A Heuristic Robust Approach for Real Estate Valuation in Areas with Few Transactions

FIG Working Week 2017

Surveying the world of tomorrow –

From digitalisation to augmented reality

Helsinki – 31 May 2017

Alexander Dorndorf

Geodetic Institute, Leibniz Universität Hannover, Germany

Founded by DFG Deutsche Forschungsgemeinschaft

Motivation

- Real estate valuation:
  - Based on statistical evaluation of purchase cases
  - Data are heterogeneous and dispersion is large
  - Data contain often outliers

Motivation

- Real estate valuation:
  - Based on statistical evaluation of purchase cases
  - Data are heterogeneous and dispersion is large
  - Data contain often outliers

- Areas with few transactions:
  - Only 20 to 30 purchase prices are available
  - Number of observations is too small for classical outlier detection

    => Detection of outliers in these data is a challenge

Aims and Methodological Approach

- **The focus of this research**

  Robust estimation of market values with a data-driven model in areas with few transactions

- **Aims of the evaluation process**

  Find the best estimation approach to deal with outliers in areas with few transactions

- **Methodological approach**

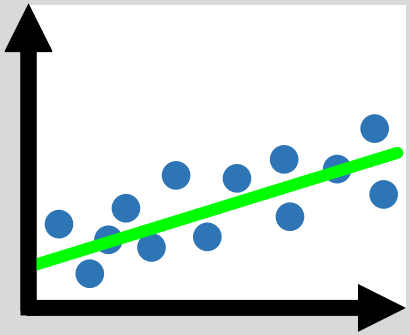
  Comparison of the results of different robust estimation approaches by means of Closed-Loop-Simulation

Source: bennettca.co.nz

Researched Estimation Approaches

Multiple Linear Regression

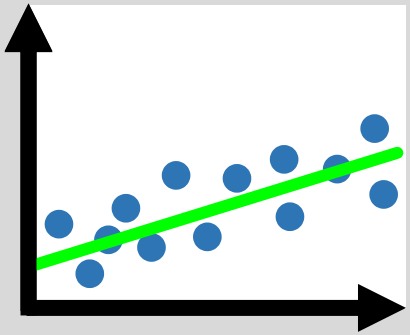
- Assumption: $e_i \sim N(0, \sigma^2)$
- Method of least square
- Is not robust



Researched Estimation Approaches

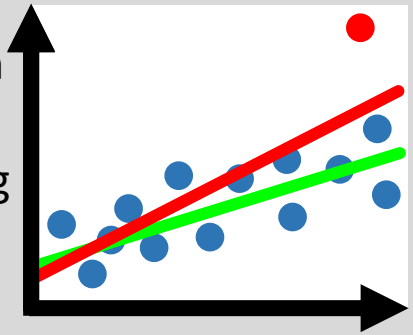
Multiple Linear Regression

- Assumption: $e_i \sim N(0, \sigma^2)$
- Method of least square
- Is not robust



Data Snooping

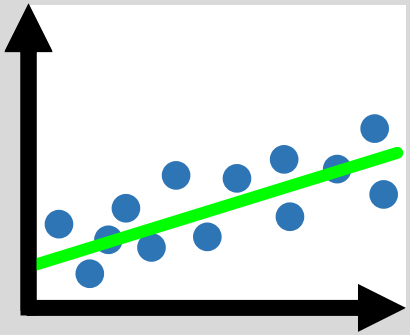
- Regression approach
- Outlier detection by hypothesis testing



Researched Estimation Approaches

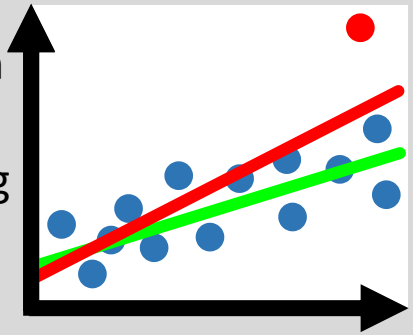
Multiple Linear Regression

- Assumption: $e_i \sim N(0, \sigma^2)$
- Method of least square
- Is not robust



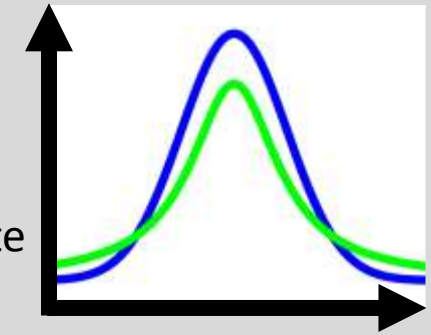
Data Snooping

- Regression approach
- Outlier detection by hypothesis testing



Robust Bayesian Approach

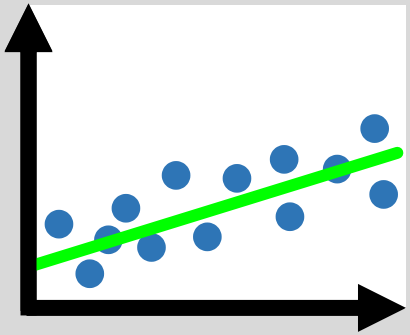
- Replace normal likelihood by student distribution
- Markov Chain Monte Carlo (MCMC)



Researched Estimation Approaches

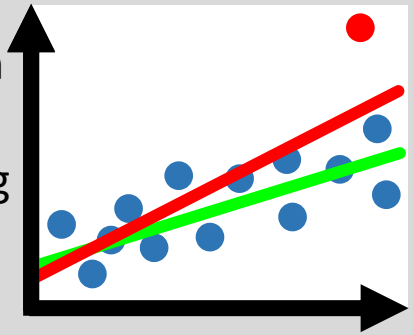
Multiple Linear Regression

- Assumption: $e_i \sim N(0, \sigma^2)$
- Method of least square
- Is not robust



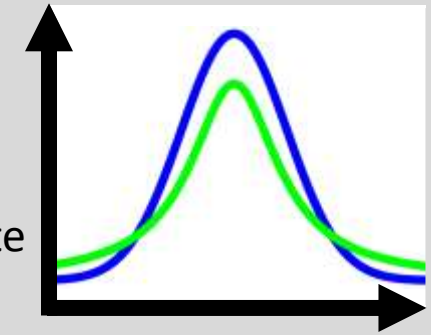
Data Snooping

- Regression approach
- Outlier detection by hypothesis testing



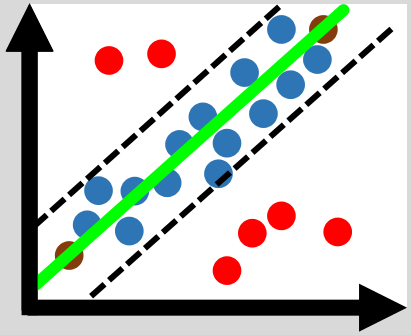
Robust Bayesian Approach

- Replace normal likelihood by student distribution
- Markov Chain Monte Carlo (MCMC)



Random Sample Consensus Set (RANSAC)

- Heuristic approach
- Search the model with the most observations



Strategy of the Investigation

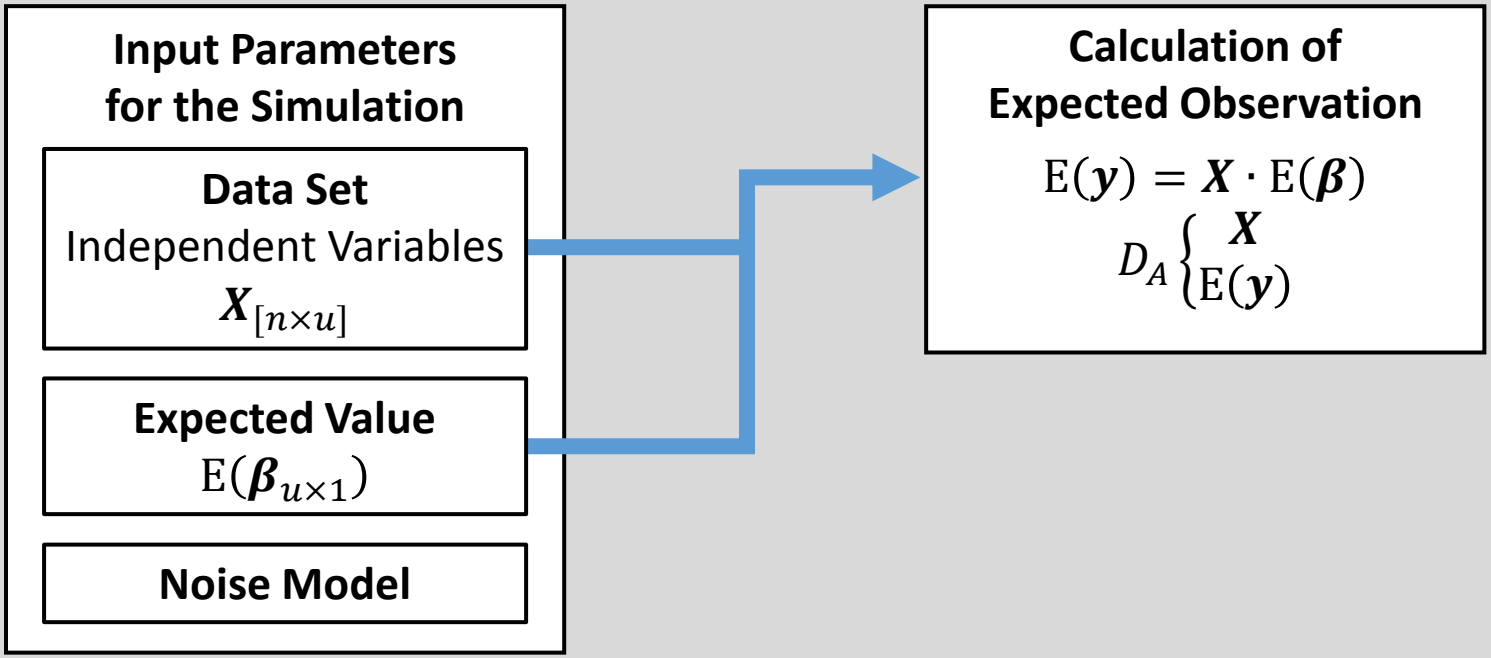
Input Parameters for the Simulation

Data Set
Independent Variables
$\boldsymbol{X}_{[n\times u]}$

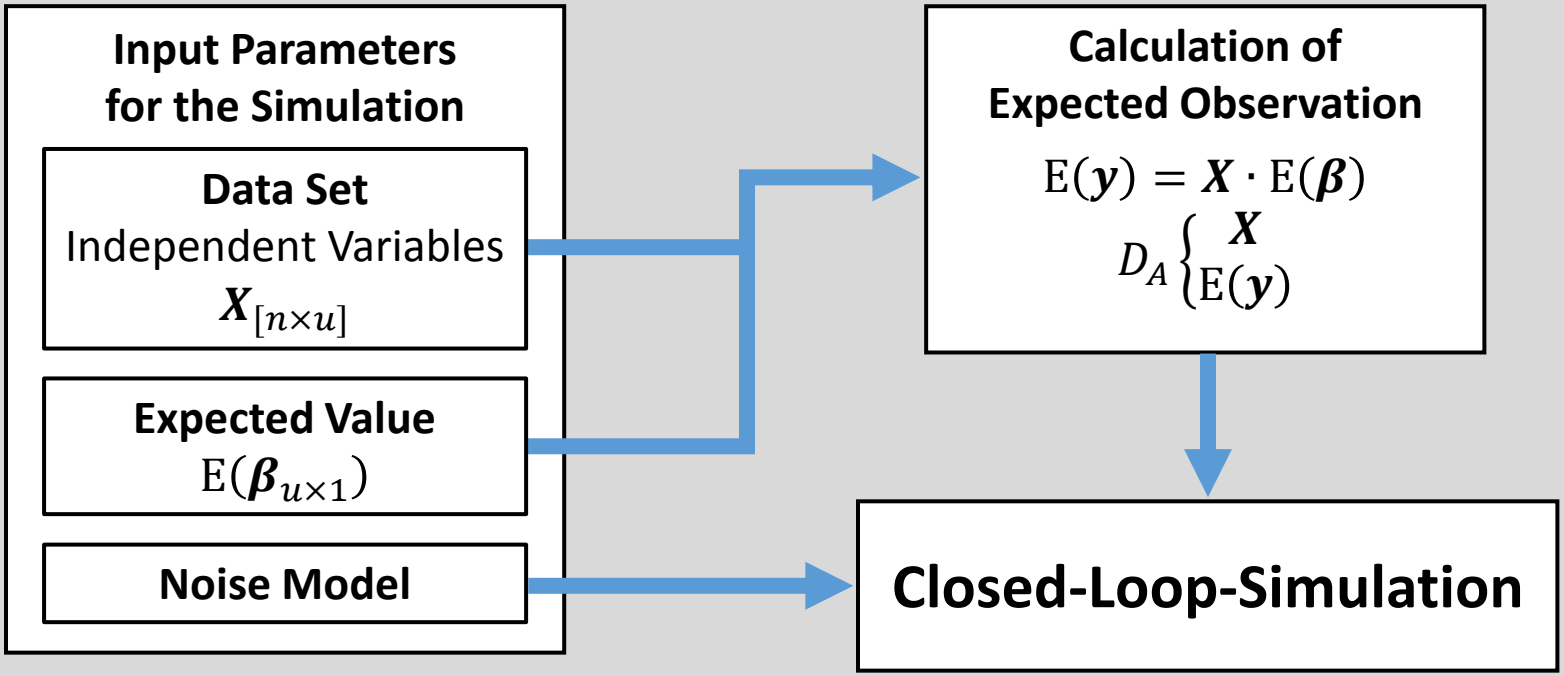
Expected Value
$\mathrm{E}(\boldsymbol{\beta}_{u\times 1})$

Noise Model

Strategy of the Investigation



Strategy of the Investigation

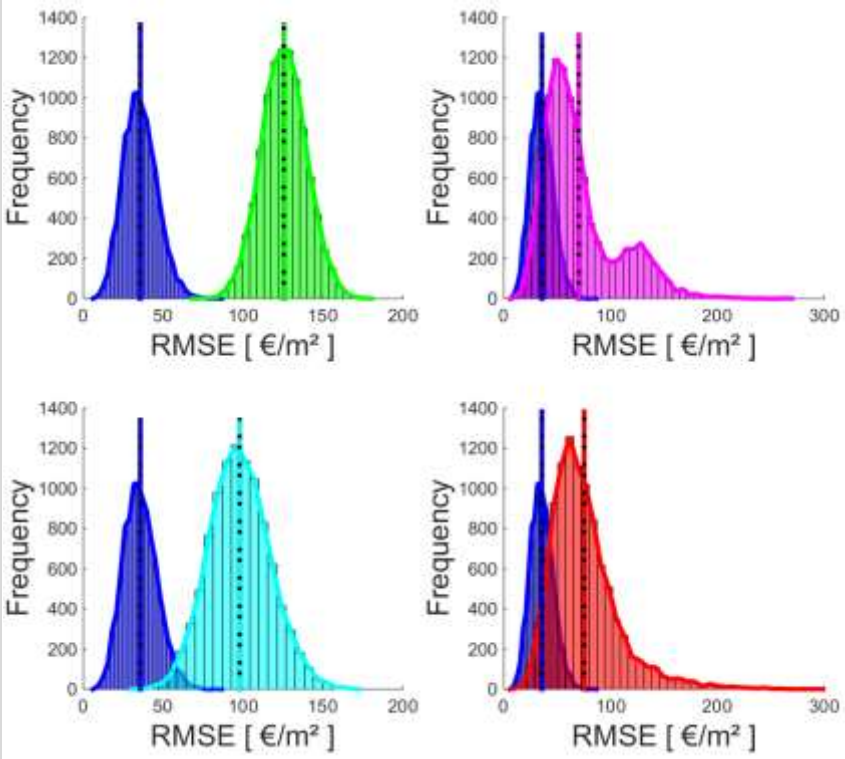


Closed-Loop-Simulation

- Benefit
  - The expected results are known
  - Position of outliers are known
  - Noise distribution is known, e. g. mixture distribution
- Procedure
  - Two scenarios:
    - Submarket with 100 purchases
    - Submarket with 30 purchases
  - Repetition of one scenario 100.000 times
  - Calculation of the Root Mean Square Error (RMSE) between adjusted and expected observations

Results of the Simulation

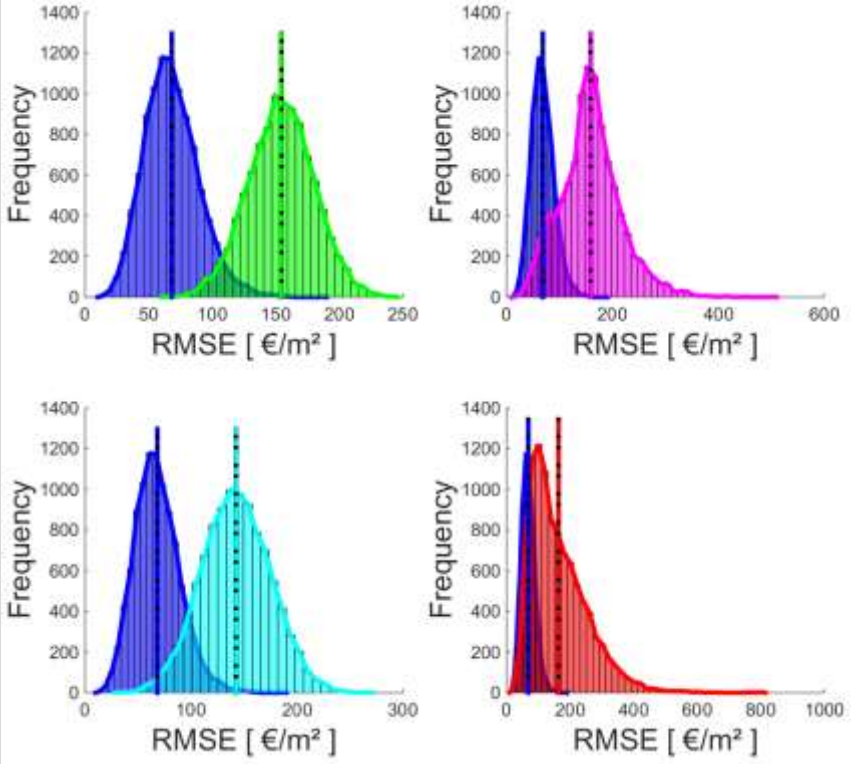
RMSE for 100 purchases



Estimation Approach	Mean [€/m²]	σ [€/m²]
Reference	36	11
Regression	126	14
Data Snooping	70	35
Robust Bayesian Approach	98	19
RANSAC	75	32

Results of the Simulation

RMSE for 30 purchases



Estimation Approach	Mean [€/m²]	σ [€/m²]
Reference	68	21
Regression	154	26
Data Snooping	159	57
Robust Bayesian Approach	143	32
RANSAC	163	89

Conclusion & Outlook

- Data Snooping and RANSAC provide results with good quality only in data sets with an adequate number of purchases
- Robust Bayesian approach has the greatest potential to deal with outliers in areas with few transaction
- In future studies:
  - Optimisation of the robust Bayesian approach to deal more efficiently with outliers
  - Combination of prior information like offer prices or expert knowledge with the available purchases by means of the Bayes theorem

TECHNISCHE UNIVERSITÄT DRESDEN

FIG WORKING WEEK 2017
Surveying the world of tomorrow – From digitalisation to augmented reality
May 29 - June 2 Helsinki Finland

Leibniz Universität Hannover

Many thanks for your attention!

Matthias Soot (M.Sc.)
Prof. Dr.-Ing. Alexandra Weitkamp

Technische Universität Dresden – Geodetic Institute
Helmholzstraße 10
Dresden D - 01069
GERMANY
Tel. +49 (0) 351-46331653
Fax + 49 (0) 351-46337190
Email: matthias.soot@tu-dresden.de
alexandra.weitkamp@tu-dresden.de

Alexander Dorndorf (M.Sc.)
Dr.-Ing. Hamza Alkhatib

Leibniz Universität Hannover – Geodetic Institute
Nienburger Str. 1
Hannover D - 30167
GERMANY
Tel. +49 (0) 511 762 2464
Fax +49 (0) 511 762 2468
Email: dorndorf@gih.uni-hannover.de
alkhatib@gih.uni-hannover.de

Founded by

DFG Deutsche Forschungsgemeinschaft

Strategy of the Investigation

Input Parameters for the Simulation

Data Set
Independent Variables
$\boldsymbol{X}_{[n\times u]}$

Expected Value
$\mathrm{E}(\boldsymbol{\beta}_{u\times 1})$

Noise Model

Data base are 260 purchases of detached houses. Following influence quantities are used:

- Living space [m²]
- Area of lot [m²]
- Standard land value [€/m²]
- Construction year [year]
- Equipping standard [without unit]

Strategy of the Investigation

Input Parameters for the Simulation

Data Set
Independent Variables
$\boldsymbol{X}_{[n\times u]}$

Expected Value
$\mathrm{E}(\boldsymbol{\beta}_{u\times 1})$

Noise Model

The estimated regression coefficients of the 260 purchases are used as expected values.

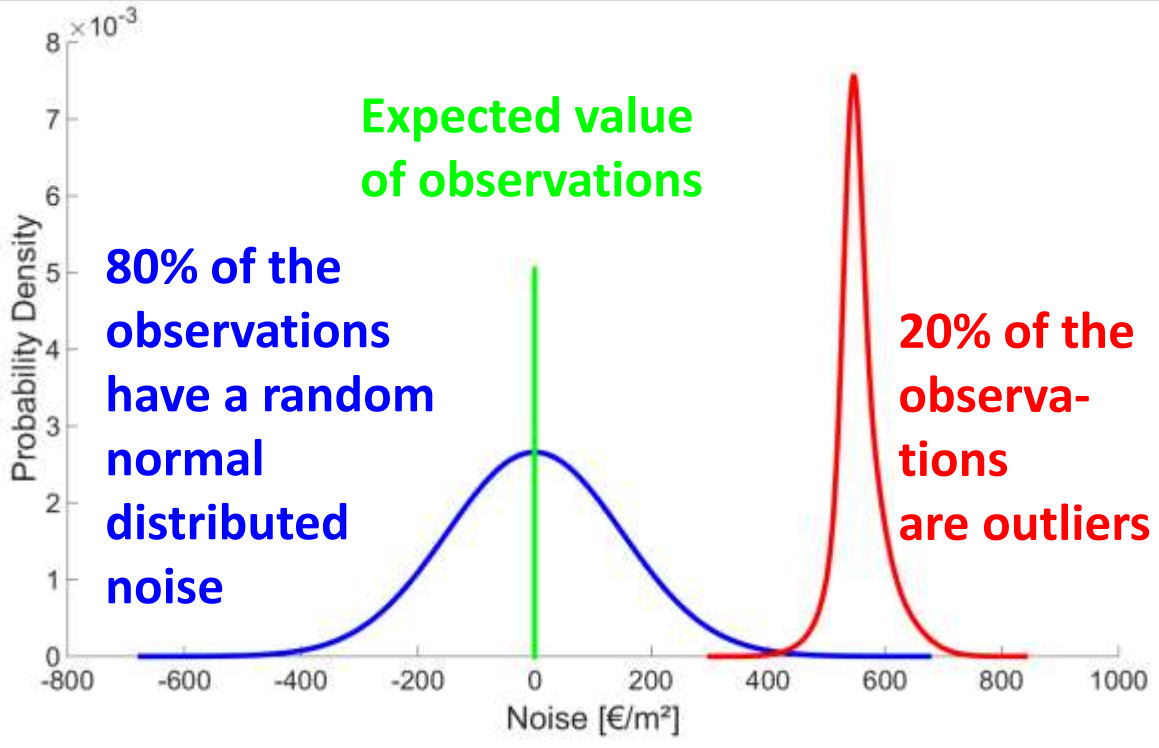
Strategy of the Investigation

Input Parameters for the Simulation

Data Set
Independent Variables
$\boldsymbol{X}_{[n\times u]}$

Expected Value
$\mathrm{E}(\boldsymbol{\beta}_{u\times 1})$

Noise Model



Closed-Loop-Simulation

Create Data Set 1

Create set $\boldsymbol{D}_R^{(j)}$ from set $\boldsymbol{D}_A^{(j)}$ by h times random drawing (e.g. h =30 purchases).

Create Observations 2

Create noise $\varepsilon^{(j)}$ with the noise model.

$$y_R^{(j)} = \mathrm{E}(\boldsymbol{y})_R^{(j)} + \varepsilon^{(j)}$$

Create Data Set without Outliers 3

Create set $\boldsymbol{D}_F^{(j)}$ from set $\boldsymbol{D}_R^{(j)}$ by removing all purchases with outliers.

Estimation 4

- Estimate $\widehat{\boldsymbol{\beta}}_l^{(j)}$ with l ={Regression, Data Snooping, robust Bayes, RANSAC, Reference}
- Calculate the Root Mean Square Error:

$$RMSE_l^{(j)} = \sqrt{\frac{1}{n_l}\sum_{i=1}^{n}\left(\left(\boldsymbol{X}_l^{(j)}\widehat{\boldsymbol{\beta}}_l^{(j)} - \mathrm{E}(\boldsymbol{y}_l)^{(j)}\right)^2\right)}$$

Iterations: $j = 1, \ldots, 100000$