Robust regression: the influence of ψ-functions, improving scale and covariance matrix estimates for tests, confidence and prediction intervals

Manuel Koller and Werner A. Stahel, ETH Zürich

June 2010, ICORS

Introduction

Notation and Assumptions

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

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

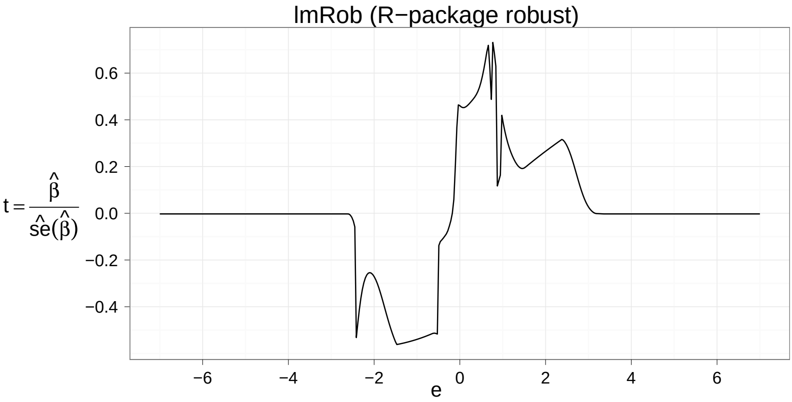
$$\frac{e_i}{\sigma} \sim F \text{ i.i.d. scale family, symmetric}$$

MM-estimation

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

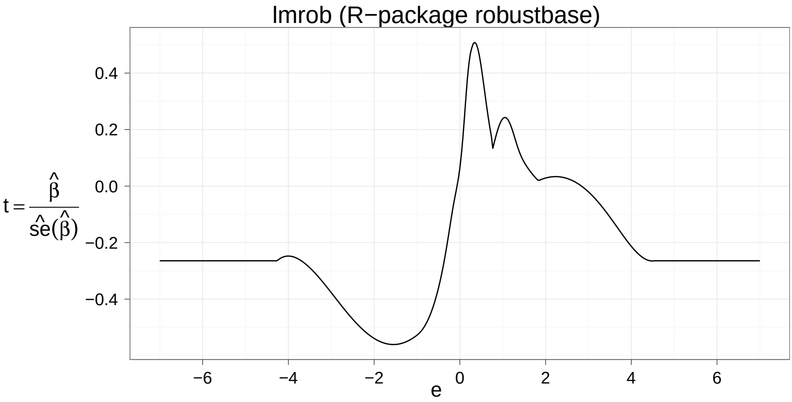
Current Methods

Sensitivity curve of the t-value of the intercept ($n = 20$, $p = 4$).



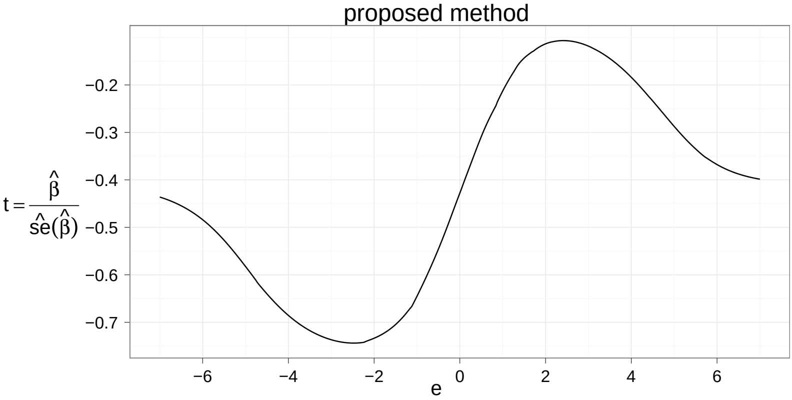
Current Methods

Sensitivity curve of the t-value of the intercept ($n = 20$, $p = 4$).



Proposed Method

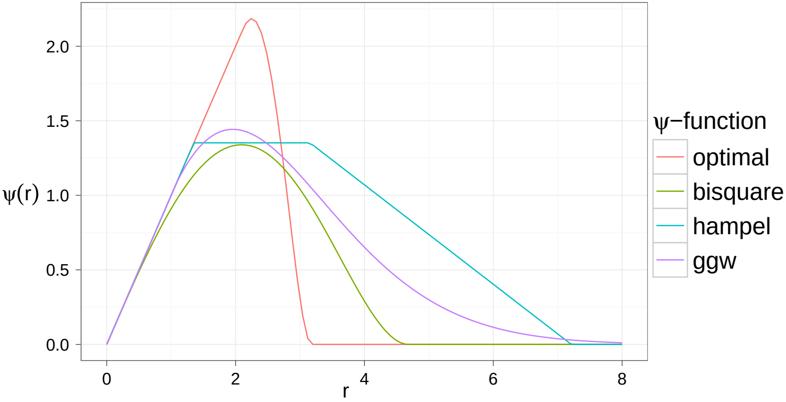
New scale and alternative covariance matrix estimate.



Outline

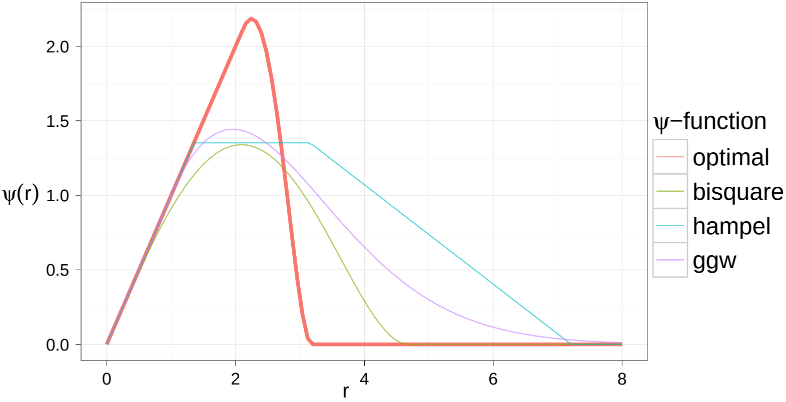
2. ψ-functions

All redescending and tuned for 95%-efficiency



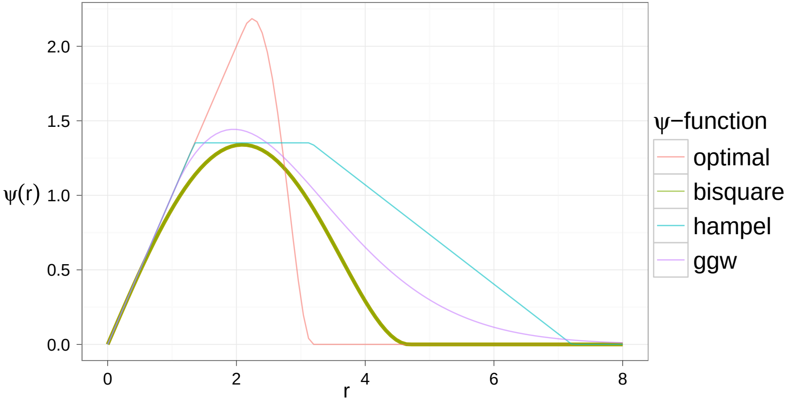
Optimality?

Optimal (as introduced by Yohai and Zamar (1997)) minimizes contamination sensitivity



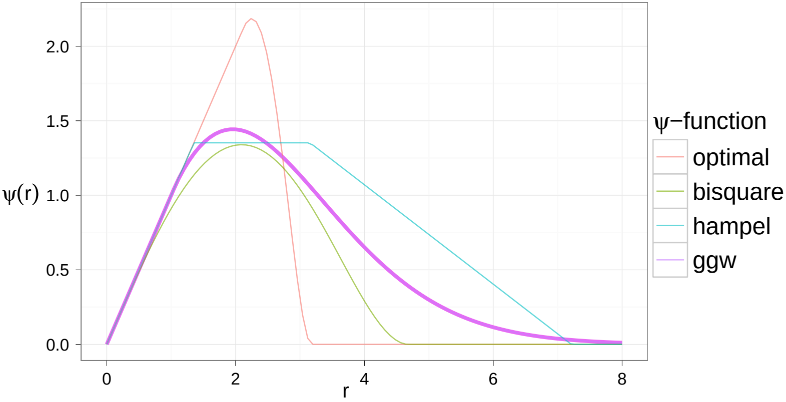
Optimality?

Bisquare is smooth



Optimality?

Generalized Gaussweight (ggw) good for inference and p/n ratios "not very small"



3. Inference for MM-estimates

For convenience: use Wald-type inference

From an estimate of the covariance matrix, we can construct tests, confidence and prediction intervals based on asymptotic normality of the estimate.

Other possibilities

use likelihood-ratio-type or saddlepoint test.

Wald type inference

Based on some regularity conditions we have asymptotic normality of $\hat{\beta}$ with covariance matrix

$$\mathrm{cov}(\hat{\beta}) = \sigma^2 \gamma \mathbf{V}_{\mathbf{X}}^{-1} = \sigma^2 \frac{\mathbf{E}\,\psi^2}{(\mathbf{E}\,\psi')^2} \left(\mathbf{E}\,\mathbf{x}\mathbf{x}'\right)^{-1} \quad (1)$$

Based on three parts:

- Scale σ
- Correction factor γ
- Matrix part $\mathbf{V}_{\mathbf{X}}$

Want to get all three right. Then expect result also to be correct.

Scale: last year

Use a standardization factor of the form

$$\tau_i = (1 - c_1 h_i)\sqrt{1 - c_2 h_i} \tag{2}$$

where h_i is the leverage of the i-th observation. c_1 and c_2 depend on ψ.

(In case of OLS: $\tau_i = \sqrt{1 - h_i}$, $\quad \sum \tau_i^2 = n - p$.)

And solve the following equation for $\hat{\sigma}$.

$$\sum_{i=1}^{n} \tau_i^2 \, \mathrm{W}\left(\frac{r_i}{\tau_i \hat{\sigma}}\right) \left[\left(\frac{r_i}{\tau_i \hat{\sigma}}\right)^2 - \kappa\right] = 0 \tag{3}$$

where $W(r) = \psi(r)/r$ is the function that produces the robustness weights.

Correction factor

In case of MM-estimation we have $\gamma = \frac{\mathbf{E}\,\psi}{(\mathbf{E}\,\psi')^2}$.

Use the standardized residuals to estimate asymptotic normality correction factor:

$$\hat{\gamma} = \frac{\frac{1}{n}\sum_{i=1}^{n} \psi\left(\frac{r_i}{\tau_i \hat{\sigma}}\right)^2}{\left[\frac{1}{n}\sum_{i=1}^{n} \psi'\left(\frac{r_i}{\tau_i \hat{\sigma}}\right)\right]^2} \qquad (4)$$

Current implementations use Huber's small sample correction:

$$\hat{\gamma} = K^2 \frac{\frac{1}{n-p}\sum_{i=1}^{n} \psi\left(\frac{r_i}{\hat{\sigma}}\right)^2}{\left[\frac{1}{n}\sum_{i=1}^{n} \psi'\left(\frac{r_i}{\hat{\sigma}}\right)\right]^2} \qquad (5)$$

Matrix part

Rejected observations should not play a role in inference. Therefore we also use the weights to calculate $\mathbf{E}\,\mathbf{x}\mathbf{x}'$:

$$\widehat{\mathbf{E}\,\mathbf{x}\mathbf{x}'} = n \left[\frac{\sum_{i=1}^{n} W\left(\frac{r_i}{\hat{\sigma}}\right) \mathbf{x}_i \mathbf{x}_i'}{\sum_{i=1}^{n} W\left(\frac{r_i}{\hat{\sigma}}\right)} \right] \tag{6}$$

where $W(r) = \psi(r)/r$ is the function that produces the robustness weights.

(This is how its usually done.)

Note: Huber's small sample correction does not take this into account.

4. Simulation Study

Criteria for choosing a ψ-function:

- maximum asymptotic bias for given efficiency
  - efficiency of the estimates for p/n "not very small" ($\hat{\beta}$)
  - good properties of the scale ($\hat{\sigma}$)
  - nominal level of the resulting tests and confidence intervals (t-statistic)
- coverage probability of the prediction intervals
- length of confidence intervals ($\rightarrow$ power!)

4. Simulation Study

Criteria for choosing a ψ-function:

- maximum asymptotic bias for given efficiency
- efficiency of the estimates for p/n "not very small" ($\hat{\beta}$)
- good properties of the scale ($\hat{\sigma}$)
- nominal level of the resulting tests and confidence intervals (t-statistic)
- coverage probability of the prediction intervals
- length of confidence intervals ($\rightarrow$ power!)

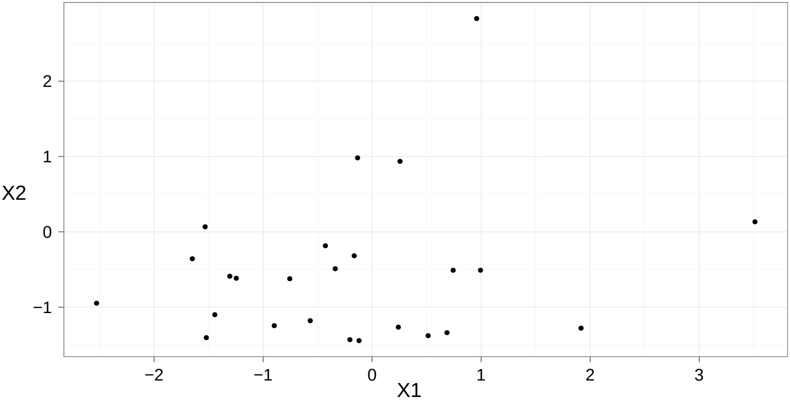
Setting

We used:

- multiple fixed as well as random designs for various n, p combinations
- a variety of error distributions (normal, t, contaminated normal, skewed t); (The same distribution was also used to generate the designs in the random design case.)
- variations of methods discussed above (based on a modified version of `lmrob`)
- for comparison we also ran the simulations on `lmRob`
- 1000 repetitions

Simulation design like in Maronna and Yohai (2009).
Only results for random designs will be shown.

Example design



For $n = 25$, $p = 2$, skew-t distribution with $df = 5$ and $\gamma = 2$.

Simulated Methods: Legend

SM

Standard MM-estimate (initial S-, final M-estimate)

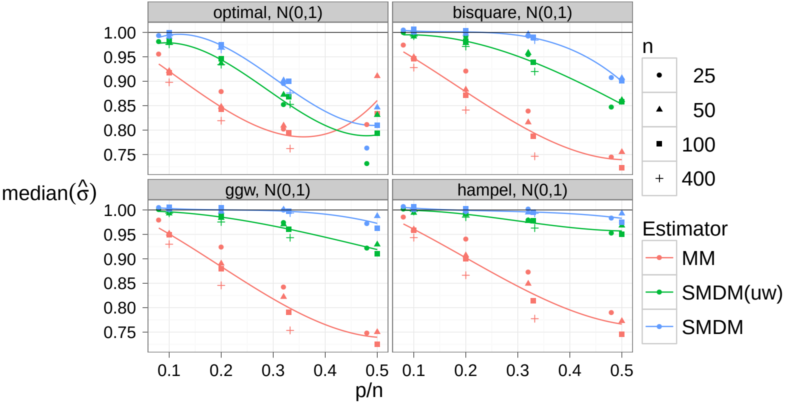
SMDM

MM-estimate, followed by D-estimate and M-estimate again.

SMDM(uw)

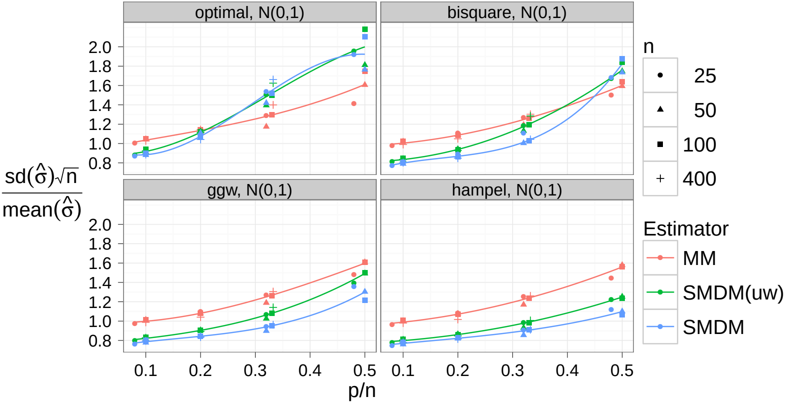
as SMDM, but using unweighted leverages

Results: Scale bias



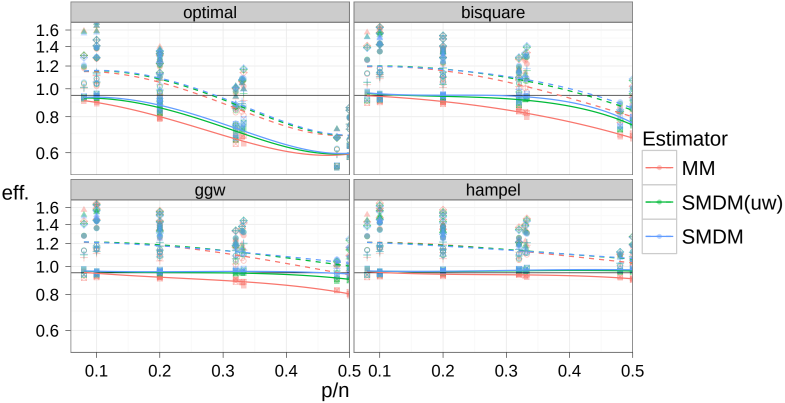
(Simulated error distribution: standard normal)

Results: Scale efficiency



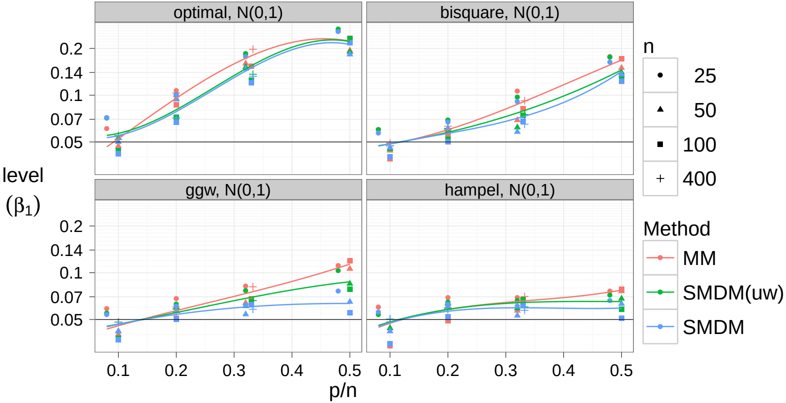
(Simulated error distribution: standard normal, calculating mean and sd with 10% trimming.)

Results: Efficiency of $\hat{\beta}$



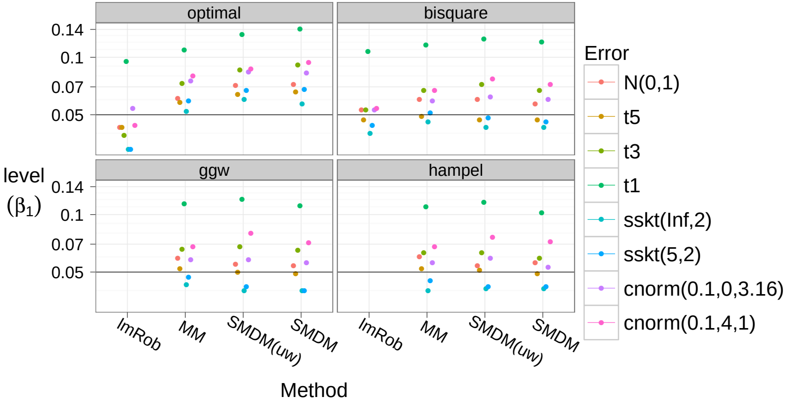
Solid line connects standard normal results, the dashed line is calculated using all results. Shape: error distribution. (Comparing to an OLS estimate and calculating the average with 10% trimming.)

Results: Empirical nominal levels



(Covariance matrix estimates as before; simulated error distribution: standard normal)

Results: Empirical nominal levels



For $n = 25$, $p = 2$, and covariance matrix estimates as before.

5. Conclusions

- ψ-function matters:
  steep descent $\rightarrow$ difficult to correct for p/n.
- the proposed "generalized gaussweight" ψ-function is continuously differentiable and slowly descending
- re-estimating σ and β (SMTM) helps keeping the efficiency of $\hat{\beta}$ at the desired level
- ψ-functions have strong influence on inference
- sensitivity curves can give insights into what goes wrong
- The proposed method is implemented in the development version of robustbase, which can be downloaded at https://r-forge.r-project.org/projects/robustbase/

References

P. J. Huber, E. M. Ronchetti.
Robust Statistics, Second Edition.
Wiley and Sons Inc., New York, 2009.

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

Maronna, R.A. and Yohai, V.J.
Correcting MM estimates for fat data sets
Computational Statistics & Data Analysis, Elsevier, 2009.

V. J. Yohai, W. A. Stahel, R. H. Zamar.
A procedure for robust estimation and inference in linear regression.
Directions in Robust Statistics and Diagnostics (Part II).
W. A. Stahel and S. Weisberg (eds.).
The IMA Volumes in Mathematics and its Applications, 365-374, Springer New York, 1991.

Yohai, V. and Zamar, R.
Optimal locally robust M-estimates of regression
Journal of Statistical Planning and Inference 64(2): 309-323, Elsevier, 1997.

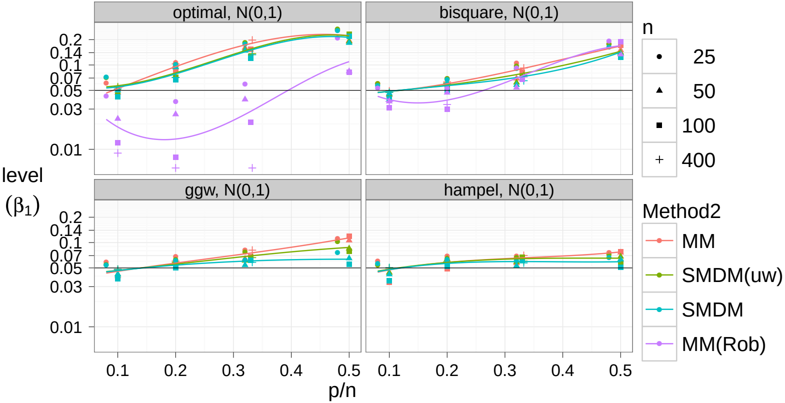
Generalized Gaussweight ψ-function

$$\psi(x, c) = \begin{cases} x & |x| \leq c \\ \exp\left(-\frac{1}{2}\frac{(|x|-c)^b}{a}\right) x & |x| > c \end{cases} \tag{7}$$

Suggested parameters:

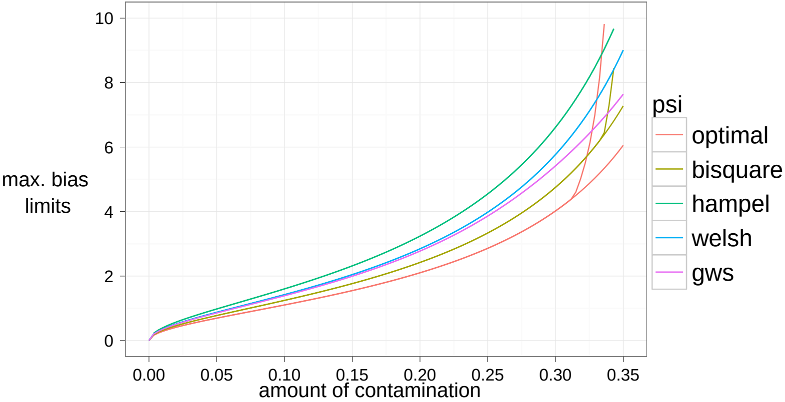
$a = 1.387$, $b = 1.5$ and $c = 1.063$ (for 95%-efficiency).

Results: Empirical nominal levels



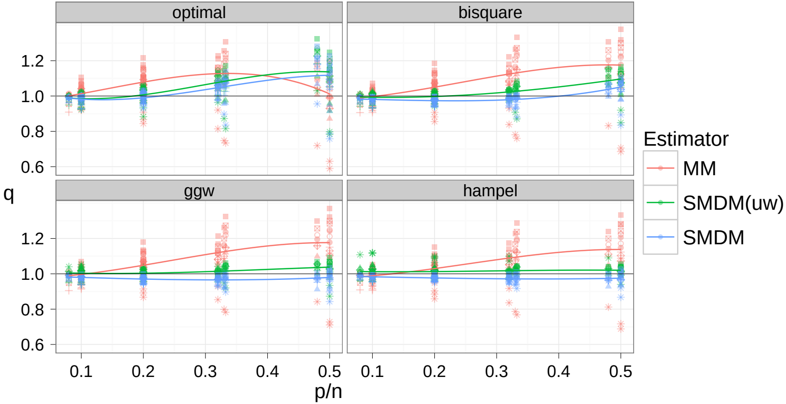
(Covariance matrix estimates as before; simulated error distribution: standard normal)

Results: Maximum asymptotic bias



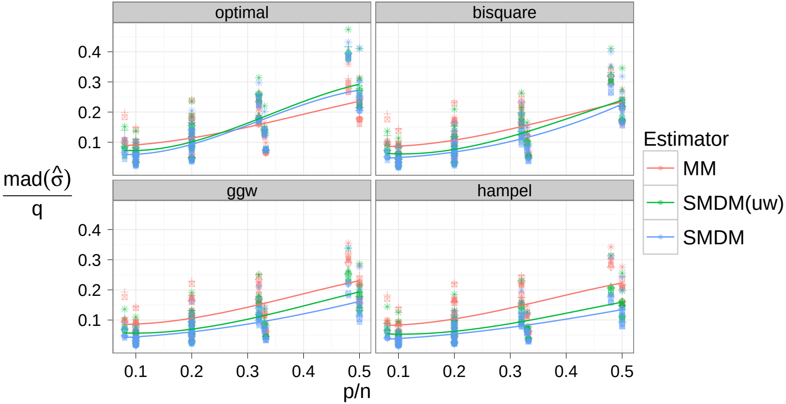
Calculations as in *Berrendero et al 2007*

Results: Scale



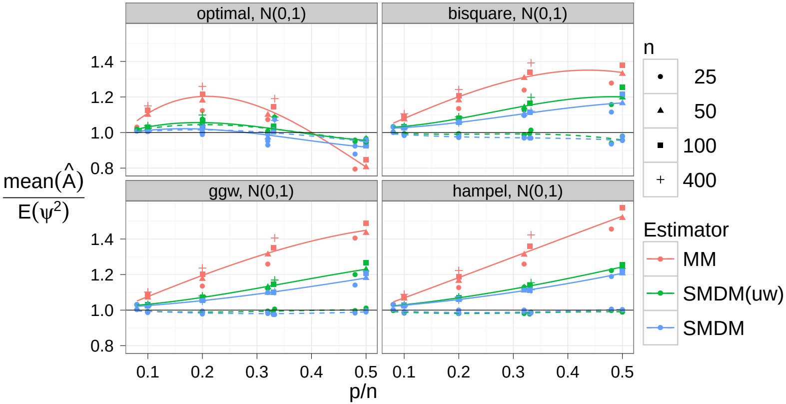
$q = \text{med}\left(\frac{\hat{\sigma}(\mathbf{1})}{\hat{\sigma}(\mathbf{X})}\right)$ shape: error distribution.

Results: Scale



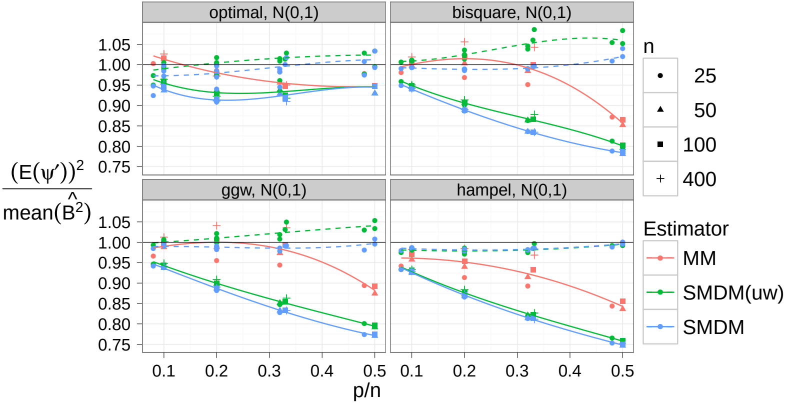
$q = \text{med}\left(\frac{\hat{\sigma}(\mathbf{1})}{\hat{\sigma}(\mathbf{X})}\right)$ shape: error distribution.

Results: Correction factor A



Solid line: corrected with $\frac{1}{1-p/n}$.
Dashed line: estimator using τ standardized residuals.

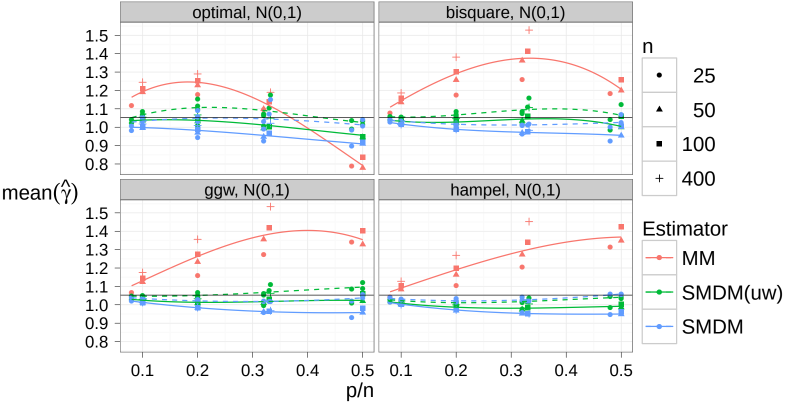
Results: Correction factor B^2



Solid line: Empirical estimator.

Dashed line: estimator using τ standarized residuals.

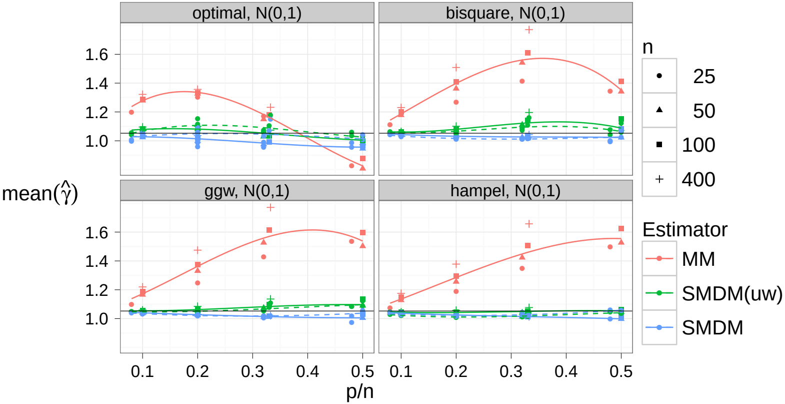
Results: Correction factor



Solid line: using $\frac{1}{1-p/n}$
Dashed line: τ standarized residuals

(Simulated error distribution: standard normal).

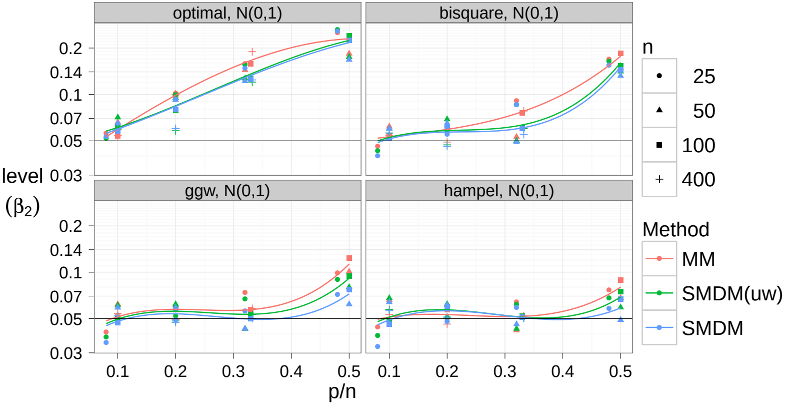
Results: Correction factor



Solid line: using $\frac{1}{1-p/n}$ and Huber's small sample correction
Dashed line: τ standarized residuals

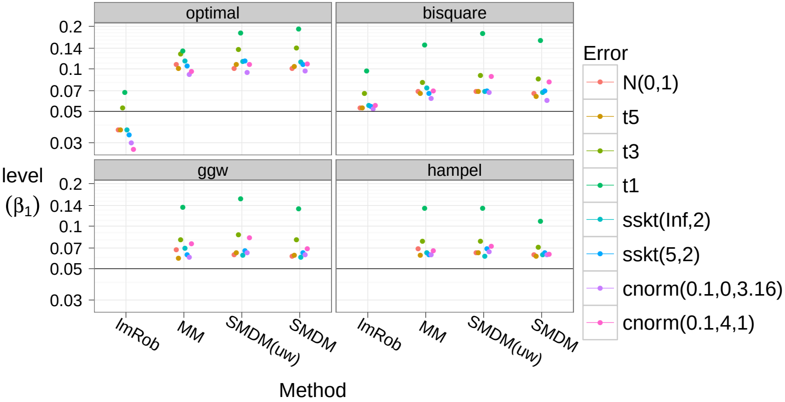
(Simulated error distribution: standard normal)

Results: Empirical nominal levels



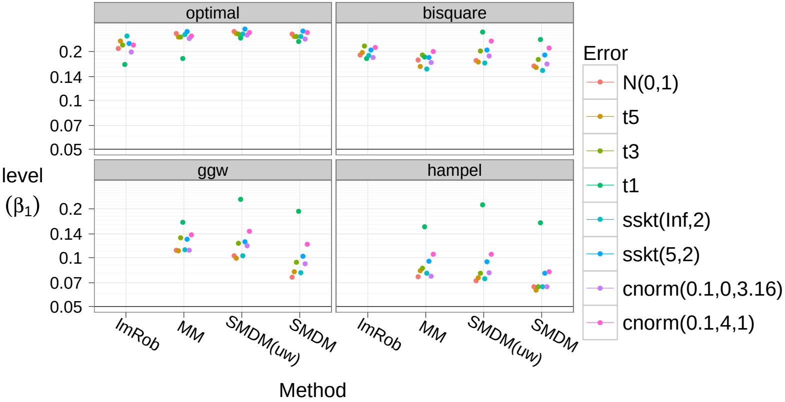
(Covariance matrix estimates as before; simulated error distribution: standard normal)

Results: Empirical nominal levels



For $n = 25$, $p = 5$, and covariance matrix estimates as before.

Results: Empirical nominal levels



For $n = 25$, $p = 12$, and covariance matrix estimates as before.