ICPE research meeting

Modern methods for robust regression

Jorge Tendeiro

14 January 2014

Literature

Presentation based on the book:

> *Andersen, R. (2008). Modern methods for robust regression. Sage University Paper Series QASS. ("Little green book" # 152)*

(Nearly) all R code that I used comes with the book and was downloaded from Sage.

Overview

Introduction

“Modern” regression

- (OLS — Ordinary Least Squares) regression:
  One of the most widely used statistical methods in social sciences.
- However, we will see that OLS regression does have limitations.
- “Modern” regression methods can be seen as improvements/ alternatives to the usual regression model.
- **Robust regression** is one such “modern” method.

Robust regression

The term *robust* has multiple interpretations in estimation frameworks:

Robust regression

The term *robust* has multiple interpretations in estimation frameworks:

- Robustness of validity: Resistance of the estimator to unusual observations.

  A *robust* estimator should not suffer big changes when small changes are made to the data.

Robust regression

The term *robust* has multiple interpretations in estimation frameworks:

- Robustness of validity: Resistance of the estimator to unusual observations.

  A *robust* estimator should not suffer big changes when small changes are made to the data.

- Robustness of efficiency: Resistance of the estimator to violations of underlying distributional assumptions.

  A *robust* estimator should keep high precision (i.e., small SEs) when distributional assumptions are violated.

Robust regression

The term *robust* has multiple interpretations in estimation frameworks:

- Robustness of validity: Resistance of the estimator to unusual observations.

  A *robust* estimator should not suffer big changes when small changes are made to the data.

- Robustness of efficiency: Resistance of the estimator to violations of underlying distributional assumptions.

  A *robust* estimator should keep high precision (i.e., small SEs) when distributional assumptions are violated.

We will mostly focus on robustness of validity in regression.

Limitations of OLS regression: Example

Jasso, G. (1985). Marital coital frequency and the passage of time: Estimating the separate effects of spouses' ages and marital duration, birth and marriage cohorts, and period influences. *American Sociological Review*, *50*(2), 224-41. doi:10.2307/2095411

- Jasso (1985) found that wife's age had a positive effect on the monthly coital frequency of married couples, after controlling for period and cohort effects.

Limitations of OLS regression: Example

Jasso, G. (1985). Marital coital frequency and the passage of time: Estimating the separate effects of spouses' ages and marital duration, birth and marriage cohorts, and period influences. *American Sociological Review*, *50*(2), 224-41. doi:10.2307/2095411

- Jasso (1985) found that wife's age had a positive effect on the monthly coital frequency of married couples, after controlling for period and cohort effects.
- Kahn and Udry (1986) questioned her findings:
  - They found four 'miscodes' in the dataset.
  - They found four additional outliers using model diagnostics.
  - They claim Jasso failed to consider a relevant interaction effect (length of marriage by wife's age): Model misspecification.

Limitations of OLS regression: Example

Jasso, G. (1985). Marital coital frequency and the passage of time: Estimating the separate effects of spouses' ages and marital duration, birth and marriage cohorts, and period influences. *American Sociological Review*, *50*(2), 224-41. doi:10.2307/2095411

- Jasso (1985) found that wife's age had a positive effect on the monthly coital frequency of married couples, after controlling for period and cohort effects.
- Kahn and Udry (1986) questioned her findings:
  - They found four 'miscodes' in the dataset.
  - They found four additional outliers using model diagnostics.
  - They claim Jasso failed to consider a relevant interaction effect (length of marriage by wife's age): Model misspecification.

Total sample size: 2062.

Limitations of OLS regression: Example

	Model 1	Model 2
Period	-0.72^{***}	-0.67^{***}
Log Wife's Age	27.61^{**}	13.56
Log Husband's Age	-6.43	7.87
Log Marital Duration	-1.50^{***}	-1.56^{***}
Wife Pregnant	-3.71^{***}	-3.74^{***}
Child Under 6	-0.56^{**}	-0.68^{***}
Wife Employed	0.37	0.23
Husband Employed	-1.28^{**}	-1.10^{**}
R^2	.0475	.0612
n	2062	2054

Note. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Model 1: Jasso's (1985) original model.

Model 2: Kahn and Udry's (1986) model excluding 4 miscodes and 4 outliers.

Limitations of OLS regression: Example

Some conclusions:

- A small number of unusual observations can have a large effect on the estimated regression coefficients.

Limitations of OLS regression: Example

Some conclusions:

- A small number of unusual observations can have a large effect on the estimated regression coefficients.
- Large samples are not immune to this problem.
  (Previous example: $8/2062 = 0.39\%$ of data.)

Limitations of OLS regression: Example

Some conclusions:

- A small number of unusual observations can have a large effect on the estimated regression coefficients.
- Large samples are not immune to this problem.
  (Previous example: $8/2062 = 0.39\%$ of data.)
- Using diagnostic tools to uncover potential problems is a crucial (and often disregarded) analysis step.

Limitations of OLS regression: Example

Some conclusions:

- A small number of unusual observations can have a large effect on the estimated regression coefficients.
- Large samples are not immune to this problem.
  (Previous example: $8/2062 = 0.39\%$ of data.)
- Using diagnostic tools to uncover potential problems is a crucial (and often disregarded) analysis step.
- The decision on what to do when influential observations are found should be based on substantive knowledge
  (i.e., no one-way-out solution exists).

Important background

Properties of estimators

Assessing whether an estimator is *robust* requires checking several mathematical properties.

Notation:

- θ population parameter that we intend to estimate
- T estimator for θ
- Y sample ($Y = (y_1, \ldots, y_n)$)
- $\widehat{\theta}$ estimate of θ ($T(Y) = \widehat{\theta}$)
- n sample size

Properties of estimators

1. Bias: *Does the estimator give, on average, the desired parameter?*

$$\text{bias} = E[\widehat{\theta} - \theta]$$

Properties of estimators

1. Bias: *Does the estimator give, on average, the desired parameter?*

$$\text{bias} = E[\widehat{\theta} - \theta]$$

2. Consistency: *Does the estimator converge to the parameter as $n \to \infty$?*

$$\lim_{n\to\infty} MSE(\widehat{\theta}) = \lim_{n\to\infty} E[(\widehat{\theta} - \theta)^2] = 0$$

Properties of estimators

1. Bias: *Does the estimator give, on average, the desired parameter?*

$$\text{bias} = E[\widehat{\theta} - \theta]$$

2. Consistency: *Does the estimator converge to the parameter as $n \to \infty$?*

$$\lim_{n\to\infty} MSE(\widehat{\theta}) = \lim_{n\to\infty} E[(\widehat{\theta} - \theta)^2] = 0$$

3. Breakdown point: *Global* measure of the resistance of an estimator

$$BDP(T, Y) = \min\left\{\frac{m}{n} : \sup_{Y'_m} \|T(Y'_m) - T(Y)\| \text{ is infinite}\right\},$$

where Y'_m is any sample derived from Y by replacing m of its n observations with arbitrary values.

$0 \leq BDP \leq .50$, the larger the better.

Properties of estimators

4. Influence function (IF): *Local* measure of the resistance of an estimator.
   Ideal: Having a bounded influence function, implying that one single observation has limited influence on the estimator.

Properties of estimators

4. Influence function (IF): *Local* measure of the resistance of an estimator.
   Ideal: Having a bounded influence function, implying that one single observation has limited influence on the estimator.

5. Relative efficiency: Ratio of *MSE*'s of the estimator with *smallest* *MSE* (say T_{Opt}) and estimator T

$$\text{Relative efficiency} = \frac{MSE(T_{\text{Opt}})}{MSE(T)}$$

$0 \leq \text{Rel. Effic.} \leq 1$, the larger the better.

Measures of location

Measure of location: Quantity that characterizes a *position* in a distribution

Statistic	Formula	*BDP*	*IF*	In current use in robust regression
Mean	$\overline{y} = \frac{\sum_i y_i}{n}$	0	Unbounded	No
α-trimmed mean	$\overline{y}_t = \frac{y_{(g+1)} + \cdots + y_{(n-g)}}{n-2g}$	α	Bounded	**Yes**
Median	$M = Q_{.50}$	.50	Bounded	**Yes**

Measures of scale

Measure of scale: Quantity that characterizes a *spread* of a distribution

Statistic	Formula	*BDP*	*IF*	In current use in robust regression
Standard deviation	$s_y = \sqrt{\frac{\sum_i (y_i - \bar{y})^2}{n-1}}$	0	Unbounded	No
Mean deviation from the mean	$MD = \frac{\sum_i \lvert y_i - \bar{y} \rvert}{n}$	0	Unbounded	No
Mean deviation from the median	$MDM = \frac{\sum_i \lvert y_i - M \rvert}{n}$	0	Unbounded	No
Interquartile range	$IQR = Q_{.75} - Q_{.25}$	.25	Bounded	**Not so often**
Median absolute deviation	$MAD = \text{median} \lvert y_i - M \rvert$	.50	Bounded	**Yes**

Robustness, resistance, and OLS regression

Classification of ‘unusual’ observations

Observation	Definition	Does it affect regression estimates?

Classification of ‘unusual’ observations

Observation	Definition	Does it affect regression estimates?
Univariate outlier	Outlier of x or y (unconditional on each other)	Not necessarily

Classification of ‘unusual’ observations

Observation	Definition	Does it affect regression estimates?
Univariate outlier	Outlier of x or y (unconditional on each other)	Not necessarily
Regression outlier	Unusual y value for a given x	Not necessarily (A)

Classification of 'unusual' observations

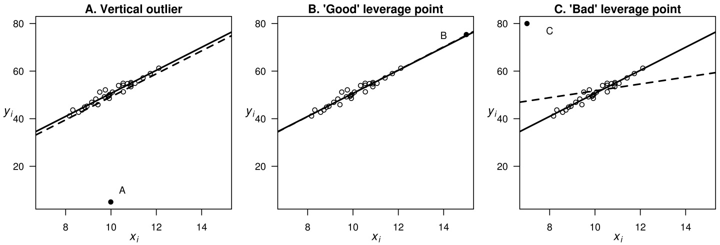
Observation	Definition	Does it affect regression estimates?
Univariate outlier	Outlier of x or y (unconditional on each other)	Not necessarily
Regression outlier	Unusual y value for a given x	Not necessarily (A)
Observation with leverage	Unusual x value	Not necessarily (B)

Classification of ‘unusual’ observations

Observation	Definition	Does it affect regression estimates?
Univariate outlier	Outlier of x or y (unconditional on each other)	Not necessarily
Regression outlier	Unusual y value for a given x	Not necessarily (A)
Observation with leverage	Unusual x value	Not necessarily (B)
Observation with influence	Regression estimates change greatly with/without them	Yes (C)

Classification of 'unusual' observations

Observation	Definition	Does it affect regression estimates?
Univariate outlier	Outlier of x or y (unconditional on each other)	Not necessarily
Regression outlier	Unusual y value for a given x	Not necessarily (A)
Observation with leverage	Unusual x value	Not necessarily (B)
Observation with influence	Regression estimates change greatly with/without them	Yes (C)



Classification of 'unusual' observations

Conclusion:

- Being x-discrepant (leverage) or y-discrepant (regression outlier) is not sufficient to identify influential observations.

Classification of 'unusual' observations

Conclusion:

- Being x-discrepant (leverage) or y-discrepant (regression outlier) is not sufficient to identify influential observations.
- However, it is a combination of both x- and y- discrepancies that determines the influence of an observation.

Classification of ‘unusual’ observations

Conclusion:

- Being x-discrepant (leverage) or y-discrepant (regression outlier) is not sufficient to identify influential observations.
- However, it is a combination of both x- and y- discrepancies that determines the influence of an observation.

Important note:

Observations with high leverages that follow the main regression trend help decreasing the SE of the estimates! ▸ Plot B

$$SE(b) = \frac{s_e}{\sqrt{\sum_i (x_i - \bar{x})^2}}$$

Classification of ‘unusual’ observations

Conclusion:

- Being x-discrepant (leverage) or y-discrepant (regression outlier) is not sufficient to identify influential observations.
- However, it is a combination of both x- and y- discrepancies that determines the influence of an observation.

Important note:
Observations with high leverages that follow the main regression trend help decreasing the SE of the estimates! Plot B

$$SE(b) = \frac{s_e}{\sqrt{\sum_i (x_i - \bar{x})^2}}$$

How do we identify regression outliers, observations with leverage, and observations with influence?

Identifying ‘unusual’ observations

$n =$ number of observations ($i = 1, \ldots, n$)
$k =$ number of predictors ($j = 1, \ldots, k$)

Detecting...	Use...	Rule of thumb? Flag if...	Test?	Plot?

Identifying ‘unusual’ observations

$n =$ number of observations ($i = 1, \ldots, n$)
$k =$ number of predictors ($j = 1, \ldots, k$)

Detecting…	Use…	Rule of thumb? Flag if…	Test?	Plot?
Leverage	Hat values h_i	$> 2(k+1)/n$ [a]	—	Index plot

a. For large samples. Use $h_i > 3(k+1)/n$ for small samples.

Identifying ‘unusual’ observations

$n =$ number of observations $(i = 1, \ldots, n)$
$k =$ number of predictors $(j = 1, \ldots, k)$

Detecting...	Use...	Rule of thumb? Flag if...	Test?	Plot?
Leverage	Hat values h_i	$> 2(k+1)/n$ [a]	—	Index plot
Regression outliers	Studentized residuals	$\lvert \cdot \rvert \geq 2$	Yes[b]	QQ-plot

a. For large samples. Use $h_i > 3(k+1)/n$ for small samples.
b. Controlling for capitalization by chance is required.

Identifying 'unusual' observations

n = number of observations ($i = 1, \ldots, n$)
k = number of predictors ($j = 1, \ldots, k$)

Detecting...	Use...	Rule of thumb? Flag if...	Test?	Plot?
Leverage	Hat values h_i	$> 2(k+1)/n$ [a]	—	Index plot
Regression outliers	Studentized residuals	$\lvert \cdot \rvert \geq 2$	Yes[b]	QQ-plot
Influence	DFBETAs	$\lvert \cdot \rvert \geq 2/\sqrt{n}$	—	Index plot
	Cook's D	$> .5$ [c] $> 4/(n-k-1)$ [d]	—	Index plot

a. For large samples. Use $h_i > 3(k+1)/n$ for small samples.
b. Controlling for capitalization by chance is required.
c. According to Cook and Weisberg (1999).
d. According to Fox (1997).

Identifying ‘unusual’ observations: Example

Weakliem, D.L., Andersen, R., & Heath, A.F. (2005). By popular demand: The effect of public opinion on income quality. *Comparative Sociology*, *4*(3), 261-284. doi:10.1163/156913305775010124

Identifying 'unusual' observations: Example

Weakliem, D.L., Andersen, R., & Heath, A.F. (2005). By popular demand: The effect of public opinion on income quality. *Comparative Sociology*, *4*(3), 261-284. doi:10.1163/156913305775010124

We will focus on a subset of the original data:

- Sample: $n = 26$ countries with democracies < 10 years.

Identifying 'unusual' observations: Example

Weakliem, D.L., Andersen, R., & Heath, A.F. (2005). By popular demand: The effect of public opinion on income quality. *Comparative Sociology*, *4*(3), 261-284. doi:10.1163/156913305775010124

We will focus on a subset of the original data:

- Sample: $n = 26$ countries with democracies < 10 years.
- Dependent variable:
  - Secpay: Mean country score on public opinion about pay inequality
    (between 0 and 1; large values reflect opinions favoring equality).

Identifying ‘unusual’ observations: Example

Weakliem, D.L., Andersen, R., & Heath, A.F. (2005). By popular demand: The effect of public opinion on income quality. *Comparative Sociology*, *4*(3), 261-284. doi:10.1163/156913305775010124

We will focus on a subset of the original data:

- Sample: $n = 26$ countries with democracies < 10 years.
- Dependent variable:
  - Secpay: Mean country score on public opinion about pay inequality
    (between 0 and 1; large values reflect opinions favoring equality).
- Predictors:
  - Gini: Income inequality
    (between 0 = ‘perfect equality’ and 1 = ‘perfect inequality’).
  - GDP: Per capita gross domestic product (scaled).

Identifying 'unusual' observations: Example

Goal: Estimate the model

$$\widehat{\text{Secpay}} = B_0 + B_1\text{Gini} + B_2\text{GDP}$$

Are there influential points?

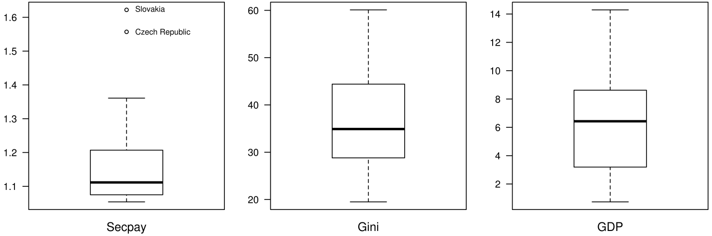
Identifying ‘unusual’ observations: Example

Goal: Estimate the model

$$\widehat{\text{Secpay}} = B_0 + B_1\text{Gini} + B_2\text{GDP}$$

Are there influential points?

First: Looking for univariate outliers:



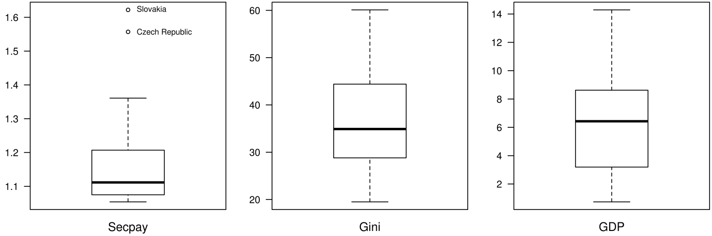
Identifying 'unusual' observations: Example

Goal: Estimate the model

$$\widehat{\text{Secpay}} = B_0 + B_1 \text{Gini} + B_2 \text{GDP}$$

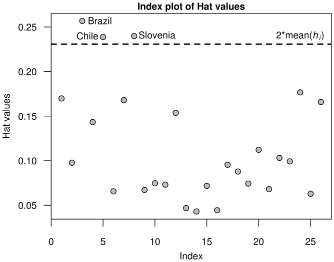
Are there influential points?

First: Looking for univariate outliers:

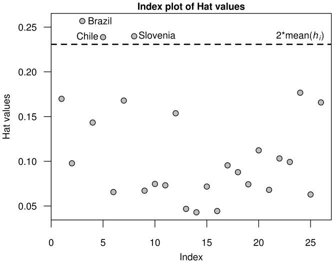


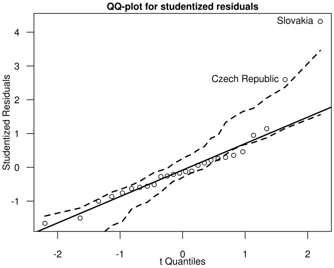
Next: Looking for leverage, regression outliers, and influence.

Identifying ‘unusual’ observations: Example

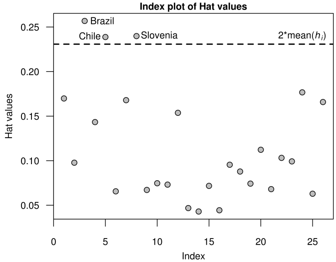


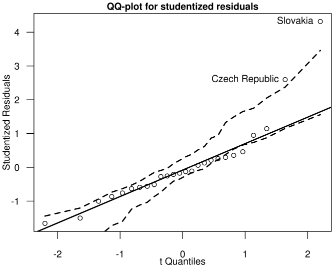
Identifying ‘unusual’ observations: Example



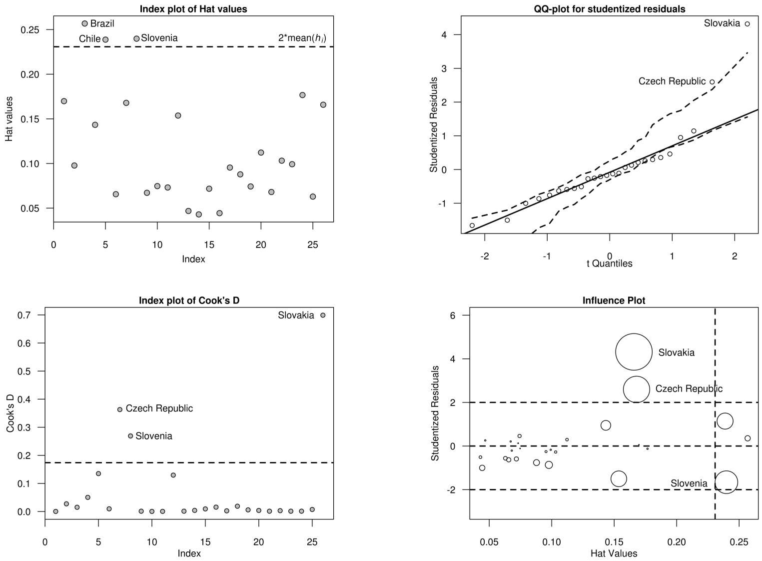


Identifying ‘unusual’ observations: Example





Identifying 'unusual' observations: Example



Identifying ‘unusual’ observations: Example

Conclusion:

- Slovakia and Czech Republic seem to have a large influence on the estimation of the regression coefficients.

Identifying 'unusual' observations: Example

Conclusion:

- Slovakia and Czech Republic seem to have a large influence on the estimation of the regression coefficients.
- This can be confirmed:

	OLS (all countries)		OLS (omitting cz and sk)	
	b	SE	*b*	SE
Intercept	.028	.128	−.107*	.058
Gini	.00074	.0028	.00527***	.0013
GDP	.0175**	.0079	.0063	.0037
s	.138		.0602	
R^2	.175		.4622	
n	26		24	

Note. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Identifying ‘unusual’ observations: Example

Conclusion:

- Slovakia and Czech Republic seem to have a large influence on the estimation of the regression coefficients.
- This can be confirmed:

	OLS (all countries)		OLS (omitting cz and sk)	
	b	SE	*b*	SE
Intercept	.028	.128	−.107*	.058
Gini	.00074	.0028	.00527***	.0013
GDP	.0175**	.0079	.0063	.0037
s	.138		.0602	
R^2	.175		.4622	
n	26		24	

Note. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Let’s now look into better alternatives to OLS!

Robust regression for the linear model

Various robust regression models

Type	Estimator	Breakdown Point	Bounded Infl. Func.	Asymptotic Efficiency
—	OLS	0	No	100%

Various robust regression models

Type	Estimator	Breakdown Point	Bounded Infl. Func.	Asymptotic Efficiency
—	OLS	0	No	100%
	LAV (Least Absolute Values)	0	Yes	64%
L-Estimators	LMS (Least Median of Squares)	.5	Yes	37%
	LTS (Least Trimmed Squares)	.5	Yes	8%

Various robust regression models

Type	Estimator	Breakdown Point	Bounded Infl. Func.	Asymptotic Efficiency
—	OLS	0	No	100%
L-Estimators	LAV (Least Absolute Values)	0	Yes	64%
	LMS (Least Median of Squares)	.5	Yes	37%
	LTS (Least Trimmed Squares)	.5	Yes	8%
R-Estimators	Bounded influence estimator	$< .2$	Yes	90%

Various robust regression models

Type	Estimator	Breakdown Point	Bounded Infl. Func.	Asymptotic Efficiency
—	OLS	0	No	100%
L-Estimators	LAV (Least Absolute Values)	0	Yes	64%
	LMS (Least Median of Squares)	.5	Yes	37%
	LTS (Least Trimmed Squares)	.5	Yes	8%
R-Estimators	Bounded influence estimator	$< .2$	Yes	90%
M-Estimators	*M*-estimates (Huber, biweight)	0	No	95%
	GM-estimates (Mal.& Schw.)	$1/(p+1)$	Yes	95%
	GM-estimates (S1S)	.5	Yes	95%

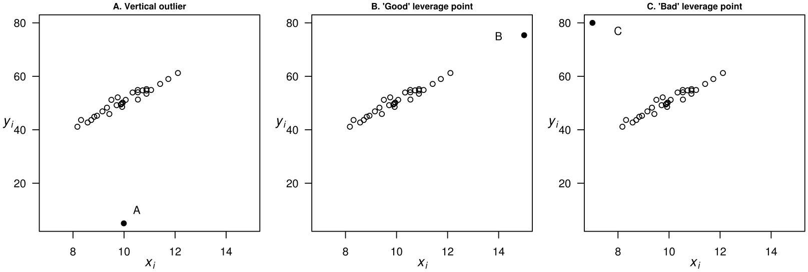
Various robust regression models

Type	Estimator	Breakdown Point	Bounded Infl. Func.	Asymptotic Efficiency
—	OLS	0	No	100%
L-Estimators	LAV (Least Absolute Values)	0	Yes	64%
	LMS (Least Median of Squares)	.5	Yes	37%
	LTS (Least Trimmed Squares)	.5	Yes	8%
R-Estimators	Bounded influence estimator	$< .2$	Yes	90%
M-Estimators	*M*-estimates (Huber, biweight)	0	No	95%
	GM-estimates (Mal.& Schw.)	$1/(p+1)$	Yes	95%
	GM-estimates (S1S)	.5	Yes	95%
S-Estimators	*S*-estimates	.5	Yes	33%
	GS-estimates	.5	Yes	67%

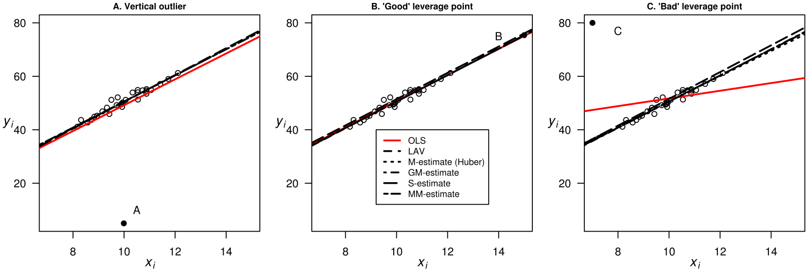
Various robust regression models

Type	Estimator	Breakdown Point	Bounded Infl. Func.	Asymptotic Efficiency
—	OLS	0	No	100%
L-Estimators	LAV (Least Absolute Values)	0	Yes	64%
	LMS (Least Median of Squares)	.5	Yes	37%
	LTS (Least Trimmed Squares)	.5	Yes	8%
R-Estimators	Bounded influence estimator	$< .2$	Yes	90%
M-Estimators	*M*-estimates (Huber, biweight)	0	No	95%
	GM-estimates (Mal.& Schw.)	$1/(p+1)$	Yes	95%
	GM-estimates (S1S)	.5	Yes	95%
S-Estimators	*S*-estimates	.5	Yes	33%
	GS-estimates	.5	Yes	67%
MM-Estimators	*MM*-estimates	.5	Yes	95%

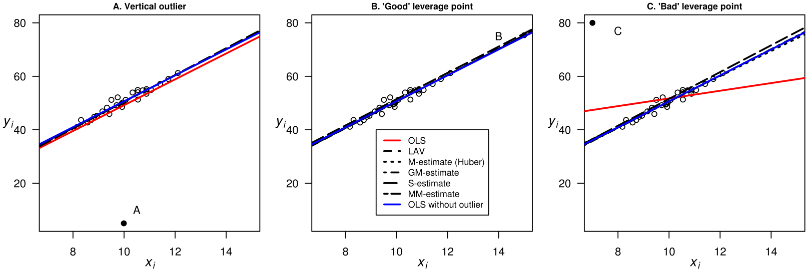
Example: Simulated data



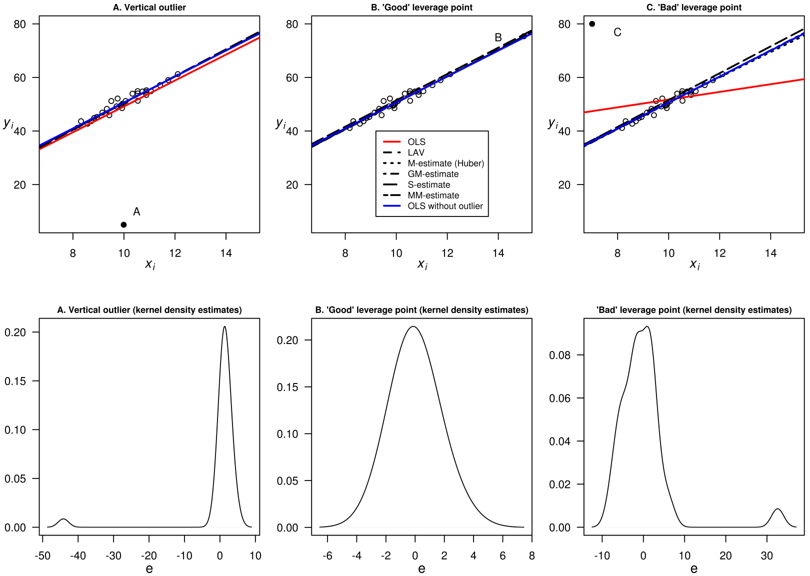
Example: Simulated data



Example: Simulated data



Example: Simulated data



Example: Public opinion about pay inequality

$$\widehat{\text{Secpay}} = B_0 + B_1\text{Gini} + B_2\text{GDP}$$

	OLS (all)	OLS (~~cz~~, ~~sk~~)	*L*-Est. (LAV)	*M*-Est. (Huber)	*M*-Est. (Biweight)	*MM*-Est.
Intercept	.0283	−.1069	−.0791	−.0632	−.0905	−.0978
Gini	.0007	.0053	.0045	.0039	.0049	.0051
GDP	.0175	.0063	.0059	.0089	.0052	.0057

- All robust regression methods give similar results.
- Once more, OLS with outliers removed gives similar results to robust regression models.

Using robust regression as regression diagnostics

Common statistics measuring influence of observations (e.g., Cook's distance) are not robust against unusual obervations.

Using robust regression as regression diagnostics

Common statistics measuring influence of observations (e.g., Cook's distance) are not robust against unusual obervations.

Q: Why?

Using robust regression as regression diagnostics

Common statistics measuring influence of observations (e.g., Cook's distance) are not robust against unusual obervations.

Q: Why?
A: Because they rely on sample mean and (co)variances, which are not robust themselves.

In particular, Cook's D suffers from a masking effect:

> *A masking effect occurs when groups of influential observations mask the influence of each other.*
> *(Rousseeuw & van Zomeren, 1990)*

Using robust regression as regression diagnostics

Common statistics measuring influence of observations (e.g., Cook's distance) are not robust against unusual obervations.

Q: Why?
A: Because they rely on sample mean and (co)variances, which are not robust themselves.

In particular, Cook's D suffers from a masking effect:

> *A masking effect occurs when groups of influential observations mask the influence of each other.*
> *(Rousseeuw & van Zomeren, 1990)*

Robust regression can be used instead.

Using robust regression as regression diagnostics

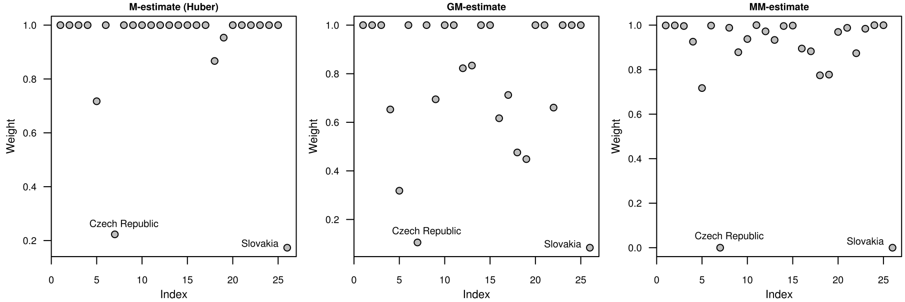
Example: Index plots of robust regression weights w_i

Idea: Weights indicate levels of 'unusualness' (y- and/or x-discrepancies) — the smaller, the more unusual.

Using robust regression as regression diagnostics

Example: Index plots of robust regression weights w_i

Idea: Weights indicate levels of 'unusualness' (y- and/or x-discrepancies) — the smaller, the more unusual.



Using robust regression as regression diagnostics

Example: RR-plots (Tukey, 1991)

Idea: Robust regression residuals are better than OLS residuals for diagnosing outliers:

> *OLS regression tries to produce normal-looking residuals even when the data themselves are not normal.*
> *(Rousseeuw & van Zomeren, 1990)*

Using robust regression as regression diagnostics

Example: RR-plots (Tukey, 1991)

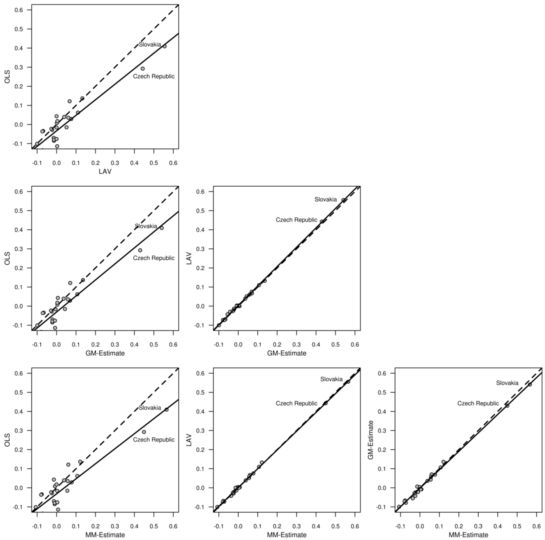
Idea: Robust regression residuals are better than OLS residuals for diagnosing outliers:

> *OLS regression tries to produce normal-looking residuals even when the data themselves are not normal.*
> *(Rousseeuw & van Zomeren, 1990)*

RR-plots: Residual-residual scatterplot matrix

- OLS assumptions hold $\Longrightarrow$ scatter around $y = x$ line (OLS *vs* robust regression residuals)

Using robust regression as regression diagnostics



Standard errors for robust regression

Standard errors for robust regression

- Analytical (asymptotic) SEs are available for only some types of robust regression:
  - ✓ *M*- and *GM*-estimation.
  - ✓ *S*- and *GS*-estimation.
  - ✓ *MM*-estimation.

  These estimates are given by

$$SE_{\widehat{\theta}} = \sqrt{\text{Diagonal}_{\text{var-cov matrix}}}.$$

Standard errors for robust regression

- Analytical (asymptotic) SEs are available for only some types of robust regression:
  - ✓ *M*- and *GM*-estimation.
  - ✓ *S*- and *GS*-estimation.
  - ✓ *MM*-estimation.

  These estimates are given by

$$SE_{\widehat{\theta}} = \sqrt{\text{Diagonal}_{\text{var-cov matrix}}}.$$

- Some problems with $SE_{\widehat{\theta}}$:
  - Unreliable for small n (say, < 40).
  - Reliability decreases as proportion of influential observation increases.

Standard errors for robust regression

- Analytical (asymptotic) SEs are available for only some types of robust regression:
  - ✓ *M*- and *GM*-estimation.
  - ✓ *S*- and *GS*-estimation.
  - ✓ *MM*-estimation.

  These estimates are given by

$$SE_{\widehat{\theta}} = \sqrt{\text{Diagonal}_{\text{var-cov matrix}}}.$$

- Some problems with $SE_{\widehat{\theta}}$:
  - Unreliable for small n (say, < 40).
  - Reliability decreases as proportion of influential observation increases.

Alternative: Bootstrapping

Computing standard errors: Bootstrapping

Bootstrapping is useful to estimate difficult/unknown sampling distributions.

Computing standard errors: Bootstrapping

Bootstrapping is useful to estimate difficult/unknown sampling distributions.

Remember: If OLS assumptions are met, then OLS estimates for the SEs are better than bootstrap estimates!

Computing standard errors: Bootstrapping

Bootstrapping is useful to estimate difficult/unknown sampling distributions.

Remember: If OLS assumptions are met, then OLS estimates for the SEs are better than bootstrap estimates!

There are two bootstrapping options in robust regression:

- Random-x bootstrapping
    - Resample from *data*.
- Fixed-x bootstrapping
    - Resample from *residuals*.

Example: Public opinion about pay inequality

Regression Coeffs.	OLS (all)	OLS (~~cz~~, ~~sk~~)	*L*-Est. (LAV)	*M*-Est. (Huber)	*M*-Est. (Biweight)	*MM*-Est.
Intercept	.0283	−.1069	−.0791	−.0632	−.0905	−.0978
Gini	.0007	.0053	.0045	.0039	.0049	.0051
GDP	.0175	.0063	.0059	.0089	.0052	.0057

Example: Public opinion about pay inequality

Regression Coeffs.	OLS (all)	OLS (~~cz~~, ~~sk~~)	*L*-Est. (LAV)	*M*-Est. (Huber)	*M*-Est. (Biweight)	*MM*-Est.
Intercept	.0283	−.1069	−.0791	−.0632	−.0905	−.0978
Gini	.0007	.0053	.0045	.0039	.0049	.0051
GDP	.0175	.0063	.0059	.0089	.0052	.0057

SEs	OLS (all)	OLS (~~cz~~, ~~sk~~)	*L*-Est. (LAV)	*M*-Est. (Huber)	*M*-Est. (Biweight)	*MM*-Est.
Intercept	.1278	.0578	.0760	.0754	.0658	.0580
Gini	.0028	.0013	.0017	.0017	.0014	.0012
GDP	.0080	.0037	.0046	.0047	.0041	.0035

Computing bootstrapping CIs

- Having estimated regression coefficients *plus* their associated (bootstrapped) SEs, it is now possible to compute $(1-\alpha)\%$ CI.

Computing bootstrapping CIs

- Having estimated regression coefficients *plus* their associated (bootstrapped) SEs, it is now possible to compute $(1-\alpha)\%$ CI.
- There are some possibilities:
  - ✓ Bootstrap *t* CI: When bias is small and the bootstrap sampling distribution is roughly normally distributed.

$$\text{CI} = \widehat{\beta} \pm t_{n-k-1,\alpha/2} SE(\widehat{\beta})$$

Computing bootstrapping CIs

- Having estimated regression coefficients *plus* their associated (bootstrapped) SEs, it is now possible to compute $(1-\alpha)\%$ CI.
- There are some possibilities:
  - ✓ Bootstrap t CI: When bias is small and the bootstrap sampling distribution is roughly normally distributed.

$$\text{CI} = \widehat{\beta} \pm t_{n-k-1,\alpha/2} SE(\widehat{\beta})$$

  - ✓ Bootstrap percentile CI: When bias is small but the bootstrap sampling distribution deviates from the normal distribution.

$$\text{CI} = (q_{\alpha/2}(\beta*), q_{1-\alpha/2}(\beta*)), \quad \beta* = \text{bootstrapped dist.}$$

Computing bootstrapping CIs

- Having estimated regression coefficients *plus* their associated (bootstrapped) SEs, it is now possible to compute $(1-\alpha)\%$ CI.
- There are some possibilities:
  - ✓ Bootstrap *t* CI: When bias is small and the bootstrap sampling distribution is roughly normally distributed.

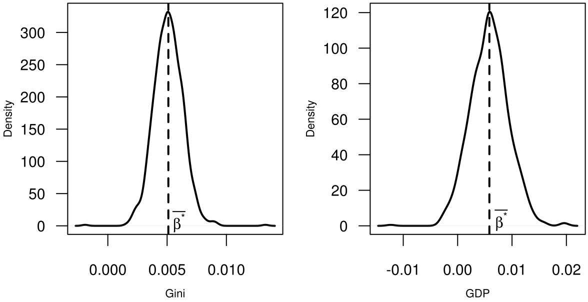
$$\text{CI} = \widehat{\beta} \pm t_{n-k-1,\alpha/2} SE(\widehat{\beta})$$

  - ✓ Bootstrap percentile CI: When bias is small but the bootstrap sampling distribution deviates from the normal distribution.

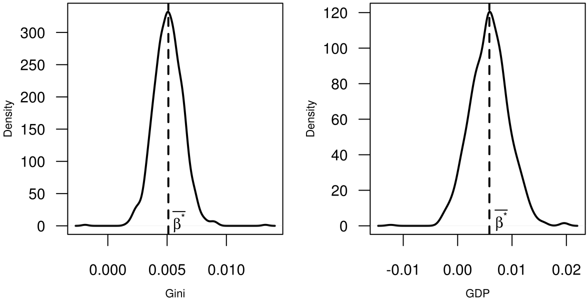
$$\text{CI} = (q_{\alpha/2}(\beta*), q_{1-\alpha/2}(\beta*)), \quad \beta* = \text{bootstrapped dist.}$$

  - ✓ Bias-corrected percentile CI: When bias is large (e.g., for small sample sizes).

Example: Public opinion about pay inequality



Example: Public opinion about pay inequality



Bootstrap *t* CIs				
	B	Boot. SE	Lower 95%	Upper 95%
Intercept	−.0978	.0580		
Gini	.0051	.0012	.0026	.0076
GDP	.0057	.0033	−.0015	.0129

Robust regression in R

Robust regression in R

Fitting regression models		
Reg. Model	R package	R command
OLS	—	`lm(SECPAY ~ gini + GDP)`
L-Estimation (LAV)	quantreg	`rq(SECPAY ~ gini + GDP)`
M-Estimation (Huber)	MASS	`rlm(SECPAY ~ gini + GDP)`
M-Estimation (Biweight)	MASS	`rlm(SECPAY ~ gini + GDP,psi=psi.bisquare)`
MM-Estimation	MASS	`rlm(SECPAY ~ gini + GDP,method="MM")`

Model diagnostics (stats package, loaded by default)		
Statistic	Diagnosing...	R command
Hat values	Leverage	`hatvalues(lm(SECPAY ~ gini + GDP))`
Studentized residual	Reg. outlier	`rstudent(lm(SECPAY ~ gini + GDP))`
DFBETA	Influence	`dfbeta(lm(SECPAY ~ gini + GDP))`
Cook's D	Influence	`cooks.distance(lm(SECPAY ~ gini + GDP))`