Inference In Robust Regression: the impact of scale estimation

Manuel Koller and Werner A. Stahel, ETH Zürich

June 2009, ICORS

1. Introduction

Notation and Assumptions

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + e_i \quad \text{for } i = 1, \ldots, n \qquad \mathbf{x}_i' = (1, x_{i1}, \ldots, x_{ip-1})$$

$$r_i(\hat{\boldsymbol{\beta}}) = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}} \qquad \boldsymbol{\beta}' = (\beta_0, \ldots, \beta_{p-1})$$

$$\frac{e_i}{\sigma} \sim F \text{ i.i.d. scale family, symmetric}$$

MM-estimation

1. high breakdown S-estimate ($\rightarrow \hat{\boldsymbol{\beta}}_S, \hat{\sigma}_S$)
2. high efficiency M-estimate ($\rightarrow \hat{\boldsymbol{\beta}}$)

(Sometimes simultaneous M-regression and M-scale estimate between 1 and 2.)

Importance of $\hat{\sigma}$

- Tests of parameters: Usually Wald-type tests based on estimated covariance matrix (from asymptotic theory): Compare $\hat{\beta}_i / \widehat{\mathrm{se}}_i(\hat{\boldsymbol{\beta}})$ to t_{n-p}.

$$\widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 \cdot \frac{\frac{n}{n-p} \operatorname{ave}_i \psi(\frac{r_i}{\hat{\sigma}})^2}{\left[\operatorname{ave}_i \psi'(\frac{r_i}{\hat{\sigma}})\right]^2} \cdot K^2 \cdot \left(\frac{\mathbf{X'WX}}{\operatorname{ave}_i w_i}\right)^{-1} \qquad (1)$$

  (likelihood ratio tests may be better)

- Prediction intervals: well estimated $\hat{\sigma}$ crucial

Outline

1. Introduction (done)
2. Design adapted scale estimate
3. Simulation study
4. Covariance matrix of $\hat{\beta}$
5. Conclusions and Outlook

2. Design adapted scale estimate

M-scale estimate

$$\underset{i}{\text{ave}}\,\chi\left(\frac{r_i}{\hat{\sigma}_S}\right) = \kappa := \int \chi(e)dF(e) \tag{2}$$

OLS: $\text{var}(r_i) = (1-h_i)\sigma^2$ $\left(\implies \mathbf{E}\left[\sum_{i=1}^n r_i^2\right] = (n-p)\sigma^2\right)$

$$\underset{i}{\text{ave}}\,\chi\left(\frac{r_i}{\tau_i\hat{\sigma}_D}\right) = \kappa \tag{3}$$

In case of OLS: $\tau_i = \sqrt{1-h_i}$

Calculation of τ_i

Use *von Mises first order expansion*:

$$r_i = y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_n = y_i - \mathbf{x}_i' \left(\boldsymbol{\beta} + \frac{1}{n} \sum_{h=1}^{n} \mathbf{IF}(e_h, \mathbf{x}_h, \sigma) + \text{remainder} \right)$$
$$\approx g(e_i, \mathbf{x}_i, \sigma) + u_{-i}(\mathbf{e}, \mathbf{X}, \sigma)$$

u_{-i}: sum of $n-1$ terms, therefore approximately normal with variance s_i^2. We get an implicit formula for τ_i:

$$\int \chi \left(\frac{g(e_i, x_i, \sigma) + u}{\tau_i \sigma} \right) dF\left(\frac{e_i}{\sigma}\right) d\,\mathcal{N}_{0,s_i^2}(u) = \kappa \qquad (4)$$

χ-Function

Natural choice: estimation of $\hat{\sigma}$ analogue to weighted least squares.

Use robustness weights from M-regression estimator:

$$\chi(r) = wr^2 = \frac{\psi(r)}{r} r^2 = \psi(r) r \tag{5}$$

$$\left(\text{use ave}_i \, \chi\left(\frac{r_i}{t_i \hat{\sigma}_D}\right) = \kappa \, \text{ave} \, w_i\right)$$

Influence on scale $\to 0$ for $r \to \infty$.

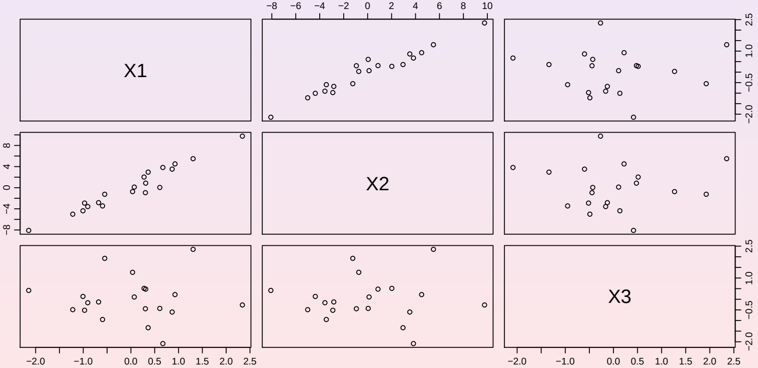
MMDW-estimation

Four step procedure:

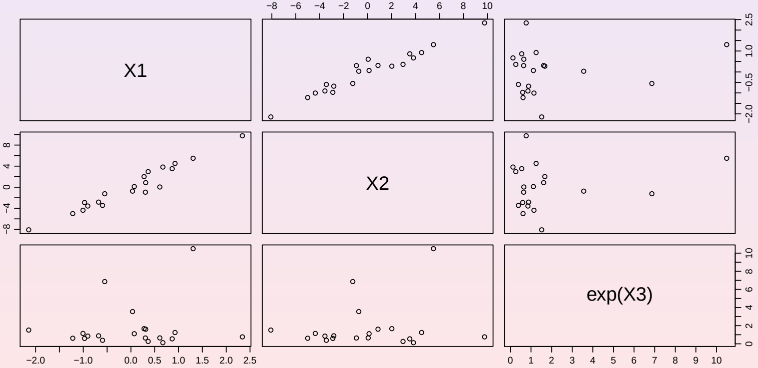
1. high breakdown S-estimate ($\rightarrow \hat{\beta}_S, \hat{\sigma}_S$)
2. high efficiency M-estimate
3. Design adapted scale estimate ($\rightarrow \hat{\sigma}_D$)
4. high efficiency M-estimate ($\rightarrow \hat{\beta}$)

3. Simulation study

Design 1 ($n = 20$)



Design 2 ($n = 20$)

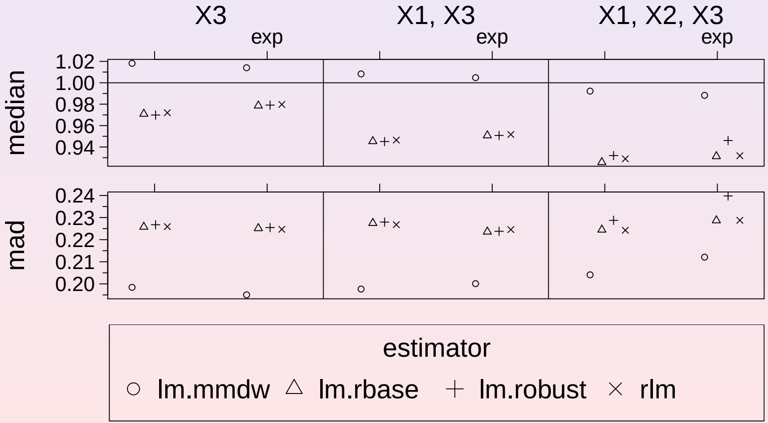

X1
X2
exp(X3)
−2.0
−1.0
0.0
0.5
1.0
1.5
2.0
2.5
−8
−6
−4
−2
0
2
4
6
8
10
0
1
2
3
4
5
6
7
8
9
10
−8
−4
0
4
8
2.5
1.0
−0.5
−2.0
0
2
4
6
8
10

Simulation parameters

- Simulate 2000 repetitions, $\boldsymbol{\beta} = \mathbf{0}$, $n = 20$.
- Error distributions: $\mathcal{N}(0, 1)$, t_3 and Cauchy
- bisquare ψ functions tuned for 95% efficiency at the normal ($c = 4.685$).
- Estimators:

`lm.mmdw`	as described above
`rlm`	from package `MASS`
`lm.rbase`	function `lmrob` from package `robustbase`
`lm.robust`	function `lmRob` from package `robust`

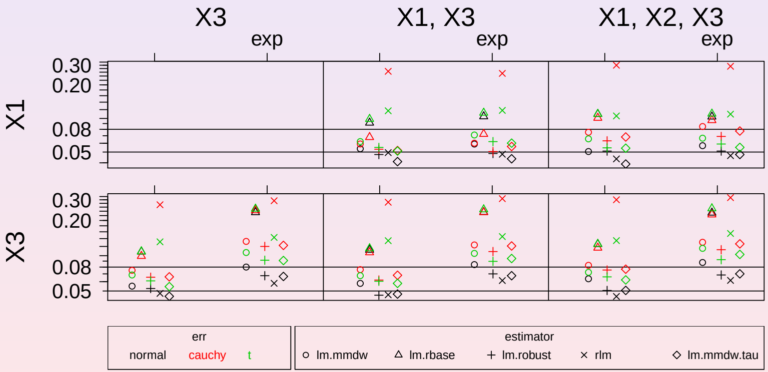
$\hat{\sigma}$ statistics: median and mad (error: $\mathcal{N}(0,1)$)



4. Covariance matrix estimate

$\widehat{\text{cov}}(\hat{\beta}) =$	$\hat{\sigma}^2$	$\frac{\frac{n}{n-p}\,\text{ave}_i\,\psi(\frac{r_i}{\hat{\sigma}})^2}{\left[\text{ave}_i\,\psi'(\frac{r_i}{\hat{\sigma}})\right]^2}$	K^2	$\text{ave}_i\, w_i (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$
`lm.mmdw`	σ_D^2	x	x	x
`rlm`	x	x	x	no weights
`lm.rbase`	x	asymmetry and heteroskedasticity		
`lm.robust`	x	x	x	x
`lm.mmdw.tau`	σ_D^2	uses τ_i		x

Level statistics: $\# \left[\left| \frac{\hat{\beta}_i}{\hat{\mathrm{se}}_i} \right| > qt(0.975, \mathbf{df}) \right] \Big/ 2000$ **(log scale)**



5. Conclusions and Outlook

- $\hat{\sigma}$ is important
- New design adapted scale estimate $\hat{\sigma}_D$:
  simulation results promising
- Likelihood ratio tests may be better.
- Estimation of covariance matrix not so clear,
  some implementations are more robust than others.

References

M. Koller (2008) Robust Statistics: Tests for Robust Linear Regression. Master's Thesis, http://stat.ethz.ch/research/dipl_arb/2008/koller.pdf.

P. J. Huber (1973). Robust regression: Asymptotics, conjectures, and Monte Carlo. Ann. Statist. 1, 799-821.

Maronna et al (2006). Robust Statistics, Theory and Methods. Wiley Series in Probability and Statistics, Wiley, New York.