	⋮	⋮
14	12.7	16.1	35	649	93.0
15	12.9	15.1	36	721	93.3
16	12.7	15.9	37	696	93.1

x_1 is spark timing x_2 air/fuel ratio
x_3 intake temperature x_4 exhaust temperature
y engine knock number

A shock and frustration - Engine Knock Data

Hettmansperger, T. P., S. J. Sheather (1992):
A Cautionary Note on the Method of Least Median Squares.
The American Statistician 46, 79–83.

Engine Knock Data $(n = 16, p = 4, h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
2	13.3	14.1	30	697	84.1
3	13.4	15.2	32	700	88.4
4	12.7	13.8	31	669	84.2
16	12.7	15.9	37	696	93.1

In fact they worked with two data sets.

x_1 is spark timing $\quad$ x_2 air/fuel ratio
x_3 intake temperature $\quad$ x_4 exhaust temperature
y engine knock number

A shock and frustration - Engine Knock Data

Hettmansperger, T. P., S. J. Sheather (1992):
A Cautionary Note on the Method of Least Median Squares.
The American Statistician 46, 79–83.

Engine Knock Data $(n = 16, p = 4, h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
2	13.3	$\boxed{14.1}$	30	697	84.1
3	13.4	15.2	32	700	88.4
4	12.7	13.8	31	669	84.2
16	12.7	15.9	37	696	93.1

In fact they worked with two data sets.

x_1 is spark timing x_2 air/fuel ratio
x_3 intake temperature x_4 exhaust temperature
y engine knock number

A shock and frustration - Engine Knock Data

Hettmansperger, T. P., S. J. Sheather (1992):
A Cautionary Note on the Method of Least Median Squares.
The American Statistician 46, 79–83.

Engine Knock Data $(n = 16, p = 4, h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
2	13.3	$\boxed{14.1}$	30	697	84.1
3	13.4	15.2	32	700	88.4
4	12.7	13.8	31	669	84.2
16	12.7	15.9	37	696	93.1

In fact they worked with two data sets.
Let's call these data "Correct".

x_1 is spark timing x_2 air/fuel ratio
x_3 intake temperature x_4 exhaust temperature
y engine knock number

A shock and frustration - Engine Knock Data

Hettmansperger, T. P., S. J. Sheather (1992):
A Cautionary Note on the Method of Least Median Squares.
The American Statistician 46, 79–83.

Engine Knock Data $(n = 16, p = 4, h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
2	13.3	15.1	30	697	84.1
3	13.4	15.2	32	700	88.4
4	12.7	13.8	31	669	84.2
16	12.7	15.9	37	696	93.1

In fact they worked with two data sets.

x_1 is spark timing x_2 air/fuel ratio
x_3 intake temperature x_4 exhaust temperature
y engine knock number

A shock and frustration - Engine Knock Data

Hettmansperger, T. P., S. J. Sheather (1992):
A Cautionary Note on the Method of Least Median Squares.
The American Statistician 46, 79–83.

Engine Knock Data $(n = 16, p = 4, h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
2	13.3	$\boxed{15.1}$	30	697	84.1
3	13.4	15.2	32	700	88.4
4	12.7	13.8	31	669	84.2
16	12.7	15.9	37	696	93.1

In fact they worked with two data sets.

Let's call these data "Damaged".

x_1 is spark timing x_2 air/fuel ratio
x_3 intake temperature x_4 exhaust temperature
y engine knock number

Engine Knock Data $(n = 16, p = 4, h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
16	12.7	15.9	37	696	93.1

This is the exact value of $\hat{\beta}^{(LTS,n,h)}$ (as # of models = $\binom{16}{11} = 4368$) !!

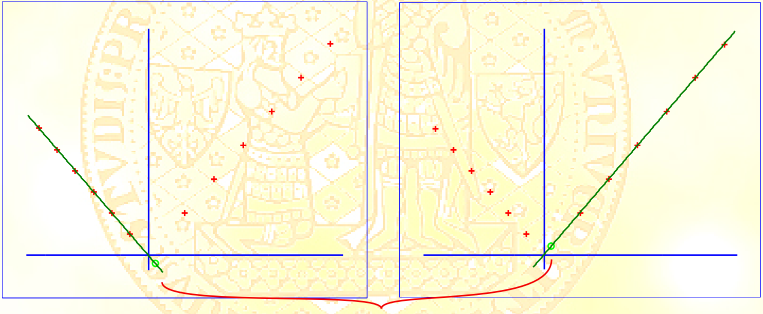
Data	Interc.	SPARK	AIR	INTK	EXHS.
Correct data ($x_{22} = 14.1$)	35.11	-0.028	2.949	0.477	-0.009
Damaged data ($x_{22} = 15.1$)	-88.7	4.72	1.06	1.57	0.068

x_1 is spark timing $\quad$ x_2 air/fuel ratio
x_3 intake temperature $\quad$ x_4 exhaust temperature
y engine knock number

An (academic) explanation by a shift of "inlier"

SENSITIVITY OF ANY HIGH-BREAKDOWN-POINT ESTIMATOR TO A SMALL CHANGE OF DATA

In both cases the model is for the majority of data

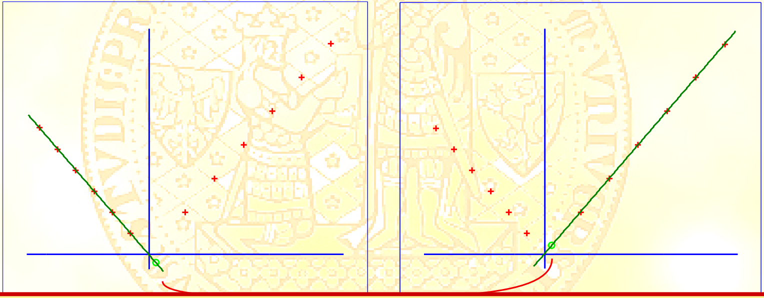


Notice: *The closer* the point (" o ") is to the y-axe,
the smaller shift causes the "switch" of the model.

An (academic) explanation by a shift of “inlier”

SENSITIVITY OF ANY HIGH-BREAKDOWN-POINT ESTIMATOR TO A SMALL CHANGE OF DATA

In both cases the model is for the majority of data



A very first idea - it is due to high breakdown point !

the smaller shift causes the “switch” of the model.

Returning to Engine Knock Data

Engine Knock Data $(n = 16, p = 4, h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
15	12.9	15.1	36	721	93.3
16	12.7	15.9	37	696	93.1

But $n = 16$ and $h = 11$!!

Data	Interc.	SPARK	AIR	INTK	EXHS.
Correct data ($x_{22} = 14.1$)	35.11	-0.028	2.949	0.477	-0.009
Damaged data ($x_{22} = 15.1$)	-88.7	4.72	1.06	1.57	0.068

x_1 is spark timing x_2 air/fuel ratio
x_3 intake temperature x_4 exhaust temperature
y engine knock number

Returning to Engine Knock Data

Engine Knock Data $(n = 16, p = 4, h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
16	12.7	15.9	37	696	93.1

But $n = 16$ and $h = 11$!!

Data	Interc.	SPARK	AIR	INTK	EXHS.

The breakdown point is about 30%, only !!!

x_1 is spark timing x_2 air/fuel ratio
x_3 intake temperature x_4 exhaust temperature
y engine knock number

Returning to Engine Knock Data

Engine Knock Data $(n = 16,\ p = 4,\ h = 11)$

c	x_1	x_2	x_3	x_4	y
1	13.3	13.9	31	697	84.4
16	12.7	15.9	37	696	93.1

But $n = 16$ and $h = 11$!!

Data	Interc.	SPARK	AIR	INTK	EXHS.

The breakdown point is about 30%, only !!!

The reason is our Godlike position - expressed by 0-1 objective function.

x_1 is spark timing x_2 air/fuel ratio
x_3 intake temperature x_4 exhaust temperature
y engine knock number

The Least Weighted Squares

Residuals $\forall \beta \in R \quad \rightarrow \qquad r_i(\beta) = Y_i - X_i'\beta$
Order statistics of squared residuals, i. e.

$$r^2_{(1)}(\beta) \leq r^2_{(2)}(\beta) \leq \ldots \leq r^2_{(n)}(\beta).$$

Definition

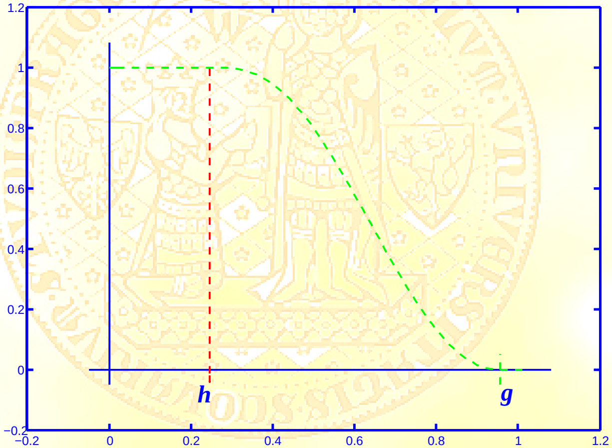
Let $w(u) : [0,1] \rightarrow [0,1], w(0) = 1$, be a (nonincreasing) weight function. Then

$$\hat{\beta}^{(LWS,n,w)} = \arg\min_{\beta \in R^p} \sum_{i=1}^n w\left(\frac{i-1}{n}\right) r^2_{(i)}(\beta)$$

will be called *the Least Weighted Squares (LWS)*.

(An example of weight function is on the next slide.)

Possible shape of weight function $w(r)$ - "one wing" of Tucekey's biweight



An algorithm for LWS

A

Find the plane through $p+1$ randomly selected observations.

Evaluate squared residuals of all observations, order them and compute $S(\hat{\beta}_{present}) = \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta)$.

Is $S(\hat{\beta}_{present})$ less than $S(\hat{\beta}_{past})$?

no → B

yes

Reorder the observations according to the order of squared residuals from the second step and establish *new* $\hat{\beta}_{present}$ just applying *WLS* on the reordered observations.

An algorithm for LWS

B

Was ℓ-times found the same model with minimal value of $S(\beta)$?

yes

no

Was already k-times repeated outer cycle ?

no → A

yes

As $\hat{\beta}^{(LWS,n,w)}$ we will assume $\beta \in R^p$ for which the functional $S(\beta)$ attained - through just described iterations - minimal value.

Disadvantages of LWS and the remedies

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta)$$

Disadvantages of LWS and the remedies

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^{2}(\beta)$$

Disadvantages of LWS and the remedies

$$\hat{\beta}^{(LWS,n,w)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta)$$

$$\boxed{\pi(\beta, j) = i \in \{1, 2, ..., n\} \quad \text{if} \quad r_j^2(\beta) = r_{(i)}^2(\beta)}$$

Disadvantages of LWS and the remedies

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta)$$

$$\boxed{\pi(\beta, j) = i \in \{1, 2, ..., n\} \quad \text{if} \quad r_j^2(\beta) = r_{(i)}^2(\beta)}$$

Disadvantages of LWS and the remedies

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\tfrac{i-1}{n}\right) r_{(i)}^2(\beta)$$

$$\boxed{\pi(\beta, j) = i \in \{1, 2, ..., n\} \quad \text{if} \quad r_j^2(\beta) = r_{(i)}^2(\beta)}$$

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{j=1}^{n} w\left(\frac{\pi(\beta, j) - 1}{n}\right) r_j^2(\beta)$$

Disadvantages of LWS and the remedies

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\tfrac{i-1}{n}\right) r_{(i)}^2(\beta)$$

$$\boxed{\pi(\beta, j) = i \in \{1, 2, ..., n\} \quad \text{if} \quad r_j^2(\beta) = r_{(i)}^2(\beta)}$$

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{j=1}^{n} w\left(\frac{\pi(\beta, j) - 1}{n}\right) r_j^2(\beta)$$

$$\frac{\pi(\beta, j) - 1}{n} = F_n(r_j^2(\beta))$$

Content

1. The history of high breakdown point estimation
   - Basic framework
   - The pursuit for high breakdown point estimator
   - Frustrations and rebirth - by the smooth depression of ...

2. S-weighted estimator
   - Combining S-estimator and the Least Weighted Squares
   - Asymptotics of SWE
   - SWE - how does it work ? The algorithm and patterns of results

S-weighted estimator

Residuals $\quad \forall \beta \in R \quad \rightarrow \quad r_i(\beta) = Y_i - X_i'\beta$

Order statistics of squared residuals, i. e.

$$r^2_{(1)}(\beta) \leq r^2_{(2)}(\beta) \leq ... \leq r^2_{(n)}(\beta).$$

S-weighted estimator

Residuals $\forall \beta \in R \quad \rightarrow \quad r_i(\beta) = Y_i - X_i'\beta$

Order statistics of squared residuals, i. e.

$$r^2_{(1)}(\beta) \leq r^2_{(2)}(\beta) \leq ... \leq r^2_{(n)}(\beta).$$

Definition

Let $w(u) : [0,1] \rightarrow [0,1], w(0) = 1$, be a (nonincreasing) weight function and ρ an objective (nondecreasing) function. Then

$$\hat{\beta}^{(SW,n,w,\rho)} = \arg\min_{\beta \in R^p} \left\{ \sigma : \sum_{i=1}^n w\left(\frac{i-1}{n}\right) \rho\left(\frac{r^2_{(i)}(\beta)}{\sigma^2}\right) = b \right\}$$

will be called *the S-weighted estimator (SWE)*.

Technical tricks (continued)

Defining the empirical d. f. of absolute values of residuals

$$F_{\beta}^{(n)}(r) = \frac{1}{n} \sum_{i=1}^{n} I\{|r_i(\beta)| < r\},$$

Technical tricks (continued)

Defining the empirical d. f. of absolute values of residuals

$$F_{\beta}^{(n)}(r) = \frac{1}{n}\sum_{i=1}^{n} I\left\{|r_i(\beta)| < r\right\},$$

we can show that $\hat{\beta}^{(SW,n,w,\rho)}$ is given as solution of

$$\boxed{\sum_{i=1}^{n} w\left(F^{(n)}_{\hat{\beta}^{(SW,n,w,\rho)}}\left(\frac{|r_i(\hat{\beta}^{(SW,n,w,\rho)})|}{\hat{\sigma}}\right)\right) X_i \psi\left(\frac{Y_i - X_i'\hat{\beta}^{(SW,n,w,\rho)}}{\hat{\sigma}}\right) = 0}$$

Technical tricks$_{(continued)}$

Defining the empirical d. f. of absolute values of residuals

$$F_{\beta}^{(n)}(r) = \frac{1}{n}\sum_{i=1}^{n} I\{|r_i(\beta)| < r\},$$

we can show that $\hat{\beta}^{(SW,n,w,\rho)}$ is given as solution of

$$\sum_{i=1}^{n} w\left(F^{(n)}_{\hat{\beta}^{(SW,n,w,\rho)}}\left(\frac{|r_i(\hat{\beta}^{(SW,n,w,\rho)})|}{\hat{\sigma}}\right)\right) X_i \psi\left(\frac{Y_i - X_i'\hat{\beta}^{(SW,n,w,\rho)}}{\hat{\sigma}}\right) = 0$$

and finally (I'll show details later, if enough time)

$$\sum_{i=1}^{n} w^*\left(F^{(n)}_{\hat{\beta}^{(SW,n,w,\rho)}}\left(\frac{|r_i(\hat{\beta}^{(SW,n,w,\rho)})|}{\hat{\sigma}}\right)\right) X_i \left(Y_i - X_i'\hat{\beta}^{(SW,n,w,\rho)}\right) = 0.$$

Kolmogorov - Smirnov

Finally, we employ

$$\forall(\varepsilon > 0) \;\; \exists(K_\varepsilon > 0, n_\varepsilon \in \mathcal{N}) \;\; \forall(n > n_\varepsilon)$$

$$\boxed{P\left(\left\{\omega \in \Omega : \sup_{v \in R^+} \sup_{\beta \in R^p} \sqrt{n}\left|F_\beta^{(n)}(v) - F_{n,\beta}(v)\right| < K_\varepsilon\right\}\right) > 1 - \varepsilon}$$

where $F_\beta^{(n)}(v)$ and $F_{n,\beta}(v)$ are empirical and theoretical d.f. of error term.

Content

Asymptotics of SWE

Conditions for consistency $\mathcal{C}1$

- $\{(X_i', e_i)'\}_{i=1}^{\infty}$ is sequence of independent, except of scale, identically distributed r. v.'s,
- d. f.'s of error term are absolutely continuous,
- $\exists\, q > 1 \;:\; E_{F_X} \|X\|^{2q} < \infty,$

Asymptotics of SWE

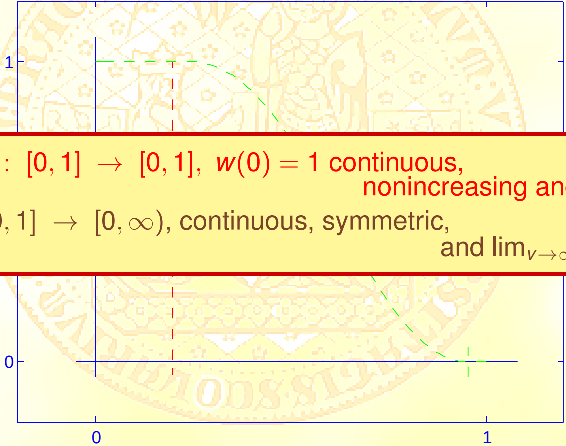
Conditions for consistency $\mathcal{C}1$

- $\{(X_i', e_i)'\}_{i=1}^{\infty}$ is sequence of independent, except of scale, identically distributed r. v.'s,
- d. f.'s of error term are absolutely continuous,
- $\exists\, q > 1 \ : \quad E_{F_X} \|X\|^{2q} < \infty$,
- there is the only one solution of the *identification condition*

$$(\beta - \beta^0)' \, E\left[w\left(F_\beta(|r(\beta)|)\right) \cdot X_1 \left(e - X_1' \left(\beta - \beta^0\right)\right)\right] = 0.$$

Asymptotics of SWE

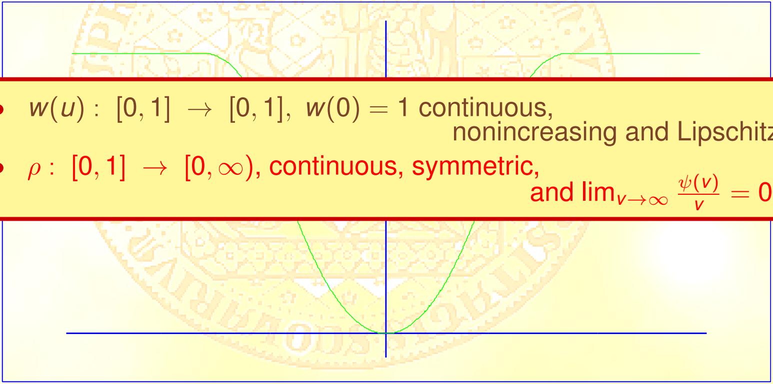
Conditions for consistency $\mathcal{C}1_{(continued)}$



- $w(u): \ [0,1] \ \rightarrow \ [0,1], \ w(0) = 1$ continuous, nonincreasing and Lipschitz,
- $\rho: \ [0,1] \ \rightarrow \ [0,\infty)$, continuous, symmetric, and $\lim_{v \rightarrow \infty} \frac{\psi(v)}{v} = 0$.

Asymptotics of SWE

Conditions for consistency $\mathcal{C}1_{(continued)}$



- $w(u): [0,1] \rightarrow [0,1],\ w(0) = 1$ continuous, nonincreasing and Lipschitz,
- $\rho: [0,1] \rightarrow [0,\infty)$, continuous, symmetric, and $\lim_{v\rightarrow\infty} \frac{\psi(v)}{v} = 0$.

Asymptotics of SWE

Consistency of the S-weighted estimator

Assertion:

Under Conditions $\mathcal{C}1$ $\hat{\beta}^{(SWE,n,w,\rho)}$ is consistent.

Proceedings of the 16th Conference on the Applied Stochastic Models, Data Analysis and Demographics 2015, 1031 - 1042.

Asymptotics of SWE

Strengthening the conditions a bit:

Conditions for $\sqrt{n}$-consistency $\mathcal{NC}1$

1. The density of error terms has bounded derivative.

Asymptotics of SWE

Strengthening the conditions a bit:

Conditions for $\sqrt{n}$-consistency $\mathcal{NC}1$

1. The density of error terms has bounded derivative.

Under Conditions $\mathcal{C}1$ and $\mathcal{NC}1$ $\hat{\beta}^{(SWE,n,w,\rho)}$ is $\sqrt{n}$-consistent.

Paper submitted to the proceedings of conference
STATISTICS AND DEMOGRAPHY 2015: THE LEGACY OF CORRADO GINI

Asymptotic of SWE

Conditions for asymptotic representation $\mathcal{AC}1$

1. The density of error term is positive in a neighborhood of zero.

Asymptotic of SWE

Conditions for asymptotic representation $\mathcal{AC}1$

1. The density of error term is positive in a neighborhood of zero.

Asymptotic representation of the S-weighted estimator

Under Conditions $\mathcal{C}1$, $\mathcal{NC}1$ and $\mathcal{AC}1$

$$\sqrt{n}\left(\hat{\beta}^{(SWE,n,w,\rho)} - \beta^0\right) = Q^{-1} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w\left(F_{\beta^0}(|e_i|)\right) \cdot X_i \psi(e_i) + o_p(1).$$

Content

Technical tricks *(continued)*

We can show that $\hat{\beta}^{(SW,n,w,\rho)}$ is given as solution of

$$\sum_{i=1}^{n} w\left(F^{(n)}_{\hat{\beta}^{(SW,n,w,\rho)}}\left(\frac{|r_i(\hat{\beta}^{(SW,n,w,\rho)})|}{\hat{\sigma}}\right)\right) X_i \psi\left(\frac{Y_i - X_i' \hat{\beta}^{(SW,n,w,\rho)}}{\hat{\sigma}}\right) = 0$$

Technical tricks *(continued)*

We can show that $\hat{\beta}^{(SW,n,w,\rho)}$ is given as solution of

$$\boxed{\sum_{i=1}^{n} w\left(F^{(n)}_{\hat{\beta}^{(SW,n,w,\rho)}}\left(\frac{|r_i(\hat{\beta}^{(SW,n,w,\rho)})|}{\hat{\sigma}}\right)\right) X_i \psi\left(\frac{Y_i - X_i'\hat{\beta}^{(SW,n,w,\rho)}}{\hat{\sigma}}\right) = 0}$$

$$\sum_{\mathcal{I}} w\left(F^{(n)}_{\hat{\beta}^{(SW,n,w,\rho)}}\left(\frac{|r_i(\hat{\beta}^{(SW,n,w,\rho)})|}{\hat{\sigma}}\right)\right) X_i \frac{\psi\left(\frac{Y_i - X_i'\hat{\beta}^{(SW,n,w,\rho)}}{\hat{\sigma}}\right)}{Y_i - X_i'\hat{\beta}^{(SW,n,w,\rho)}} \left(Y_i - X_i'\hat{\beta}^{(SW,n,w,\rho)}\right) = 0$$

with $\mathcal{I} = \left\{ i : Y_i - X_i'\hat{\beta}^{(SW,n,w,\rho)} \neq 0 \right\}$

Technical tricks$_{(continued)}$

We can show that $\hat\beta^{(SW,n,w,\rho)}$ is given as solution of

$$\sum_{i=1}^n w\left(F^{(n)}_{\hat\beta^{(SW,n,w,\rho)}}\left(\frac{|r_i(\hat\beta^{(SW,n,w,\rho)})|}{\hat\sigma}\right)\right) X_i \psi\left(\frac{Y_i - X_i'\hat\beta^{(SW,n,w,\rho)}}{\hat\sigma}\right) = 0$$

$$\sum_{\mathcal{I}} w\left(F^{(n)}_{\hat\beta^{(SW,n,w,\rho)}}\left(\frac{|r_i(\hat\beta^{(SW,n,w,\rho)})|}{\hat\sigma}\right)\right) X_i \frac{\psi\left(\frac{Y_i - X_i'\hat\beta^{(SW,n,w,\rho)}}{\hat\sigma}\right)}{Y_i - X_i'\hat\beta^{(SW,n,w,\rho)}} \left(Y_i - X_i'\hat\beta^{(SW,n,w,\rho)}\right) = 0$$

with $\mathcal{I} = \left\{i : Y_i - X_i'\hat\beta^{(SW,n,w,\rho)} \neq 0\right\}$

and finally

$$\sum_{i=1}^n w^*\left(F^{(n)}_{\hat\beta^{(SW,n,w,\rho)}}\left(\frac{|r_i(\hat\beta^{(SW,n,w,\rho)})|}{\hat\sigma}\right)\right) X_i \left(Y_i - X_i'\hat\beta^{(SW,n,w,\rho)}\right) = 0.$$

Technical tricks$_{(continued)}$

$$\sum_{i=1}^{n} w^{*}\left(F^{(n)}_{\hat{\beta}^{(SW,n,w,\rho)}}\left(\frac{|r_i(\hat{\beta}^{(SW,n,w,\rho)})|}{\hat{\sigma}}\right)\right) X_i \left(Y_i - X_i' \hat{\beta}^{(SW,n,w,\rho)}\right) = 0$$

Technical tricks$_{(continued)}$

$$\sum_{i=1}^{n} w^* \left(F^{(n)}_{\hat\beta^{(SW,n,w,\rho)}} \left(\frac{|r_i(\hat\beta^{(SW,n,w,\rho)})|}{\hat\sigma} \right) \right) X_i \left(Y_i - X_i' \hat\beta^{(SW,n,w,\rho)} \right) = 0$$

$$\hat\beta^{(SW,n,w,\rho)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} w^* \left(F^{(n)}_{\hat\beta^{(SW,n,w,\rho)}} \left(\frac{|r_i(\hat\beta^{(SW,n,w,\rho)})|}{\hat\sigma} \right) \right) \times \times \left(Y_i - X_i' \hat\beta^{(SW,n,w,\rho)} \right)^2$$

with

$$\hat\sigma^2 = \frac{\sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r^2_{(i)}(\hat\beta^{(SW,n,w,\rho)})}{\sum_{i=1}^{n} w\left(\frac{i-1}{n}\right)} .$$

An algorithm for S-weighted estimator

A

Find the plane through $p+1$ randomly selected observations.

Evaluate squared residuals of all observations, order them and compute $S(\hat{\beta}_{present}) = \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) \rho\left(\frac{r_{(i)}^2(\beta)}{\hat{\sigma}^2}\right)$.

Is $S(\hat{\beta}_{present})$ less than $S(\hat{\beta}_{past})$?

no

B

yes

Reorder the observations according to the order of squared residuals from the second step and establish *new* $\hat{\beta}_{present}$ just applying weighted M_ρ-estimator on the reordered observations.

An algorithm for *S*-weighted estimator

B

Was ℓ-times found the same model with minimal value of $S(\beta)$?

no

Was already k-times repeated outer cycle ?

no

A

yes

yes

As $\hat{\beta}^{(SW,n,w,\rho)}$ we will assume $\beta \in R^p$ for which the functional $S(\beta)$ attained - through just described iterations - minimal value.

Numerical study - comparison of OLS, LWS, S and SW-estimators

The framework:

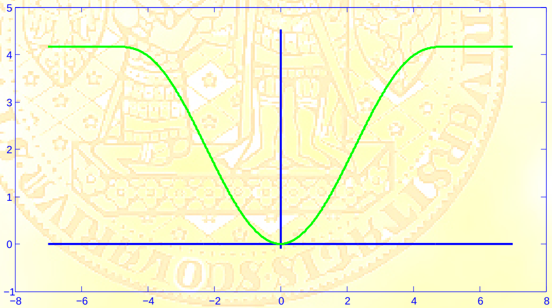
- 500 data sets, each data set contains 500 observations.

Numerical study - comparison of OLS, LWS, S and SW-estimators

***The framework*:**

- 500 data sets, each data set contains 500 observations.
- The constant c of Tukey's biweight function for *S*- and *SW*-estimator

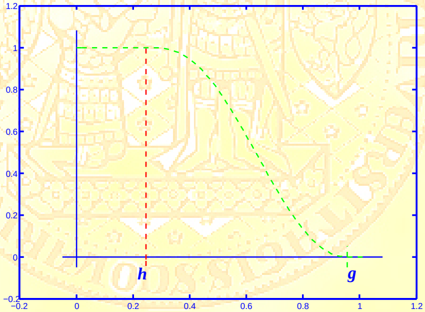
$$\rho_c(x) = \frac{x^2}{2} - \frac{x^4}{2 \cdot c^2} + \frac{x^6}{6 \cdot c^4} \quad \text{for} \quad |x| \leq c \quad \text{and} \quad \rho_c(x) = \frac{c^2}{6} \text{ otherwise,}$$



Numerical study - comparison of OLS, LWS, S and SW-estimators

The framework:

- 500 data sets, each data set contains 500 observations.
- The constant c of Tukey's biweight function for S- and SW-estimator and h and g of Tuckey-like weight function for LWS and S-weighted estimator are given in the heads of tables.



Numerical study - comparison of OLS, LWS, S and SW-estimators

The framework:

- 500 data sets, each data set contains 500 observations.
- The constant c of Tukey's biweight function for S- and SW-estimator and h and g of Tuckey-like weight function for LWS and S-weighted estimator are given in the heads of tables.
- The objective function for LWS was quadratic.

Numerical study - comparison of OLS, LWS, S and SW-estimators

The framework:

- 500 data sets, each data set contains 500 observations.
- The constant c of Tukey's biweight function for *S*- and *SW*-estimator and h and g of Tuckey-like weight function for *LWS* and *S*-weighted estimator are given in the heads of tables.
- The objective function for *LWS* was quadratic.
- The error terms are heteroscedastic ($0.5 \leq \sigma_i^2 \leq 5$) and independent from explanatory variables.

Numerical study - comparison of OLS, LWS, S and SW-estimators

The framework:

- 500 data sets, each data set contains 500 observations.
- The constant c of Tukey's biweight function for *S*- and *SW*-estimator and h and g of Tuckey-like weight function for *LWS* and *S*-weighted estimator are given in the heads of tables.
- The objective function for *LWS* was quadratic.
- The error terms are heteroscedastic $(0.5 \le \sigma_i^2 \le 5)$ and independent from explanatory variables.
- Exhibited are

$$\hat{\beta}_j^{(method)} = \frac{1}{500} \sum_{k=1}^{500} \hat{\beta}_j^{(method,k)}$$

and

$$\widehat{\mathrm{MSE}}\left(\hat{\beta}_j^{(method)}\right) = \frac{1}{500} \sum_{k=1}^{500} \left[\hat{\beta}_j^{(method,k)} - \beta_j^0\right]^2.$$

Numerical study - comparison of OLS, LWS, S and SW-estimators

The framework:

- 500 data sets, each data set contains 500 observations.
- The constant c of Tukey's biweight function for *S*- and *SW*-estimator and h and g of Tuckey-like weight function for *LWS* and *S*-weighted estimator are given in the heads of tables.
- The objective function for *LWS* was quadratic.
- The error terms are heteroscedastic ($0.5 \leq \sigma_i^2 \leq 5$) and independent from explanatory variables.
- Exhibited are

$$\hat{\beta}_j^{(method)} = \frac{1}{500} \sum_{k=1}^{500} \hat{\beta}_j^{(method,k)}$$

and

$$\widehat{\mathrm{MSE}}\left(\hat{\beta}_j^{(method)}\right) = \frac{1}{500} \sum_{k=1}^{500} \left[\hat{\beta}_j^{(method,k)} - \beta_j^0\right]^2.$$

Everything else will be clear from the heads of the next tables.

Numerical study - comparison of OLS, LWS, S and SW-estimators

TABLE 1

Data were not contaminated - but we did not know it.

Hence we took measures against an unknown level of contamination and then step by step accomodated all constants.

$h = 0.995, \quad g = 1, \quad c = 24, \quad n = 500$

True β^0	**1**	**- 2**	**-3**	**- 4**	**-5**
$\hat{\beta}^{(OLS)}_{(MSE)}$	$0.99_{(0.026)}$	$2.00_{(0.022)}$	$-3.01_{(0.021)}$	$3.99_{(0.023)}$	$-5.01_{(0.030)}$
$\hat{\beta}^{(LWS)}_{(MSE)}$	$0.99_{(0.025)}$	$2.00_{(0.020)}$	$-3.01_{(0.021)}$	$3.99_{(0.023)}$	$-5.01_{(0.029)}$
$\hat{\beta}^{(S)}_{(MSE)}$	$0.99_{(0.025)}$	$2.00_{(0.021)}$	$-3.01_{(0.021)}$	$3.99_{(0.022)}$	$-5.01_{(0.029)}$
$\hat{\beta}^{(SW)}_{(MSE)}$	$0.99_{(0.022)}$	$2.00_{(0.018)}$	$-3.01_{(0.018)}$	$4.00_{(0.021)}$	$-5.00_{(0.026)}$

Numerical study - comparison of OLS, LWS, S and SW-estimators

TABLE 2

The contamination was created by bad leverage points and outliers on the level 2+2%.

$$h = 0.92, \quad g = 0.945, \quad c = 7, \quad n = 500$$

True β^0	**1**	**2**	**- 3**	**4**	**- 5**
$\hat{\beta}^{(OLS)}_{(MSE)}$	$0.91_{(0.820)}$	$-2.98_{(30.578)}$	$3.96_{(56.626)}$	$-6.10_{(108.145)}$	$6.89_{(148.750)}$
$\hat{\beta}^{(LWS)}_{(MSE)}$	$0.99_{(0.022)}$	$1.97_{(0.022)}$	$-2.99_{(0.024)}$	$3.98_{(0.024)}$	$-4.99_{(0.023)}$
$\hat{\beta}^{(S)}_{(MSE)}$	$1.00_{(0.017)}$	$1.98_{(0.018)}$	$-2.99_{(0.023)}$	$3.97_{(0.025)}$	$-4.98_{(0.020)}$
$\hat{\beta}^{(SW)}_{(MSE)}$	$0.99_{(0.016)}$	$1.98_{(0.016)}$	$-2.99_{(0.018)}$	$3.98_{(0.021)}$	$-4.99_{(0.016)}$

Numerical study - comparison of OLS, LWS, S and SW-estimators

TABLE 3

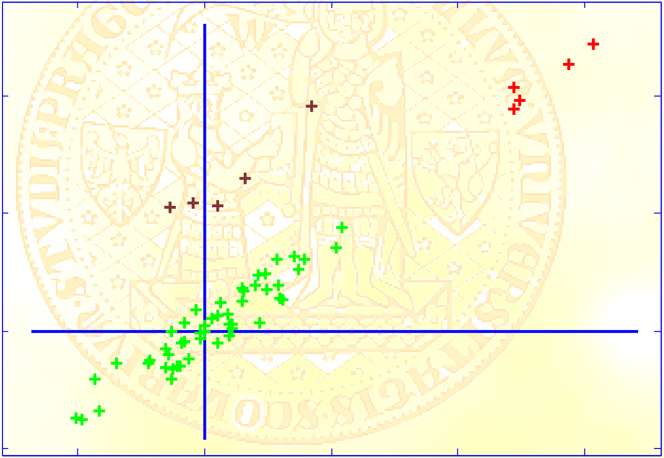
The contamination was created by outliers on the level 5%.

Data contain good leverage points on the level 5%.

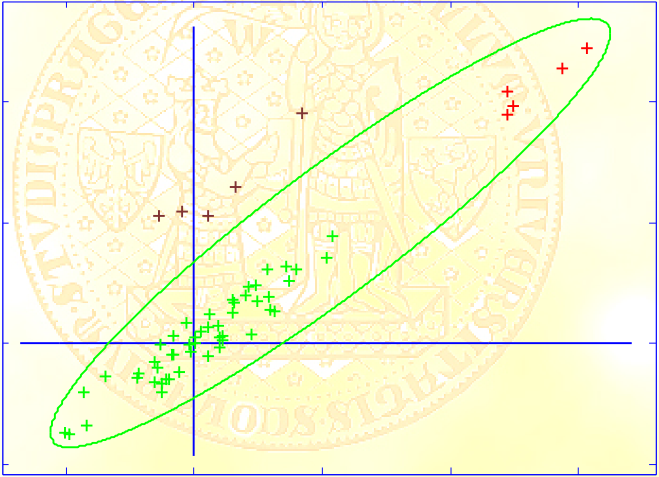
$h = 0.92, \quad g = 0.945, \quad c = 7, \quad n = 500$

True β^0	**1**	**2**	**- 3**	**4**	**- 5**
$\hat{\beta}^{(OLS)}_{(MSE)}$	$0.66_{(0.306)}$	$1.95_{(0.005)}$	$-2.94_{(0.005)}$	$3.94_{(0.006)}$	$-4.91_{(0.012)}$
$\hat{\beta}^{(LWS)}_{(MSE)}$	$1.00_{(0.021)}$	$1.99_{(0.001)}$	$-2.99_{(0.001)}$	$4.00_{(0.001)}$	$-4.99_{(0.001)}$
$\hat{\beta}^{(S)}_{(MSE)}$	$0.98_{(0.025)}$	$1.97_{(0.016)}$	$-2.98_{(0.027)}$	$3.96_{(0.022)}$	$-4.95_{(0.025)}$
$\hat{\beta}^{(SW)}_{(MSE)}$	$1.00_{(0.016)}$	$1.99_{(0.001)}$	$-2.99_{(0.001)}$	$4.00_{(0.001)}$	$-4.99_{(0.001)}$

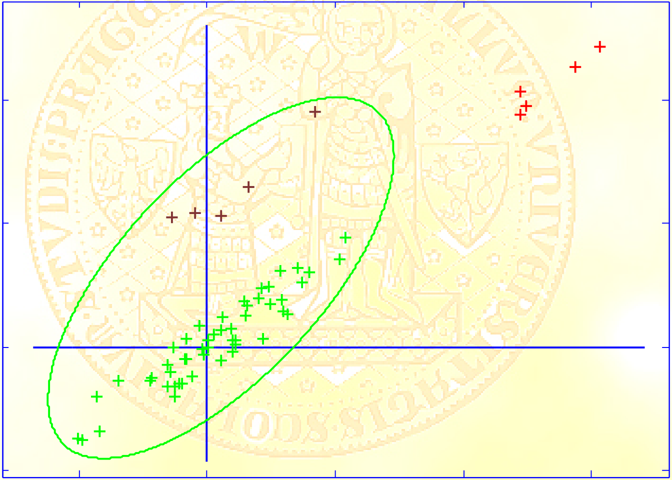
Topology of data with good leverage points and some outliers



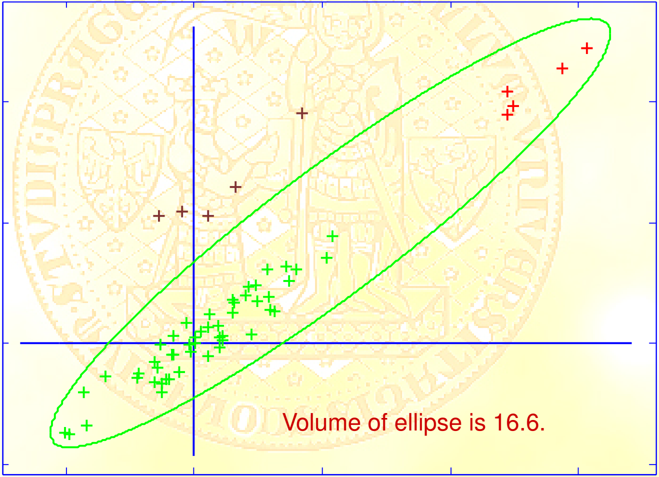
An attempt to find minimal volume with given number of points



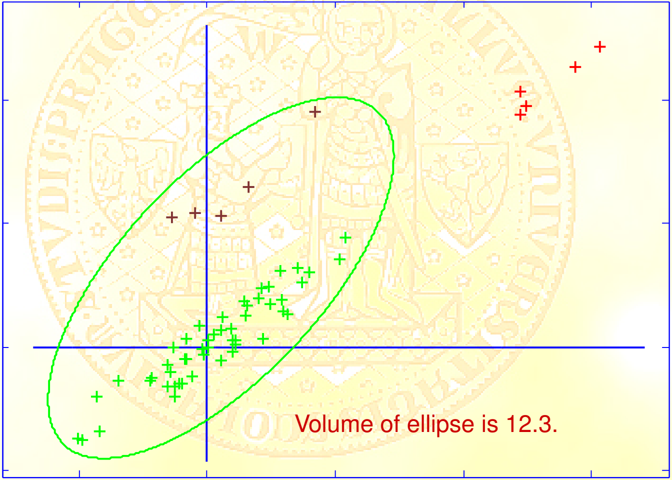
Estimate of scale is smaller but still not minimal



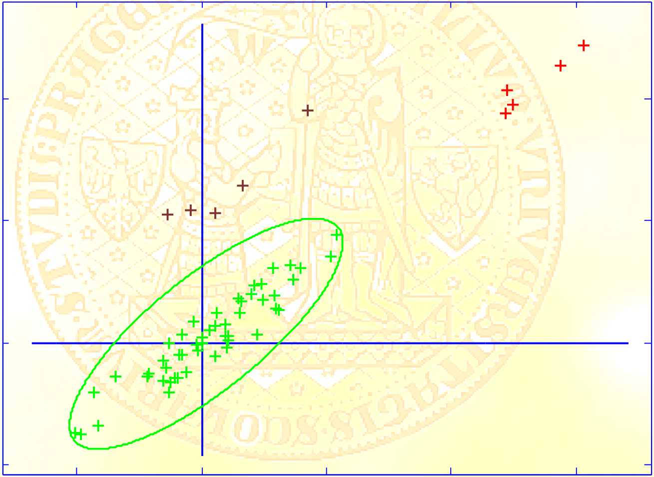
An attempt to find minimal volume with given number of points



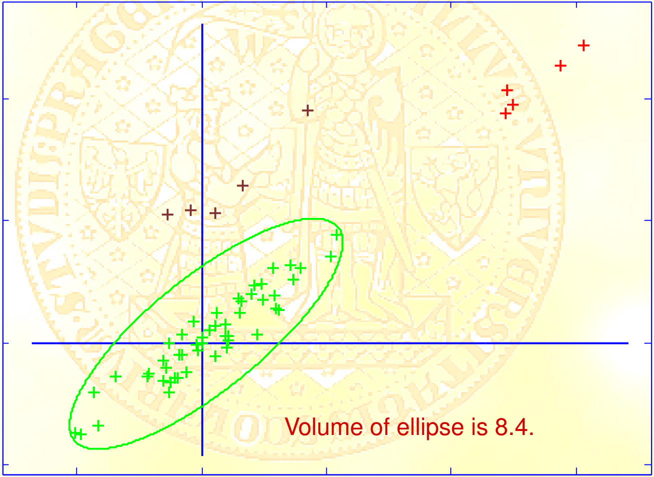
Estimate of scale is smaller but still not minimal



Estimate of scale is smallest but some information is lost



Estimate of scale is smallest but some information is lost

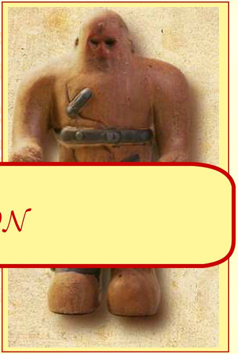


In compliance with JS @JŠ ...

Až půjdete někdy kolem, zastavte se na chvilku ...

(When You'll go around, stop for a moment...)

In compliance with JS @JŠ ...



THANKS FOR ATTENTION

Až půjdete někdy kolem, zastavte se na chvilku ...

(When You'll go around, stop for a moment...)

Charles University

The emblem of

Charles University

- the foundation documents were symbolically
and evidently humbly handed over

by Charles the IV., the Emperor of Roman Empire,
so probably the most powerfull man of those days,

to the representative of Higher Power, the Saint Venceslav.

THANKS FOR ATTENTION

once again