Robustification of Statistical and Econometrical Regression Methods

Tomáš Jurczyk

Charles University in Prague,
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

ROBUST 2016 13. 9. 2016

Motivation (goal)

To push method in wide usage, it is not sufficient to have good method. You need to have the whole package of methods, diagnostics and tests around it.

We want to build this "package" around **LWS** method (which is method for robust regression).

Goals:

- We want to find variant of this method in case of multicollinearity
- Investigate and understand the situation in data with contamination together multicollinearity

Regression task

Notation:

Linear regression model:

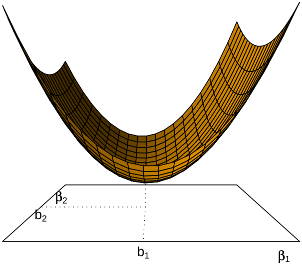
$$Y_i = \sum_{j=1}^{p} x_{ij}\beta_j^0 + e_i, \quad i = 1, 2, \ldots, n.$$

Classical method:

Least Squares (LS)

$$\boldsymbol{b} = \underset{\beta \in \mathcal{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} r_i^2(\boldsymbol{\beta}).$$

Example (Loss function of the LS estimate, 2 regressors)



- $\longleftarrow$ $\sum_{i=1}^{n}(Y_i - x_{1i}\beta_1 - x_{2i}\beta_2)^2 = \sum_{i=1}^{n} r_i^2(\boldsymbol{\beta})$
- $\longleftarrow$ Loss function of LS is quadratic
- $\longleftarrow$ $(b_1, b_2)' = \boldsymbol{b}$ denote LS estimate – the minimal value of loss function

Assumptions

LS estimator is simple and widely used

It is also due to nice properties of this estimator.
Under following assumptions LS is BLUE:

1) $e_i, \ \ i = 1, \dots, n$ are independent
2) $Ee_i = 0$ for all i, which means $E\boldsymbol{Y} = \boldsymbol{X}_i'\boldsymbol{\beta}^0$
3) The rank of the matrix $\boldsymbol{X}$ is full.
4) Variance $var e_i = \sigma^2, \ \ i = 1, \dots, n$.
5) $e_i, \ \ i = 1, \dots, n$ has normal distribution.

Assumptions

LS estimator is simple and widely used

It is also due to nice properties of this estimator.
Under following assumptions LS is BLUE:

1) $e_i, \quad i = 1, \ldots, n$ are independent
2) $Ee_i = 0$ for all i, which means $E\boldsymbol{Y} = \boldsymbol{X}_i'\boldsymbol{\beta}^0$
3) The rank of the matrix $\boldsymbol{X}$ is full.
4) Variance $var e_i = \sigma^2, \quad i = 1, \ldots, n$.
5) $e_i, \quad i = 1, \ldots, n$ has normal distribution.

Typicaly not all assumptions are fulfilled.

Here are possible data problems:

- Observation are not independent
- Heteroscedasticity (different variances of components of error term)
- **Multicollinearity** (problem with dependence of regressors)
- Not normal distribution of error term
- Presence of **outlying observations**

Problems of data - Multicollinearity

Multicollinearity

- situation when regressors are “nearly" linear dependend

Problems of data - Multicollinearity

Multicollinearity

- situation when regressors are “nearly" linear dependend

Consequences for the least squares method (LS)

- Matrix $\boldsymbol{X'X}$ is almost singular
- The smallest eigenvalue t_p^2 of the matrix $\boldsymbol{X'X}$ is close to 0.
- Numerical solution of normal equation is not stable.
- Multicollinearity induces large expected value of the length of the LS estimate ($\boldsymbol{b}$).
  $E\left\|\boldsymbol{b}\right\|^2 - \left\|\beta^{\mathbf{0}}\right\|^2 = E\left\|\boldsymbol{b} - \beta^{\mathbf{0}}\right\|^2 = tr(var(\boldsymbol{b})) = \sigma^2 tr(\boldsymbol{X'X})^{-1} = \sigma^2 \sum_{i=1}^{p}(1/t_i)^2$
- It may cause large variance of b_j.
  $var\boldsymbol{b} = \sigma^2(\boldsymbol{X'X})^{-1} = \sigma^2 \sum_{i=1}^{p} t_i^{-2}\boldsymbol{q}_i\boldsymbol{q}_i'$

where: $\boldsymbol{X} = \boldsymbol{PTQ}'$, $\boldsymbol{P'P} = \boldsymbol{QQ}' = \boldsymbol{I}$, $\boldsymbol{T} = diag(t_1, t_2, \ldots, t_p)$, t_i^2 eigen value of $\boldsymbol{X'X}$

Problems of data - Multicollinearity

Multicollinearity

- situation when regressors are “nearly" linear dependend

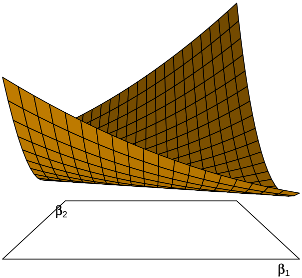
Consequences for the least squares method (LS)

- Matrix $\boldsymbol{X}'\boldsymbol{X}$ is almost singular
- The smallest eigenvalue t_p^2 of the matrix $\boldsymbol{X}'\boldsymbol{X}$ is close to 0.
- Numerical solution of normal equation is not stable.
- Multicollinearity induces large expected value of the length of the LS estimate ($\boldsymbol{b}$).
- It may cause large variance of b_j.

Example (Loss function of the LS estimate, 2 regressors, data with multicollinearity)

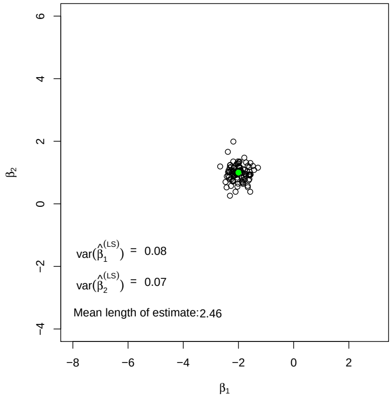
→ Typical shape for data with strong multicollinearity

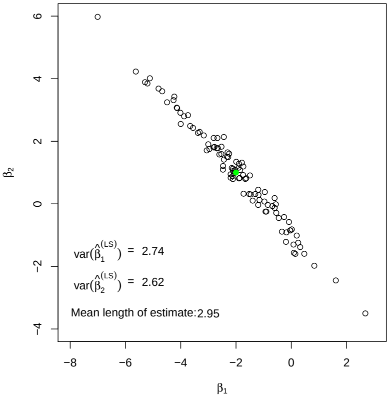
→ There are almost the same values of loss function along some line in parameter space ⇒ unstable behaviour ⇒ large variance and large expected length of $\boldsymbol{b}$



Demonstration of multicollinearity for LS

LS estimate on independent regressors (left graph) and LS estimate for multicollinear regressors (right graph). Each point is LS estimate for one run of simulated dataset (correlation of regressors in right graph is around 0.99). Green dot is theoretical value of β^0.





Ridge regression

We show estimate used instead of the least squares when multicollinearity is present.

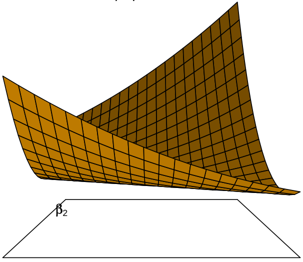
Ridge Regression (RR)

$$\boldsymbol{b}_\delta = \underset{\beta \in \mathcal{R}^p}{\operatorname{argmin}} \left(\sum_{i=1}^{n} r_i^2(\boldsymbol{\beta}) + \delta \sum_{j=1}^{p} \beta_j^2 \right).$$

Example (2 regressors, data with multicollinearity)

Loss function of the LS estimate

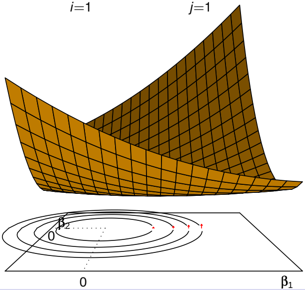
$$\sum_{i=1}^{n} r_i^2(\boldsymbol{\beta})$$



$\rightarrow$

Loss function of the ridge estimate

$$\sum_{i=1}^{n} r_i^2(\boldsymbol{\beta}) + \delta \sum_{j=1}^{p} \beta_j^2$$



Problem of data - contamination

Basic goal of robust statistics

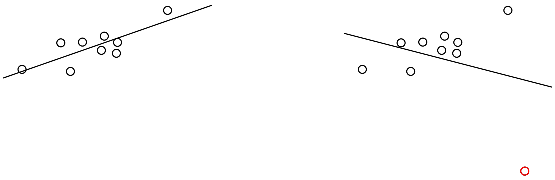
- Finding models which correspond to the structure of the majority of the data.
- The outlier detection is closely connected with this objective.

Outlier in our context - observation which does not follow the regression model.

Problems with outlier presence

- Already one outlier is able to change the value of LS estimate essentially.

Example (Influence of one outlier to LS estimate)



Approaches to overcome contamination

Reduction of influence of contaminating points:

- Different loss function (M-estimates, Least Absolute Deviation regression,...)
- Implicit residual weighting (LTS, LMS, LWS)

Representative of robust methods

Least Trimmed Squares (LTS)

Let $n/2 \leq h \leq n$. Then

$$\boldsymbol{b}^{(LTS,n,h)} = \underset{\beta \in \mathcal{R}^p}{\operatorname{argmin}} \sum_{i=1}^{h} r_{(i)}^2(\boldsymbol{\beta})$$

is called Least Trimmed Squares Estimator, $r_{(j)}^2(\boldsymbol{\beta})$ is the j-th order statistic among the squared residuals.

Generalization of LTS:

Least Weighted Squares (LWS)

Let $1 = w_1 \geq w_2 \geq \ldots . \geq w_n \geq 0$ are weights. Then

$$\boldsymbol{b}^{(LWS,w)} = \underset{\beta \in \mathcal{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} w_i r_{(i)}^2(\boldsymbol{\beta})$$

is called Least Weighted Squares estimate.

Multicollinearity in company with outliers

We need to stress that **contamination and multicollinearity are different in their essence**. Multicollinearity is a problem of regressors, whereas the presence of outliers is the problem with non-compliance of the regression model.

This induces lot of problematic combined situations.

Known problems with outliers in connection with multicollinearity

- Outlier can affect value of estimate of classical methods (like RR).
- In addition already one outlier can hide or create multicollinearity for classical methods for multicollinearity detection.
- Therefore we need some robust detector of multicollinearity.

Multicollinearity in company with outliers

We need to stress that **contamination and multicollinearity are different in their essence**. Multicollinearity is a problem of regressors, whereas the presence of outliers is the problem with non-compliance of the regression model.

This induces lot of problematic combined situations.

Known problems with outliers in connection with multicollinearity

- Outlier can affect value of estimate of classical methods (like RR).
- In addition already one outlier can hide or create multicollinearity for classical methods for multicollinearity detection.
- Therefore we need some robust detector of multicollinearity.

First idea and approach also presented in literature is to use high breakdown methods for outlier detection and after revelation of outliers use some classical multicollinearity diagnostics on non-contaminated data.

Will such approach work?

Multicollinearity in company with outliers

We need to stress that **contamination and multicollinearity are different in their essence**. Multicollinearity is a problem of regressors, whereas the presence of outliers is the problem with non-compliance of the regression model.

This induces lot of problematic combined situations.

Known problems with outliers in connection with multicollinearity

- Outlier can affect value of estimate of classical methods (like RR).
- In addition already one outlier can hide or create multicollinearity for classical methods for multicollinearity detection.
- Therefore we need some robust detector of multicollinearity.

First idea and approach also presented in literature is to use high breakdown methods for outlier detection and after revelation of outliers use some classical multicollinearity diagnostics on non-contaminated data.

Will such approach work?

You will see in the poster...

Poster content

Outline what you will see in the poster:

- Simple examples for understanding multicollinearity and outliers separately
- Investigation of functionality of recent proposals for regression methods suitable for combined outlier-multicollinearity problem – some important results have been done in this area
- Proposal of new regression method called Ridge Least Weighted Squares
- Properties of this estimator
- Using new estimate for diagnostics

Strategy

Least squares $\longrightarrow$ *Ridge regression*

$\downarrow$

Least weighted squares

Strategy

Least squares	$\longrightarrow$	*Ridge regression*
$\downarrow$		$\downarrow$
Least weighted squares	$\longrightarrow$	*Robust ridge regression*

Strategy

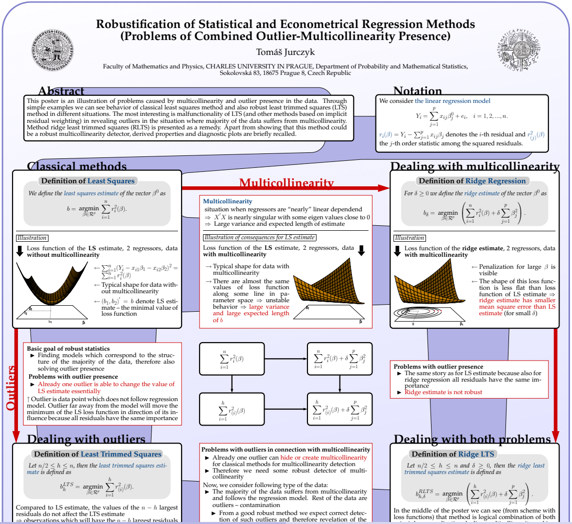
$$\operatorname*{argmin}_{\beta \in \mathcal{R}^p} \sum_{i=1}^{n} r_i^2(\boldsymbol{\beta}) \longrightarrow \operatorname*{argmin}_{\beta \in \mathcal{R}^p} \left(\sum_{i=1}^{n} r_i^2(\boldsymbol{\beta}) + \delta \sum_{j=1}^{p} \beta_j^2 \right)$$

$$\downarrow \qquad\qquad \downarrow$$

$$\operatorname*{argmin}_{\beta \in \mathcal{R}^p} \sum_{i=1}^{n} w_i r_{(i)}^2(\boldsymbol{\beta}) \longrightarrow \operatorname*{argmin}_{\beta \in \mathcal{R}^p} \left(\sum_{i=1}^{n} w_i r_{(i)}^2(\boldsymbol{\beta}) + \delta \sum_{j=1}^{p} \beta_j^2 \right)$$

How to find poster

Follow the same color as this presentation...



My articles concerning this work

[1] T. Jurczyk, High Breakdown Point in Regression, *in WDS'08 Proceedings of Contributed Papers: Part I - Mathematics and Computer Sciences (eds. J. Safrankova and J. Pavlu), Prague, Matfyzpress*, pp. 94-–99, 2008.

[2] T. Jurczyk, Ridge least weighted squares, *Acta Universitatis Carolinae, Mathematica et Physica*, 52, 1, 15-26, 2011.

[3] T. Jurczyk, Weak consistency and weak $\sqrt{n}$-consistency of ridge least weighted squares, *Proceedings of the 14th Applied Stochastic Models and Data Analysis (AMSDA 2011) Conference, Faculty of Economics of the University of Rome, Rome, Italy*, pp. 635–643, 2011.

[4] T. Jurczyk, Trimmed Estimators in Regression Framework, *Acta Univ. Palacki. Olomuc., Fac.rer. nat., Mathematica 50, 2,* 45–53, 2011.

[5] T. Jurczyk, Outlier detection under multicollinearity, *Journal of Statistical Computation and Simulation vol. 82, no. 2,* pp. 261–278, 2012.