Session 6: Robust inference
Winter course, CMStatistics 2016

Mia Hubert, Peter Rousseeuw, Stefan Van Aelst

Department of Mathematics
KU Leuven, Belgium

December 6–7, 2016

KU LEUVEN

Outline of the course

- 1. General notions of robustness
- 2. Robustness for univariate data
- 3. Robust multivariate methods
- 4. Robust regression
- 5. Robust principal component analysis
- 6. Inference
- 7. Multivariate and functional depth
- 8. High dimensional data and sparsity
- 9. Cellwise outliers

Inference: Outline

1. Robust regression
2. Robust inference
3. Fast and robust bootstrap
4. Robust multivariate location and scatter
5. Inference for robust PCA
6. Inference for robust multivariate regression
7. Robust bootstrap tests in regression
8. Robust multigroup inference
9. Robust model selection
10. Software

Regression model

- Dataset $\mathcal{Z}_n = \{(y_1, \boldsymbol{x}_1), \ldots (y_n, \boldsymbol{x}_n)\} \subset \mathbb{R}^{p+1}$.
- Linear regression model:

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

- $E(\varepsilon_i) = 0$ and $\text{Var}(\varepsilon_i) = \sigma^2$
- Residuals $r_i(\hat{\boldsymbol{\beta}}) = y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}} \qquad i = 1, \ldots, n$

Regression S-estimators

Regression S-estimator

$$\hat{\boldsymbol{\beta}}_S = \underset{\boldsymbol{\beta}}{\text{argmin}}\, \hat{\sigma}(\boldsymbol{\beta})$$

where $\hat{\sigma}(\boldsymbol{\beta})$ is given by

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0\left(\frac{r_i(\boldsymbol{\beta})}{\hat{\sigma}(\boldsymbol{\beta})}\right) = \delta_0$$

with ρ_0 a smooth bounded ρ-function.

Regression MM estimates

Regression MM-estimators

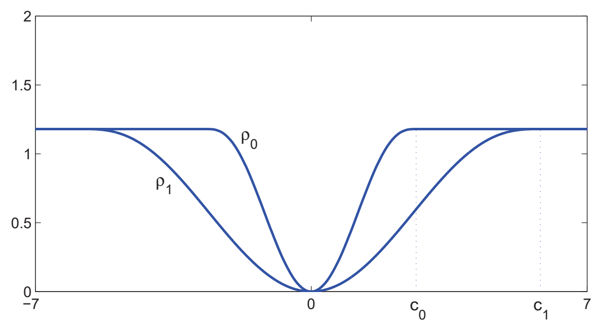
1. Compute an initial regression S-estimator $\hat{\boldsymbol{\beta}}_S$ with high breakdown value, and its corresponding scale estimate $\hat{\sigma}_S = \hat{\sigma}(\hat{\boldsymbol{\beta}}_S)$.
2. Compute a regression M-estimator with fixed scale $\hat{\sigma}_S$ and initial estimate $\hat{\boldsymbol{\beta}}_S$ but now using a bounded ρ-function with high efficiency.

$$\hat{\boldsymbol{\beta}}_{MM} = \underset{\boldsymbol{\beta}}{\text{argmin}}\, \frac{1}{n}\sum_{i=1}^{n} \rho_1\left(\frac{r_i(\boldsymbol{\beta})}{\hat{\sigma}_S}\right)$$

Combines robustness and efficiency

Example: loss functions

Tukey biweight ρ function for 50% breakdown point, and 95% efficiency:



Inference: Outline

Inference for robust estimators

- Reflect precision of parameter estimates
- Reflect existence and strength of observed effects
- Take into account that observed data contains outliers

$\longrightarrow$ Inference should be robust!

Inference for robust estimators

- Parametric inference: based on asymptotic distribution
  - Derived under ideal, outlier-free assumptions
  - No robustness guaranteed
- Bootstrap inference: less assumptions, but
  - High computational cost
  - Loss of robustness
- Computationally feasible and robust bootstrap inference?

$\boxed{\longrightarrow \text{Fast and Robust Bootstrap}}$

Inference: Outline

MM estimating equations

The estimates $\hat{\boldsymbol\beta}_{MM}$, $\hat\sigma_S$ and $\hat{\boldsymbol\beta}_S$ satisfy the equations

$$
\begin{aligned}
\hat{\boldsymbol\beta}_{MM} &= \left[\sum_{i=1}^n w_i^1\, \boldsymbol{x}_i\, \boldsymbol{x}_i'\right]^{-1} \sum_{i=1}^n w_i^1\, \boldsymbol{x}_i\, y_i \ \text{ with } w_i^1 = \frac{\psi_1(r_i(\hat{\boldsymbol\beta}_{MM})/\hat\sigma_S)}{r_i(\hat{\boldsymbol\beta}_{MM})}\,, \\
\hat\sigma_S &= \sum_{i=1}^n v_i\,(y_i - \hat{\boldsymbol\beta}_S{}'\boldsymbol{x}_i) \qquad \text{ with } v_i = \frac{\hat\sigma_S}{n\delta_0}\,\frac{\rho_0((r_i(\hat{\boldsymbol\beta}_S)/\hat\sigma_S)}{\tilde r_i(\hat{\boldsymbol\beta}_S)}\,, \\
\hat{\boldsymbol\beta}_S &= \left[\sum_{i=1}^n w_i^0\, \boldsymbol{x}_i\, \boldsymbol{x}_i'\right]^{-1} \sum_{i=1}^n w_i^0\, \boldsymbol{x}_i\, y_i \ \text{ with } w_i^0 = \frac{\psi_0(r_i(\hat{\boldsymbol\beta}_S)/\hat\sigma_S)}{r_i(\hat{\boldsymbol\beta}_S)}\,,
\end{aligned}
$$

Bootstrap sample MM estimating equations

- A bootstrap sample $\mathcal{Z}_n^* = \{(\mathbf{x}_1^*, y_1^*), \ldots, (\mathbf{x}_n^*, y_n^*)\}$ consists of n observations drawn from $\mathcal{Z}_n$ with replacement.
- The MM-estimates $\hat{\boldsymbol{\beta}}^*_{MM}$ and $\hat{\sigma}^*_S$ for a bootstrap sample $\mathcal{Z}_n^*$ satisfy the equations

$$\hat{\boldsymbol{\beta}}^*_{MM} = \left[\sum_{i=1}^n w_i^{1*}\, \boldsymbol{x}_i^*\, \boldsymbol{x}_i^{*\prime}\right]^{-1} \sum_{i=1}^n w_i^{1*}\, \boldsymbol{x}_i^*\, y_i^* \quad \text{with } w_i^{1*} = \frac{\psi_1(r_i(\hat{\boldsymbol{\beta}}^*_{MM})/\hat{\sigma}^*_S)}{r_i(\hat{\boldsymbol{\beta}}^*_{MM})},$$

$$\hat{\sigma}^*_S = \sum_{i=1}^n v_i^*\, (y_i^* - \hat{\boldsymbol{\beta}}^{*\prime}_S \boldsymbol{x}_i^*) \quad \text{with } v_i^* = \frac{\hat{\sigma}^*_S}{n\delta_0} \frac{\rho_0((r_i(\hat{\boldsymbol{\beta}}^*_S)/\hat{\sigma}^*_S)}{\tilde{r}_i(\hat{\boldsymbol{\beta}}^*_S)},$$

$$\hat{\boldsymbol{\beta}}^*_S = \left[\sum_{i=1}^n w_i^{0*}\, \boldsymbol{x}_i^*\, \boldsymbol{x}_i^{*\prime}\right]^{-1} \sum_{i=1}^n w_i^{0*}\, \boldsymbol{x}_i^*\, y_i^* \quad \text{with } w_i^{0*} = \frac{\psi_0(r_i(\hat{\boldsymbol{\beta}}^*_S)/\hat{\sigma}^*_S)}{r_i(\hat{\boldsymbol{\beta}}^*_S)},$$

First order approximation for a bootstrap sample

For a bootstrap sample $\mathcal{Z}_n^* = \{(\mathbf{x}_1^*, y_1^*), \ldots, (\mathbf{x}_n^*, y_n^*)\}$ calculate the estimates

$$\tilde{\boldsymbol{\beta}}^*_{MM} = \left[\sum_{i=1}^n \tilde{w}_i^{1*}\, \boldsymbol{x}_i^*\, \boldsymbol{x}_i^{*\prime}\right]^{-1} \sum_{i=1}^n \tilde{w}_i^{1*}\, \boldsymbol{x}_i^*\, y_i^* \quad \text{with } \tilde{w}_i^{1*} = \frac{\psi_1(r_i(\hat{\boldsymbol{\beta}}_{MM})/\hat{\sigma}_S)}{r_i(\hat{\boldsymbol{\beta}}_{MM})},$$

$$\tilde{\sigma}^*_S = \sum_{i=1}^n \tilde{v}_i^*\, (y_i^* - \hat{\boldsymbol{\beta}}_S{}' \boldsymbol{x}_i^*) \quad \text{with } \tilde{v}_i^* = \frac{\hat{\sigma}_S}{n\delta_0} \frac{\rho_0((r_i(\hat{\boldsymbol{\beta}}_S)/\hat{\sigma}_S)}{\tilde{r}_i(\hat{\boldsymbol{\beta}}_S)},$$

$$\tilde{\boldsymbol{\beta}}^*_S = \left[\sum_{i=1}^n \tilde{w}_i^{0*}\, \boldsymbol{x}_i^*\, \boldsymbol{x}_i^{*\prime}\right]^{-1} \sum_{i=1}^n \tilde{w}_i^{0*}\, \boldsymbol{x}_i^*\, y_i^* \quad \text{with } \tilde{w}_i^{0*} = \frac{\psi_0(r_i(\hat{\boldsymbol{\beta}}_S)/\hat{\sigma}_S)}{r_i(\hat{\boldsymbol{\beta}}_S)},$$

Note that $\hat{\boldsymbol{\beta}}_{MM}$, $\hat{\sigma}_S$ and $\hat{\boldsymbol{\beta}}_S$ are not recalculated!

Linear correction: FRB estimates

The first order approximations $\tilde{\boldsymbol{\beta}}^*_{MM}$, $\tilde{\sigma}^*_S$ and $\tilde{\boldsymbol{\beta}}^*_S$ underestimate the sampling variability!

$\Rightarrow$ apply a linear correction:

- Put $\widehat{\Theta} = (\hat{\boldsymbol{\beta}}_{MM}, \hat{\sigma}_S, \hat{\boldsymbol{\beta}}_S)$, then we have functions $\mathbf{g}_n$, such that

$$\mathbf{g}_n(\widehat{\Theta}) = \widehat{\Theta}$$

- Taylor expansion about estimands Θ:

$$\widehat{\Theta} = \mathbf{g}_n(\Theta) + \nabla\mathbf{g}_n(\Theta)(\widehat{\Theta} - \Theta) + R$$

- With R small, rewrite:

$$(\widehat{\Theta} - \Theta) \approx [\mathbf{I} - \nabla\mathbf{g}_n(\Theta)]^{-1}(\mathbf{g}_n(\Theta) - \Theta)$$

- Taking bootstrap equivalents:

$$(\widehat{\Theta}^* - \widehat{\Theta}) \approx [\mathbf{I} - \nabla\mathbf{g}_n(\widehat{\Theta})]^{-1}(\mathbf{g}^*_n(\widehat{\Theta}) - \widehat{\Theta})$$

Properties of fast and robust bootstrap

Fast and robust bootstrap distribution (Salibian-Barrera and Zamar, 2002)

The fast and robust bootstrap estimates for $\widehat{\Theta} = (\hat{\boldsymbol{\beta}}_{MM}, \hat{\sigma}_S, \hat{\boldsymbol{\beta}}_S)$ are given by

$$\widehat{\Theta}^*_{FRB} = \widehat{\Theta} + [\mathbf{I} - \nabla\mathbf{g}_n(\widehat{\Theta})]^{-1}(\widetilde{\Theta}^* - \widehat{\Theta})$$

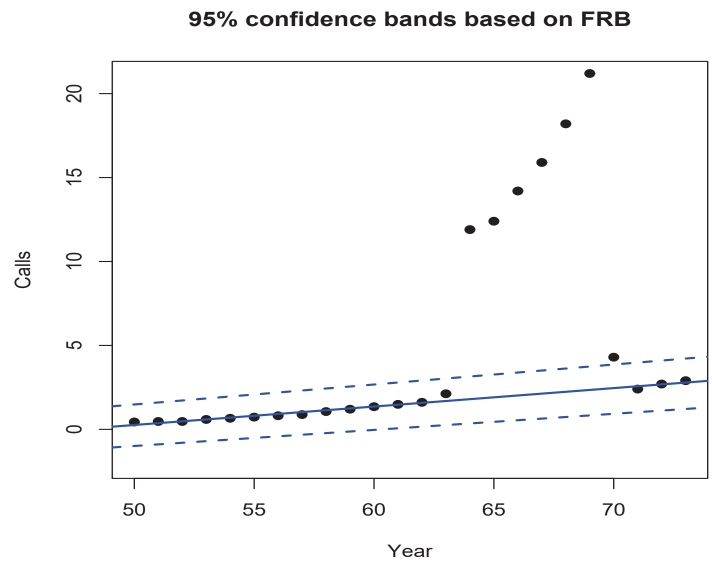
where $\widetilde{\Theta}^* = (\tilde{\boldsymbol{\beta}}^*_{MM}, \tilde{\sigma}^*_S, \tilde{\boldsymbol{\beta}}^*_S) = \mathbf{g}^*_n(\widehat{\Theta})$

Consistency: Under regularity conditions the FRB distribution and the sample distribution of the estimators converge to the same limiting distribution.

Computational efficiency: The FRB estimates are solutions of a system of linear equations.

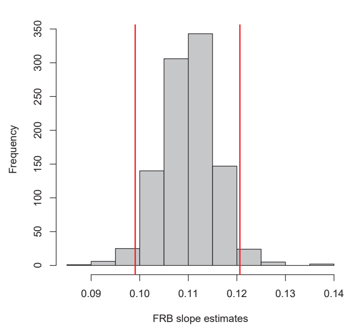
Robustness: The FRB estimates use the weights of the MM/S-estimates at the original sample. FRB quantiles have maximal breakdown point.

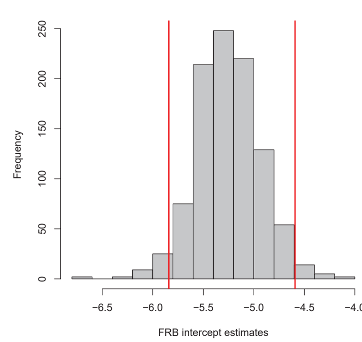
Example: Telephone data



Robust inference: Telephone data

Yearly number of international calls from Belgium, from 1950 to 1973





Inference: Outline

Multivariate location and scatter model

- Dataset $\mathcal{X}_n = \{(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^p$.
- Multivariate location and scatter model:

$$\boldsymbol{x}_i = \boldsymbol{\mu} + \Sigma^{1/2}\,\boldsymbol{\varepsilon}_i \qquad i = 1, \ldots, n$$

- $E(\boldsymbol{\varepsilon}_i) = 0$ and $\mathrm{Cov}(\boldsymbol{\varepsilon}_i) = I_p$
- Distances $d_i(\boldsymbol{\mu}, \Sigma) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})} \qquad i = 1, \ldots, n$

Multivariate S-estimates

S-estimator of location and scatter

$$(\hat{\boldsymbol{\mu}}_S, \hat{\Sigma}_S) = \underset{\boldsymbol{\mu},\Sigma}{\text{argmin}}\, |\Sigma|$$

over all $\boldsymbol{\mu} \in \mathbb{R}^p$ and symmetric positive definite Σ that satisfy

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0(d_i(\boldsymbol{\mu}, \Sigma)) = \delta$$

with ρ_0 a smooth *bounded* ρ-function.

Multivariate MM-estimates

MM-estimator of location and scatter

1. Denote $\hat{\sigma}^2 = |\hat{\Sigma}_S|^{1/p}$, the S-estimate of the generalized scale.
2. The MM-estimator for location and shape $(\hat{\boldsymbol{\mu}}_{MM}, \hat{\Gamma}_{MM})$ minimizes

$$\frac{1}{n}\sum_{i=1}^{n} \rho_1\left(\frac{\sqrt{(\boldsymbol{x}_i - \boldsymbol{\mu})'\Gamma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})}}{\hat{\sigma}}\right) \tag{1}$$

   among all $\boldsymbol{\mu} \in \mathbb{R}^p$ and symmetric positive definite Γ with $|\Gamma| = 1$.

   The MM-estimator for the covariance matrix is then $\hat{\Sigma}_{MM} = \hat{\sigma}^2 \hat{\Gamma}_{MM}$.

Combines robustness and efficiency

Multivariate MM estimating equations

The estimates $\hat{\boldsymbol{\mu}}_{MM}$, $\hat{\Gamma}_{MM}$, $\hat{\boldsymbol{\mu}}_S$, and $\hat{\Sigma}_S$ satisfy the equations

$$\hat{\boldsymbol{\mu}}_{MM} = \Big(\sum_{i=1}^{n} \frac{\psi_1(d_{i,MM}/\hat{\sigma})}{d_{i,MM}}\Big)^{-1}\Big(\sum_{i=1}^{n} \frac{\psi_1(d_{i,MM}/\hat{\sigma})}{d_{i,MM}}\boldsymbol{x}_i\Big)$$

$$\widehat{\Gamma}_{MM} = G\Big(\sum_{i=1}^{n} \frac{\psi_1(d_{i,MM}/\hat{\sigma})}{d_{i,MM}}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_{MM})(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_{MM})'\Big)$$

$$\hat{\boldsymbol{\mu}}_S = \Big(\sum_{i=1}^{n} \frac{\psi_0(d_{i,S})}{d_{i,S}}\Big)^{-1}\Big(\sum_{i=1}^{n} \frac{\psi_0(d_{i,S})}{d_{i,S}}\boldsymbol{x}_i\Big)$$

$$\hat{\Sigma}_S = \frac{1}{n\delta}\Big(\sum_{i=1}^{n} p\frac{\psi_0(d_{i,S})}{d_{i,S}}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_S)(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_S)' + \Big(\sum_{i=1}^{n} v_i\Big)\hat{\Sigma}_S\Big)$$

where $G(A) = |A|^{-1/p}A$, $v_i = \rho_0(d_{i,S}) - \psi_0(d_{i,S})d_{i,S}$, $\hat{\sigma} = |\hat{\Sigma}_S|^{1/p}$ and $d_{i,MM}^2 = (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_{MM})'\widehat{\Gamma}_{MM}^{-1}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_{MM})$, $d_{i,S}^2 = (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_S)'\hat{\Sigma}_S^{-1}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_S)$.

First order approximation for a bootstrap sample

$$\tilde{\boldsymbol{\mu}}^*_{MM} = \Big(\sum_{i=1}^{n} \frac{\psi_1(\tilde{d}^*_{i,MM}/\hat{\sigma})}{\tilde{d}^*_{i,MM}}\Big)^{-1}\Big(\sum_{i=1}^{n} \frac{\psi_1(\tilde{d}^*_{i,MM}/\hat{\sigma})}{\tilde{d}^*_{i,MM}}\boldsymbol{x}^*_i\Big)$$

$$\widetilde{\Gamma}^*_{MM} = G\Big(\sum_{i=1}^{n} \frac{\psi_1(\tilde{d}^*_{i,MM}/\hat{\sigma})}{\tilde{d}^*_{i,MM}}(\boldsymbol{x}^*_i - \hat{\boldsymbol{\mu}}_{MM})(\boldsymbol{x}^*_i - \hat{\boldsymbol{\mu}}_{MM})'\Big)$$

$$\tilde{\boldsymbol{\mu}}^*_S = \Big(\sum_{i=1}^{n} \frac{\psi_0(\tilde{d}^*_{i,S})}{\tilde{d}^*_{i,S}}\Big)^{-1}\Big(\sum_{i=1}^{n} \frac{\psi_0(\tilde{d}^*_{i,S})}{\tilde{d}^*_{i,S}}\boldsymbol{x}^*_i\Big)$$

$$\widetilde{\Sigma}^*_S = \frac{1}{n\delta}\Big(\sum_{i=1}^{n} p\frac{\psi_0(\tilde{d}^*_{i,S})}{\tilde{d}^*_{i,S}}(\boldsymbol{x}^*_i - \hat{\boldsymbol{\mu}}_S)(\boldsymbol{x}^*_i - \hat{\boldsymbol{\mu}}_S)' + \Big(\sum_{i=1}^{n} \tilde{v}_i\Big)\hat{\Sigma}_S\Big)$$

where $G(A) = |A|^{-1/p}A$, $\tilde{v}_i = \rho_0(\tilde{d}^*_{i,S}) - \psi_0(\tilde{d}_{i,S})^*\tilde{d}^*_{i,S}$, $\hat{\sigma} = |\hat{\Sigma}_S|^{1/p}$ and $(\tilde{d}^*_{i,MM})^2 = (\boldsymbol{x}^*_i - \hat{\boldsymbol{\mu}}_{MM})'\widehat{\Gamma}_{MM}^{-1}(\boldsymbol{x}^*_i - \hat{\boldsymbol{\mu}}_{MM})$, $(\tilde{d}^*_{i,S})^2 = (\boldsymbol{x}^*_i - \hat{\boldsymbol{\mu}}_S)'\hat{\Sigma}_S^{-1}(\boldsymbol{x}^*_i - \hat{\boldsymbol{\mu}}_S)$.

Properties of fast and robust bootstrap

Fast and robust bootstrap distribution (Salibian-Barrera et al., 2006)

The fast and robust bootstrap estimates for $\widehat{\Theta} = (\hat{\boldsymbol{\mu}}_{MM}, \hat{\Gamma}_{MM}, \hat{\boldsymbol{\mu}}_S, \hat{\Sigma}_S)$ are given by

$$\widehat{\Theta}^*_{FRB} = \widehat{\Theta} + [\mathbf{I} - \nabla \mathbf{g}_n(\widehat{\Theta})]^{-1}(\widetilde{\Theta}^* - \widehat{\Theta})$$

where $\widetilde{\Theta}^* = (\tilde{\boldsymbol{\mu}}^*_{MM}, \widetilde{\Gamma}^*_{MM}, \tilde{\boldsymbol{\mu}}^*_S, \widetilde{\Sigma}^*_S)$.

Consistency: Under regularity conditions the FRB distribution and the sample distribution of the estimators converge to the same limiting distribution.

Computational efficiency: The FRB estimates are solutions of a system of linear equations.

Robustness: The FRB estimates use the weights of the MM/S-estimates at the original sample. FRB quantiles have maximal breakdown point.

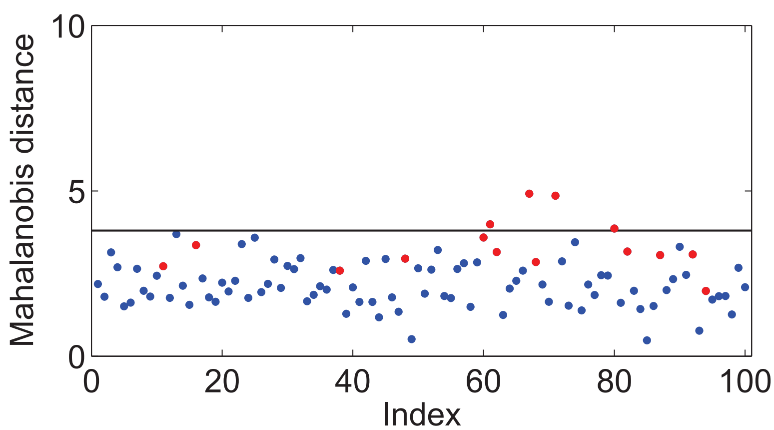
Inference: Outline

1. Robust regression
2. Robust inference
3. Fast and robust bootstrap
4. Robust multivariate location and scatter
5. Inference for robust PCA
6. Inference for robust multivariate regression
7. Robust bootstrap tests in regression
8. Robust multigroup inference
9. Robust model selection
10. Software

Example: Forged Swiss bank notes data

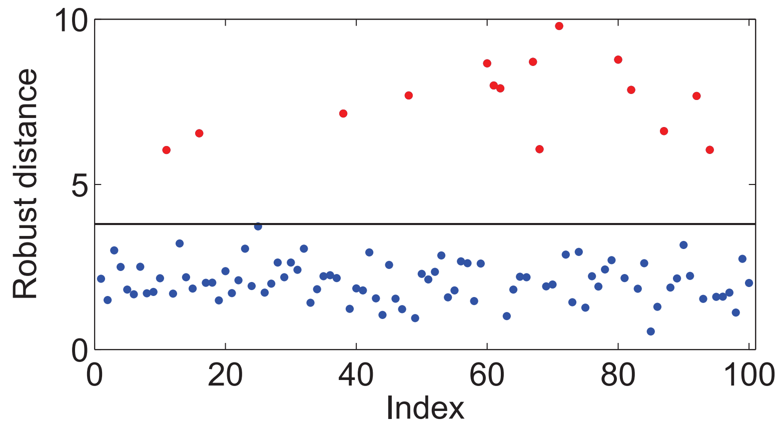
- $n = 100$ forged Swiss bank notes
- 6 variables:
  - V1: length of the bill
  - V2: height of the bill, measured on the left
  - V3: height of the bill, measured on the right
  - V4: distance of inner frame to the lower border
  - V5: distance of inner frame to the upper border
  - V6: length of diagonal

Outliers?



Mahalanobis distances: $MD(\boldsymbol{x}_i) = [(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_n)' S_n^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_n)]^{1/2}$

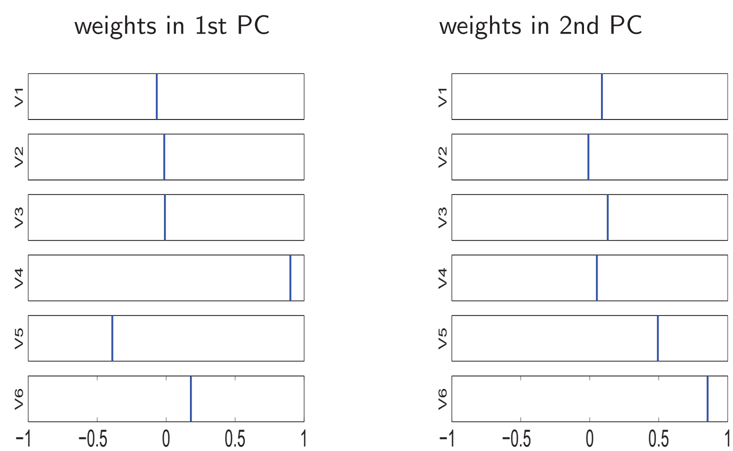
Multivariate outlier detection



⇒ group of 15 outliers

PCA

Classical PC estimates

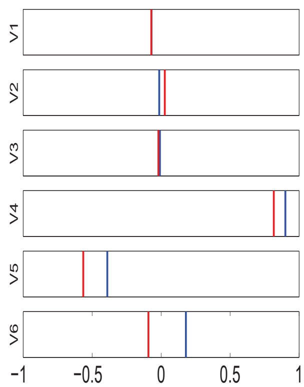


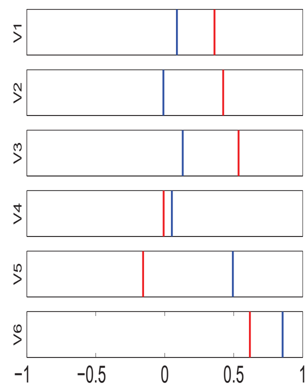
Robust PCA with MM-estimates

Classical + robust (MM) PC estimates

weights in 1st PC

weights in 2nd PC

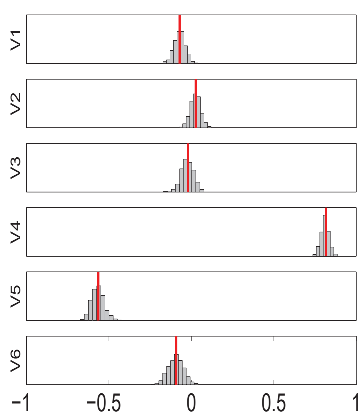




Forged Swiss bank notes data

Histograms of FRB estimates of the weights

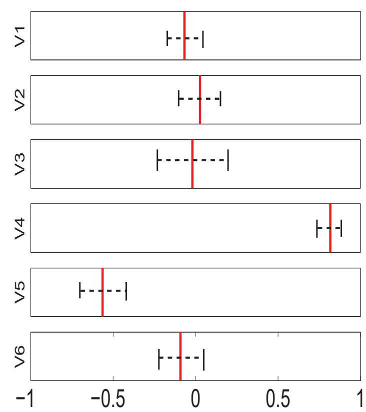
weights in first PC



→ Accuracy of loadings in the first PC

Forged Swiss bank notes data

FRB confidence intervals for the weights in first PC



Forged Swiss bank notes data

Stability of PCA based on MM-estimates?

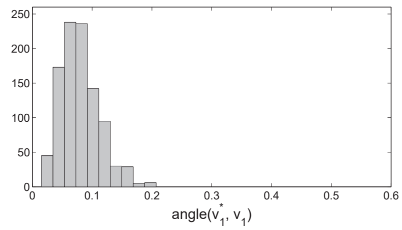
$\rightarrow$ Investigate sampling distribution of angles between $\widehat{v}_1$ and *true* component v_1: **distribution of acos($|v_1'\widehat{v}_1|$)**

$\triangleright$ can be estimated through bootstrap values acos($|\widehat{v}_1'\widehat{v}_1^*|$) $\triangleleft$

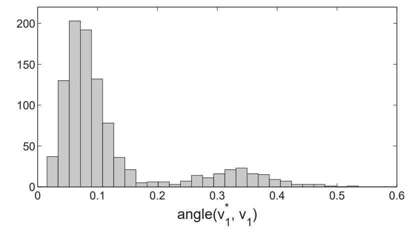
Forged Swiss bank notes data

angles between $\widehat{v}_1^*$ and $\widehat{v}_1$ ($\in [0, \pi/2]$)

FRB



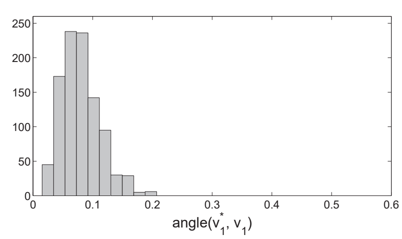
classical bootstrap



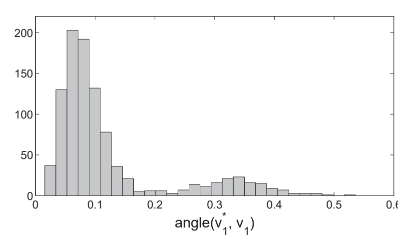
Forged Swiss bank notes data

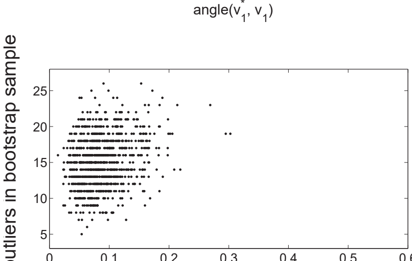
angles between $\widehat{v}_1^*$ and $\widehat{v}_1$ ($\in [0, \pi/2]$)

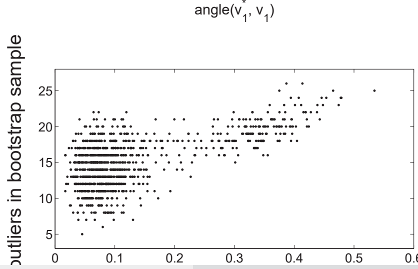
FRB



classical bootstrap



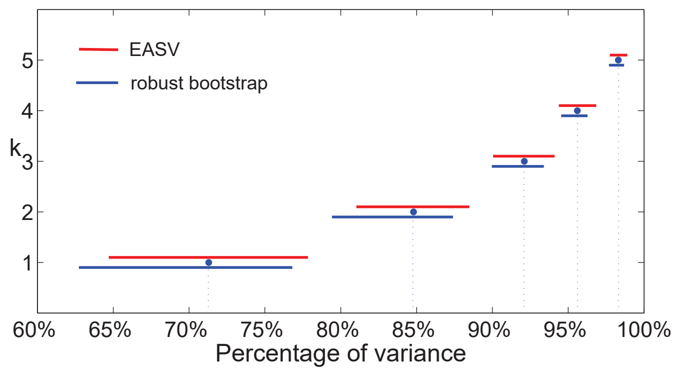




Forged Swiss bank notes data

Percentage of variance explained

95% confidence intervals: robust bootstrap and asymptotic normality



Inference: Outline

1. Robust regression
2. Robust inference
3. Fast and robust bootstrap
4. Robust multivariate location and scatter
5. Inference for robust PCA
6. Inference for robust multivariate regression
7. Robust bootstrap tests in regression
8. Robust multigroup inference
9. Robust model selection
10. Software

Example: School data

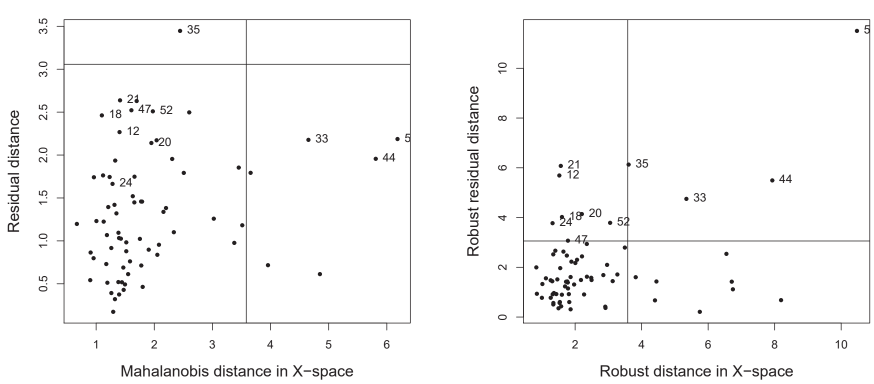
- Explain scores on 3 different tests from 70 school sites by means of 5 explanatory variables
- Responses: reading (y_1), mathematics (y_2), and selfesteem (y_3) score
- Predictors: education level of mother (x_1), highest occupation of a family member (x_2), parent counseling index (x_3), number of teachers (x_4), parental visits index (x_5)
- Model:

$$
\begin{aligned}
y_1 &= \beta_{11} + \beta_{21}x_1 + \beta_{31}x_2 + \beta_{41}x_3 + \beta_{51}x_4 + \beta_{51}x_5 + \epsilon_1 \\
y_2 &= \beta_{12} + \beta_{22}x_1 + \beta_{32}x_2 + \beta_{42}x_3 + \beta_{52}x_4 + \beta_{52}x_5 + \epsilon_2 \\
y_3 &= \beta_{13} + \beta_{23}x_1 + \beta_{33}x_2 + \beta_{43}x_3 + \beta_{53}x_4 + \beta_{53}x_5 + \epsilon_3
\end{aligned}
$$

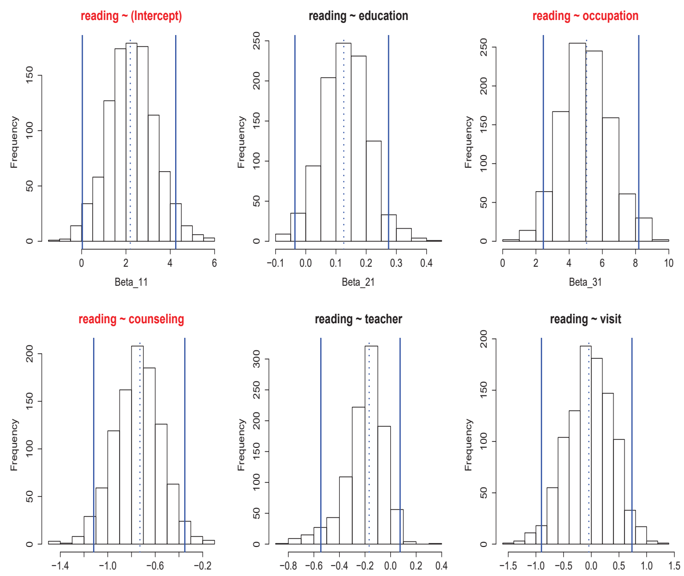
The errors $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \epsilon_3)'$ have center zero and some positive definite scatter matrix $\boldsymbol{\Sigma}$

Multivariate regression example: School data

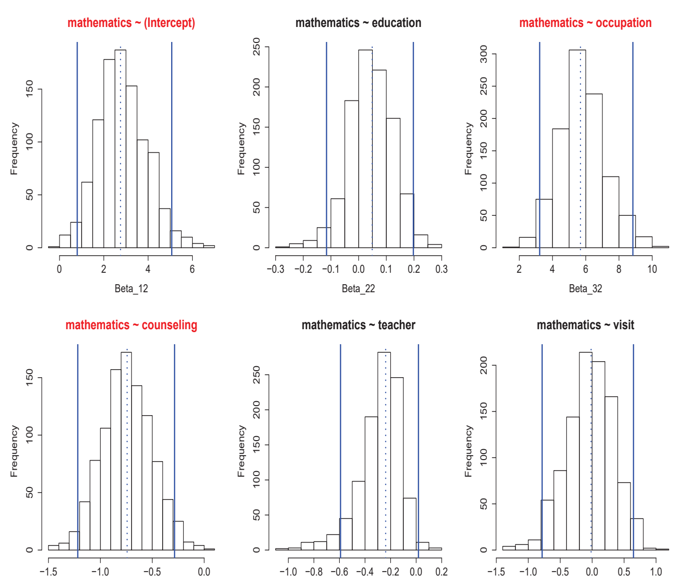
Diagnostic plot: Outlier detection



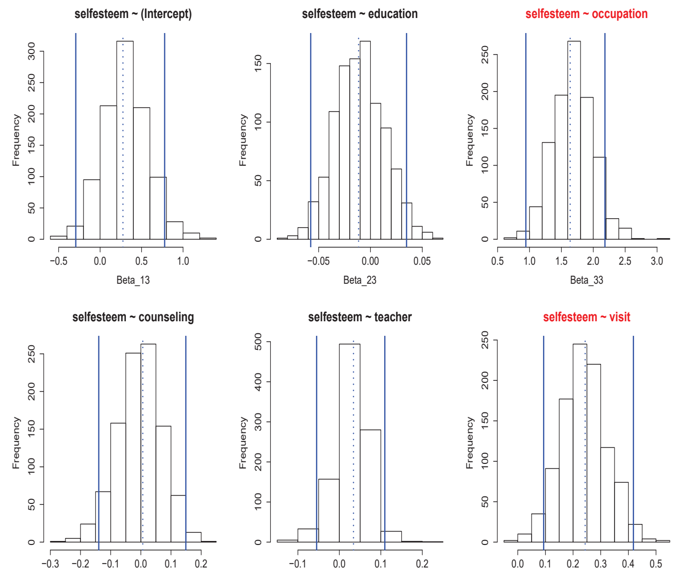
Example: School data



Example: School data



Example: School data



Inference: Outline

1. Robust regression
2. Robust inference
3. Fast and robust bootstrap
4. Robust multivariate location and scatter
5. Inference for robust PCA
6. Inference for robust multivariate regression
7. Robust bootstrap tests in regression
8. Robust multigroup inference
9. Robust model selection
10. Software

Tests for the regression model

- Linear regression model:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

- Denote $\boldsymbol{\beta} = ((\boldsymbol{\beta}^{(1)})', (\boldsymbol{\beta}^{(2)})')'$ with $\boldsymbol{\beta}^{(1)} \in \mathbb{R}^q$ and $\boldsymbol{\beta}^{(2)} \in \mathbb{R}^{p-q}$
- Test for a linear hypothesis

$$H_0 \colon \boldsymbol{\beta}_0^{(2)} = \mathbf{0}$$
$$H_a \colon \boldsymbol{\beta}_0^{(2)} \neq \mathbf{0}$$

Classical approach: the F-test

- $\hat{\boldsymbol{\beta}}_{LS}$, $\hat{\sigma}_{LS}$: least squares estimates in the full model
- $\hat{\boldsymbol{\beta}}_{LS,r}$, $\hat{\sigma}_{LS,r}$: least squares estimates in the reduced model under H_0

$$F = \frac{(\mathsf{SSE}(\hat{\boldsymbol{\beta}}_{LS,r}) - \mathsf{SSE}(\hat{\boldsymbol{\beta}}_{LS}))/(p-q)}{\mathsf{SSE}(\hat{\boldsymbol{\beta}}_{LS})/(n-p)}$$

- Under H_0 : $F \sim F_{p-q,n-p}$ if errors are normal
- Under H_0 : $(p-q)F \sim \chi^2_{p-q}$ asymptotically

$$(p-q)F \approx n\left(\frac{\hat{\sigma}^2_{LS,r} - \hat{\sigma}^2_{LS}}{\hat{\sigma}^2_{LS}}\right)$$

Robust test statistic

Let $\hat{\sigma}_S$ and $\hat{\sigma}_{S,r}$ be the scale estimates corresponding to the S/MM-estimates in the full and reduced model.

Consider the test statistic

$$L_n = n\left(\frac{\hat{\sigma}^2_{S,r} - \hat{\sigma}^2_S}{\hat{\sigma}^2_S}\right)$$

Asymptotic null distribution

Asymptotic null distribution (Salibian-Barrera et al., 2016)

Asymptotically the distribution of the test statistic L_n under H_0 is given by

$$\frac{2DB}{H}L_n \xrightarrow{\mathcal{D}} \chi^2_{p-q}$$

with

$$B = E(\psi_0(u)u)$$
$$D = E(\psi_0'(u))$$
$$H = E(\psi_0^2(u))$$

Robust test

- The asymptotic approximation of the null distribution
  - Only holds well for large samples
  - Is worse for contaminated data
- Can we estimate the null distribution by fast and robust bootstrap?

Fast and robust bootstrap test

- For the full model, set $\widehat{\Theta} = (\hat{\sigma}_S, \hat{\boldsymbol{\beta}}_S)$ and $\widetilde{\Theta}^* = (\tilde{\sigma}^*_S, \tilde{\boldsymbol{\beta}}^*_S)$, then

$$\widehat{\Theta}^*_{FRB} = \widehat{\Theta} + [\mathbf{I} - \nabla \mathbf{g}_n(\widehat{\Theta})]^{-1}(\widetilde{\Theta}^* - \widehat{\Theta})$$

- For the reduced model, set $\widehat{\Theta}_r = (\hat{\sigma}_{S,r}, \hat{\boldsymbol{\beta}}_{S,r})$ and $\widetilde{\Theta}^*_r = (\tilde{\sigma}^*_{S,r}, \tilde{\boldsymbol{\beta}}^*_{S,r})$, then

$$\widehat{\Theta}^*_{r,FRB} = \widehat{\Theta}_r + [\mathbf{I} - \nabla \mathbf{g_r}_n(\widehat{\Theta}_r)]^{-1}(\widetilde{\Theta}^*_r - \widehat{\Theta}_r)$$

- However, the distribution of

$$\tilde{L}^\star_n = n\left(\frac{(\hat{\sigma}^\star_{S,r,FRB})^2 - (\hat{\sigma}^\star_{S,FRB})^2}{(\hat{\sigma}^\star_{S,FRB})^2}\right)$$

  is **inconsistent** because it converges at a faster rate $(1/n)$!

FRB estimate of the null distribution

- Consider a test statistic $L_n = h_n(\widehat{\boldsymbol{\alpha}}_n)$
- We now need that

$$h_n^\star(\widehat{\boldsymbol{\alpha}}_{n,FRB}^\star) \;=\; h_n^\star(\widehat{\boldsymbol{\alpha}}_n^\star) + o_p(1/n)$$

- Taylor expansion of $h_n^\star(\widehat{\boldsymbol{\alpha}}_{n,FRB}^\star)$:

$$h_n^\star(\widehat{\boldsymbol{\alpha}}_{n,FRB}^\star) \;= h_n^\star(\widehat{\boldsymbol{\alpha}}_n^\star) + \nabla h_n^\star(\widehat{\boldsymbol{\alpha}}_n^\star)(\widehat{\boldsymbol{\alpha}}_n^{R\star} - \widehat{\boldsymbol{\alpha}}_n^\star) + o_P(n^{-1})$$

- $\widehat{\boldsymbol{\alpha}}_{n,FRB}^\star - \widehat{\boldsymbol{\alpha}}_n^\star = O_P(n^{-1})$

$\Rightarrow$ we need that

$$\nabla h_n^\star(\widehat{\boldsymbol{\alpha}}_n^\star) = o_P(1)$$

Test statistics for FRB

- Rewrite the test statistics L_n as

$$L_n = n\left(\frac{\hat{\sigma}_{S,r}^2 - \hat{\sigma}_S^2}{\hat{\sigma}_S^2}\right) = n\left(\frac{\hat{\sigma}^2(\hat{\boldsymbol{\beta}}_{S,r}) - \hat{\sigma}^2(\hat{\boldsymbol{\beta}}_S)}{\hat{\sigma}^2(\hat{\boldsymbol{\beta}}_S)}\right) = h_n(\hat{\boldsymbol{\beta}}_{S,r}, \hat{\boldsymbol{\beta}}_S)$$

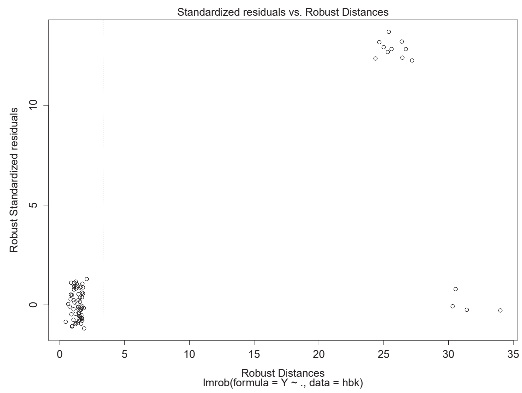
- $\nabla h_n^\star(\hat{\boldsymbol{\beta}}_{S,r}^\star, \hat{\boldsymbol{\beta}}_S^\star) = o_P(1)$
- Set

$$L_{n,FRB}^\star = h_n^\star(\hat{\boldsymbol{\beta}}_{S,r,FRB}^\star, \hat{\boldsymbol{\beta}}_{S,FRB}^\star)$$

  $\Rightarrow$ The distribution of $L_{n,FRB}^\star$ consistently estimates the null distribution of the test statistic L_n.
- Limited extra computational cost

Example: Hawkins-Bradu-Kass data

Cases 1–10 are bad leverage points, cases 11–14 are good leverage points, and the remainder is regular ($n = 75$, $p = 3$).



Robust estimation of multivariate location and scatter

Example: Hawkins-Bradu-Kass data

Model: $y = \beta_{01}x_{i1} + \beta_{02}x_{i2} + \beta_{03}x_{i3} + \beta_{04} + \epsilon_i$

$H_0: \ \beta_{01} = \beta_{03} = 0$

	F-test	L_n (asymp)	$L_{n,FRB}$
Full data	0.00035	0.786	0.564
Regular data	0.252		

Inference: Outline

Multigroup model

- Observations $\boldsymbol{x}_{ji} \in \mathbb{R}^p$ with $j = 1, \ldots, k; \quad i = 1, \ldots, n_j$
- For each group j:

$$\boldsymbol{x}_{ji} = \boldsymbol{\mu}_j + \Sigma^{1/2}\,\boldsymbol{\varepsilon}_{ji} \qquad i = 1, \ldots, n_j$$

- $E(\boldsymbol{\varepsilon}_{ji}) = 0$ and $\text{Cov}(\boldsymbol{\varepsilon}_{ji}) = I_p$
- Distances $d_{ji}(\boldsymbol{\mu}_j, \Sigma) = \sqrt{(\boldsymbol{x}_{ji} - \boldsymbol{\mu}_j)'\Sigma^{-1}(\boldsymbol{x}_{ji} - \boldsymbol{\mu}_j)}$

k-sample S-estimators

k-sample S-estimators (He and Fung, 2000)

The S-estimator of the k locations $\hat{\boldsymbol{\mu}}_{S,1}, \ldots, \hat{\boldsymbol{\mu}}_{S,k}$ and common scatter $\hat{\Sigma}_S^{(k)}$ minimize $|\Sigma|$ subject to

$$\frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\rho_0(d_{ji}\boldsymbol{\mu}_j, \Sigma) = \delta$$

among all $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k \in \mathbb{R}^p$ and symmetric positive definite Σ.

k-sample MM-estimators

k-sample MM-estimators (Van Aelst and Willems, 2011)

1. Put $(\hat{\sigma}_S^{(k)})^2 = |\hat{\Sigma}_S^{(k)}|^{1/p}$, the S-estimate of the generalized scale
2. the MM-estimator of the k locations $\hat{\boldsymbol{\mu}}_{MM,1}, \ldots, \hat{\boldsymbol{\mu}}_{MM,k}$ and common shape $\widehat{\Gamma}_{MM}^{(k)}$ minimize

$$\frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\rho_1\left(\frac{\sqrt{(\boldsymbol{x}_{ji}-\boldsymbol{\mu}_j)^t\Gamma^{-1}(\boldsymbol{x}_{ji}-\boldsymbol{\mu}_j)}}{\hat{\sigma}_S^{(k)}}\right)$$

among all $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k \in \mathbb{R}^p$ and symmetric positive definite Γ with $|\Gamma| = 1$. The MM-estimator of the common covariance matrix is then

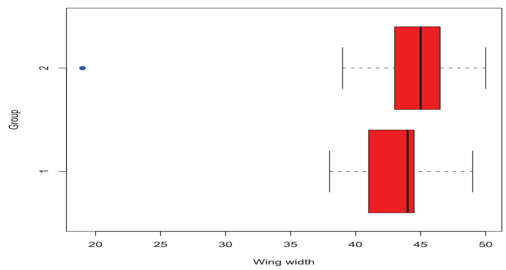
$$\hat{\Sigma}_{MM}^{(k)} = (\hat{\sigma}_S^{(k)})^2\,\widehat{\Gamma}_{MM}^{(k)}$$

.

Example: robust LDA

Biting flies data

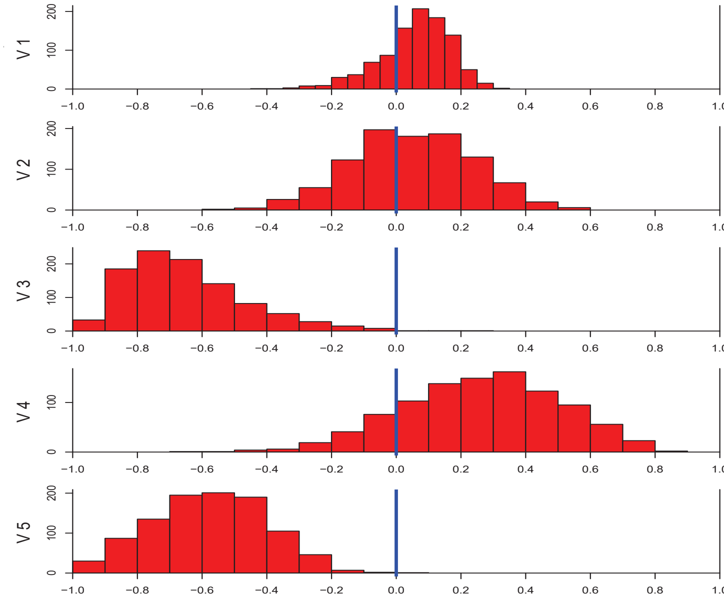
- Two groups of 35 flies (Leptoconops torrens and Leptoconops carteri)
- Measurements of
  - wing length
  - wing width
  - third palp length
  - third palp width
  - fourth palp length



Biting Flies: LDA

- Robust LDA
- Simultaneous two-sample MM-estimates
- FRB inference for the canonical variate
- Variable selection using backward elimination
- Selection criterion: Significance of the discriminant coordinate coefficients

Biting Flies: FRB



Biting Flies: Backward elimination

	Variable				
Model	1	2	3	4	5
1	0.490	0.817	0.006	0.296	0.002
2	0.306	-	0.016	0.216	0.000
3	-	-	0.016	0.096	0.000
4	-	-	0.006	-	0.000

One-way MANOVA model

- Observations $\boldsymbol{x}_{ji} \in \mathbb{R}^p$ with $j = 1, \ldots, k; \quad i = 1, \ldots, n_j$
- For each group j:

$$\boldsymbol{x}_{ji} = \boldsymbol{\mu}_j + \Sigma^{1/2}\,\boldsymbol{\varepsilon}_{ji} \qquad i = 1, \ldots, n_j$$

- $E(\boldsymbol{\varepsilon}_{ji}) = 0$ and Cov$(\boldsymbol{\varepsilon}_{ji}) = I_p$
- MANOVA test:

$$\begin{aligned} H_0 :& \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_k \\ H_a :& \boldsymbol{\mu}_r \neq \boldsymbol{\mu}_s \text{ for at least one } r \neq s \end{aligned}$$

Classical test: Wilk's Lambda

$$\Lambda_n = \frac{|\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\boldsymbol{x}_{ji} - \overline{\boldsymbol{x}}_j)(\boldsymbol{x}_{ji} - \overline{\boldsymbol{x}}_j)^t|}{|\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\boldsymbol{x}_{ji} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{ji} - \overline{\boldsymbol{x}})^t|}$$

- LRT assuming that $F_j = N(\boldsymbol{\mu}_j, \Sigma)$
- Asymptotic $\chi^2_{p(k-1)}$ distribution
- Sensitive to outliers

A robust one-way MANOVA test statistic

Based on the one-sample and k-sample S-estimates, consider the test statistic

$$\Lambda_n = \frac{|\hat{\Sigma}_S^{(k)}|}{|\hat{\Sigma}_S^{(1)}|} \equiv \frac{\hat{\sigma}_S^{(k)}}{\hat{\sigma}_S^{(1)}} = \frac{\hat{\sigma}(\hat{\boldsymbol{\mu}}_{S,1}, \ldots, \hat{\boldsymbol{\mu}}_{S,k}, \widehat{\Gamma}_S^{(k)})}{\hat{\sigma}(\hat{\boldsymbol{\mu}}_S, \widehat{\Gamma}_S^{(1)})}$$
$$= h_n(\hat{\boldsymbol{\mu}}_{S,1}, \ldots, \hat{\boldsymbol{\mu}}_{S,k}, \widehat{\Gamma}_S^{(k)}, \hat{\boldsymbol{\mu}}_S, \widehat{\Gamma}_S^{(1)})$$

The distribution of

$$\Lambda^\star_{n,FRB} = h^\star_n(\hat{\boldsymbol{\mu}}^\star_{S,1,FRB}, \ldots, \hat{\boldsymbol{\mu}}^\star_{S,k,FRB}, \widehat{\Gamma}^{(k)^\star}_{S,FRB}, \hat{\boldsymbol{\mu}}^\star_{S,FRB}, \widehat{\Gamma}^{(1)^\star}_{S,FRB})$$

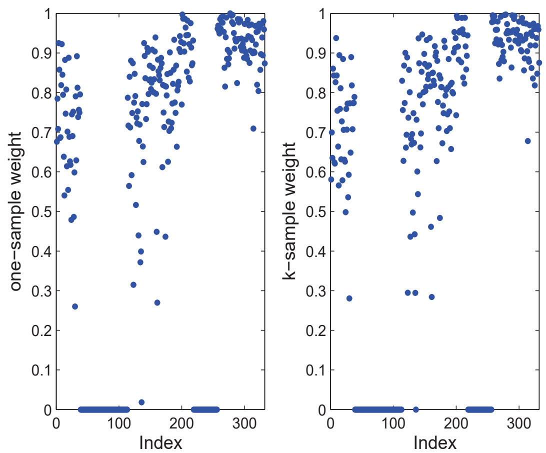
consistently estimates the null distribution of the test statistic Λ_n.

Example: Oslo transect data

Oslo transect data

- 360 samples of different plant species collected along a transect running through Oslo, Norway.
- Factor `lithology` which consists of four levels ($k = 4$)
- Three elements: `P`, `K` and `Zn` ($p = 3$)

MM-estimator weights (90% efficiency)



MANOVA results

	Clas	$\Lambda_{n,FRB}$	$S\Lambda_n^{2a}$	$S\Lambda_n^{2b}$	$MM\Lambda_n^a$	$MM\Lambda_n^b$
p	.704	.016	.015	.014	.018	.019

Inference: Outline

Selecting a linear regression model

- Dataset $\mathcal{Z}_n = \{(y_i, x_{i1}, \ldots, x_{ip}) = (y_i, \boldsymbol{x}_i); i = 1, \ldots, n\} \subset \mathbb{R}^{p+1}$
- $X_1, \ldots, X_p$ are the candidate regressors
- The full model: $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \qquad i = 1, \ldots, n$
- An estimate $\hat{\boldsymbol{\beta}}$ may become unstable if there are
  - several noise variables among the candidate regressors
  - highly correlated predictors (multicollinearity)
- Model selection can improve the unstable coefficient estimates
  - Trade (a little) bias for (a large) variability reduction!
  - Enhance interpretability: most relevant effects

Submodels

- Dataset $\mathcal{Z}_n = \{(y_i, x_{i1}, \ldots, x_{ip}); i = 1, \ldots, n\} \subset \mathbb{R}^{p+1}$.
- Let $\alpha \subset \{1, \ldots, p\}$ denote the predictors included in a submodel
- The corresponding submodel is:

$$y_i = \boldsymbol{x}'_{\alpha i}\boldsymbol{\beta}_\alpha + \varepsilon_{\alpha i} \qquad i = 1, \ldots, n.$$

A selected model is considered a good model if

- It is parsimonious
- It fits the data well
- It yields good predictions for similar data

$\rightarrow$ Use selection criteria!

Final prediction error

- Final prediction error $FPE(\alpha) = \frac{1}{\sigma^2}\sum_{i=1}^{n} E\left[(z_i - \boldsymbol{x}'_{\alpha i}\hat{\boldsymbol{\beta}}_\alpha)^2\right]$
- Based on LS, $FPE(\alpha)$ can be estimated by

$$\widehat{FPE}(\alpha) = \frac{RSS(\alpha)}{\hat{\sigma}^2_{LS}} + 2d(\alpha)$$

- $\hat{\sigma}_{LS}$ is the residual scale estimate in the "full" model α_f.
  Usually, $\alpha_f = \{1, \ldots, p\}$

Robust FPE

- Final prediction error $FPE(\alpha) = \dfrac{1}{\sigma^2}\sum_{i=1}^{n} E\left[(z_i - \boldsymbol{x}'_{\alpha i}\hat{\boldsymbol{\beta}}_\alpha)^2\right]$
- Robust final prediction error:

$$RFPE(\alpha) = \sum_{i=1}^{n} E\left[\rho\left(\frac{z_i - \boldsymbol{x}'_{\alpha i}\hat{\boldsymbol{\beta}}_\alpha}{\sigma}\right)\right]$$

  with ρ a bounded loss function
- An estimate of $RFPE(\alpha)$ is given by

$$\widehat{RFPE}(\alpha) = \sum_{i=1}^{n} \rho(r_i(\hat{\boldsymbol{\beta}}_\alpha)/\hat{\sigma}_n) + p(\alpha)\frac{\sum_{i=1}^{n}\psi^2(r_i(\hat{\boldsymbol{\beta}}_\alpha)/\hat{\sigma}_n)}{\sum_{i=1}^{n}\psi'(r_i(\hat{\boldsymbol{\beta}}_\alpha)/\hat{\sigma}_n)}$$

- $\hat{\sigma}_n$ is the robust scale estimate of a 'full' model α_f.
  Usually, $\alpha_f = \{1,\dots,p\}$

Bootstrap based selection criteria

- The (conditional) expected prediction error:

$$\text{PE}(\alpha) = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(z_i - \boldsymbol{x}'_{\alpha i}\hat{\boldsymbol{\beta}}_\alpha\right)^2 \middle| y, X\right],$$

- Estimates of PE(α) can be obtained by bootstrap
- A more advanced selection criterion takes both goodness-of-fit and PE into account:

$$\text{PPE}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \boldsymbol{x}'_{\alpha i}\hat{\boldsymbol{\beta}}_\alpha\right)^2 + f(n)\,p(\alpha) + E^\star\left[\frac{1}{n}\sum_{i=1}^{n}\left(z_i - \boldsymbol{x}'_{\alpha i}\hat{\boldsymbol{\beta}}_\alpha\right)^2 \middle| y, X\right]$$

Robust bootstrap selection criteria

Robust equivalents of the bootstrap based selection criteria:

$$\widehat{RPE}(\alpha) = \frac{\hat{\sigma}_n^2}{n} E^\star \left[\sum_{i=1}^{n} \rho \left(\frac{z_i - \boldsymbol{x}'_{\alpha i}\hat{\boldsymbol{\beta}}_\alpha}{\hat{\sigma}_n} \right) \middle| y, X \right]$$

$$\widehat{PRPE}(\alpha) = \frac{\hat{\sigma}_n^2}{n} \left\{ \sum_{i=1}^{n} \rho \left(\frac{y_i - \boldsymbol{x}'_{\alpha i}\hat{\boldsymbol{\beta}}_\alpha}{\hat{\sigma}_n} \right) + f(n)\, p(\alpha) \right\} + \widehat{RPE}(\alpha)$$

- ρ is a bounded loss function
- $f(n)d(\alpha)$ is the penalty term with e.g. $f(n) = 2\log n$
- $\hat{\sigma}_n$ is the robust scale estimate of a 'full' model α_f. Usually, $\alpha_f = \{1, \ldots, p\}$
- $E^\star$ is a bootstrap estimate of the expected value

FRB based selection criteria

$$\widehat{RPE}(\alpha) = \frac{\hat{\sigma}_n^2}{n} E^\star_{FRB} \left[\sum_{i=1}^{n} \rho \left(\frac{z_i - \boldsymbol{x}'_{\alpha i}\hat{\boldsymbol{\beta}}_\alpha}{\hat{\sigma}_n} \right) \middle| y, X \right]$$

$$\widehat{PRPE}(\alpha) = \frac{\hat{\sigma}_n^2}{n} \left\{ \sum_{i=1}^{n} \rho \left(\frac{y_i - \boldsymbol{x}'_{\alpha i}\hat{\boldsymbol{\beta}}_\alpha}{\hat{\sigma}_n} \right) + f(n)\, p(\alpha) \right\} + \widehat{RPE}(\alpha)$$

- ρ is the MM loss function and $\hat{\beta}_\alpha$ is the MM estimate
- $E^\star_{FRB}$ is a bootstrap estimate of the expected value using FRB estimates of $\boldsymbol{\beta}_\alpha$ based on bootstrap samples of size $m \leq n$

Consistent model selection

Suppose a true model $\alpha_0 \subset \{1, \ldots, p\}$ exists and is included in the set $\mathcal{A}$ of models considered.

If we select the model that minimizes $\widehat{RPE}(\alpha)$ or $\widehat{PRPE}(\alpha)$, that is

$$\hat{\alpha}_{m,n} = \text{argmin}_{\alpha \in \mathcal{A}} \widehat{RPE}(\alpha) \text{ and } \tilde{\alpha}_{m,n} = \text{argmin}_{\alpha \in \mathcal{A}} \widehat{PRPE}(\alpha),$$

then, under appropriate regularity conditions, the model selection criteria are consistent in the sense that

$$\lim_{n \to \infty} P(\hat{\alpha}_{m,n} = \alpha_0) = 1 \text{ and } \lim_{n \to \infty} P(\tilde{\alpha}_{m,n} = \alpha_0) = 1\,.$$

Two conditions have practical consequences

- $m = o(n)$ (m out of n bootstrap)
- $f(n) = o(n/m)$

Variable selection strategies

1. Choose a selection criterion
2. Follow a model selection strategy
   - All subsets $\rightarrow$ **too time consuming**
   - Backward elimination
   - Forward selection
   - Stepwise selection
3. Select optimal model(s)
4. Evaluate their performance

Examples: backward elimination

- We compare the full model with models selected by backward elimination based on
  - $\widehat{RFPE}(\alpha)$
  - $\widehat{RPE}(\alpha)$
  - $\widehat{PRPE}(\alpha)$ with $f(n) = \log(n)$
- For each of the models we report an adjusted robust R^2
- To compare predictive power we consider the robust 5-fold CV MSPE (5% trimming)

Example: Los Angeles Ozone Pollution Data

- 366 observations (different days) on 9 variables
- Response: temperature (degrees F) at El Monte, CA
- Covariates: Measurements of temperature, pressure, humidity, ozone, etc at other places in CA.
- We start from the full quadratic model ($p = 45$)

model	$p(\alpha)$	RR_a^2	5% Trimmed MSPE
Full	45	0.8660	10.78
RFPE	23	0.8174	10.66
RPE	10	0.7583	11.67
PRPE	7	0.7643	10.45

Example: Diabetes data

- 442 observations on 16 variables.
- Response: Measure of disease progression after one year
- Covariates: 10 baseline variables (age, sex, BMI, , ...)
- We start from a quadratic model with some interactions ($p = 65$)

model	$p(\alpha)$	RR_a^2	5% Trimmed MSPE
Full	65	0.7731	4988.1
RFPE	16	0.6045	2231.2
RPE	11	0.5127	2657.2
PRPE	7	0.5302	2497.0

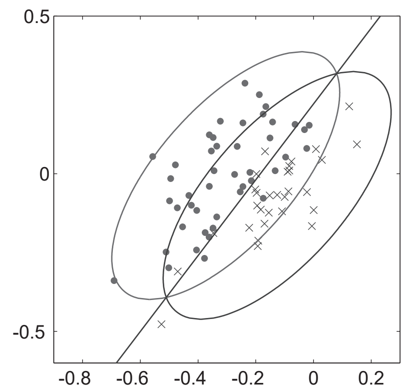
Example: Linear discriminant analysis

Hemophilia data

- 2 groups
- $n = 75$ training samples
  - $n_1 = 30$ controls
  - $n_2 = 45$ hemophilia A carriers
- 2 variables

Robust LDA based on S-estimates

Discriminant line and 97.5% tolerance ellipses

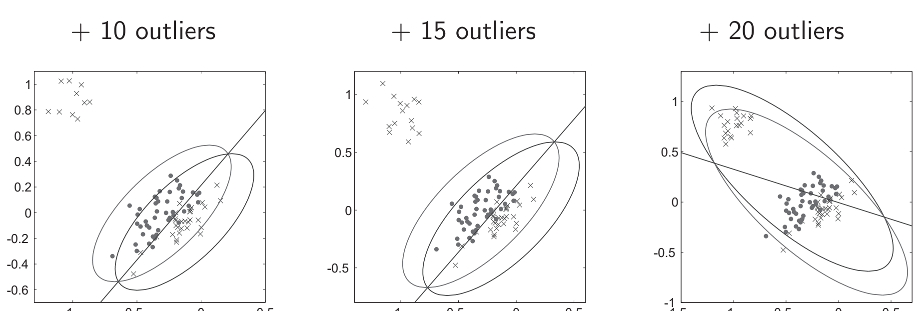


No outliers

Robust estimation of multivariate location and scatter

Robust LDA based on S-estimates

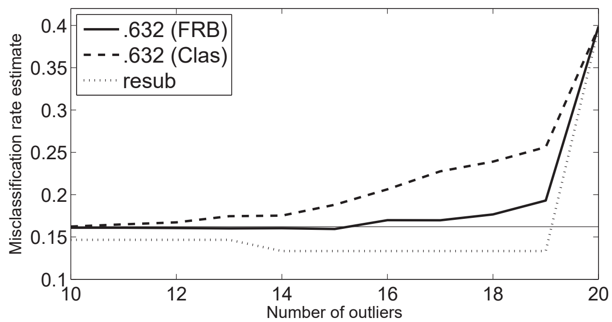
Discriminant line and 97.5% tolerance ellipses



FRB error rates for LDA based on S-estimates

We use the 0.632 estimator to estimate the error rate:

$$\widehat{\text{err}}_{.632} = .632\ \widehat{\text{err}}_{\text{boot}} + .368\ \widehat{\text{err}}_{\text{resub}}$$



Inference: Outline

1. Robust regression
2. Robust inference
3. Fast and robust bootstrap
4. Robust multivariate location and scatter
5. Inference for robust PCA
6. Inference for robust multivariate regression
7. Robust bootstrap tests in regression
8. Robust multigroup inference
9. Robust model selection
10. Software

Software

Software for FRB inference: R package *FRB* described in Van Aelst and Willems (2013).

References

- He, X. and Fung, W.K. (2000). "High breakdown estimation for multiple populations with applications to discriminant analysis," *Journal of Multivariate Analysis*, 72, 151–162.
- Salibian-Barrera, M. and Zamar, R.H. (2002). "Bootstrapping robust estimates of regression," *The Annals of Statistics*, 30, 556–582.
- Salibian-Barrera, M., and Van Aelst, S. (2008). "Robust model selection using fast and robust bootstrap," *Computational Statistics & Data Analysis*, 52, 5121–5135.
- Salibian-Barrera, M., Van Aelst, S. and Willems, G. (2006). "PCA based on multivariate MM-estimators with fast and robust bootstrap," *Journal of the American Statistical Association*, 101, 1198–1211.
- Salibian-Barrera, M., Van Aelst, S. and Willems, G. (2008). "Fast and robust bootstrap," *Statistical Methods and Applications*, 17, 41–71.

References

- Salibian-Barrera, M., Van Aelst, S., and Yohai, V. (2016). "Robust tests for linear regression models based on τ-estimates," *Computational Statistics & Data Analysis*, 93, 436–455.
- Van Aelst, S. and Willems, G. (2011). "Robust and efficient one-way MANOVA tests," *Journal of the American Statistical Association*, 106, 706–718.
- Van Aelst, S., and Willems, G. (2013), "Fast and robust bootstrap for multivariate inference: the R package FRB," *Journal of Statistical Software*, 53, 1–32.