Robust methods for asymmetric responses and applications to the analysis of hospital costs

Alfio Marazzi
University of Lausanne
Institute for social and preventive medicine
Bugnon 17, CH 1005 Lausanne

alfio.marazzi@hospvd.ch

1 PCS and Casemix hospital funding

A *patient classification system* (PCS) is a tool for hospital management.

Using a given PCS and the corresponding *grouper* (i.e., a computer program) the patient population (or *casemix*) is subdivided into groups (*diagnosis related groups* or *DRGs*) homogeneous wrt:

- resource consumption,
- clinical criteria.

Main purpose:

budgeting

determine prospective payments or reimbursement rates:

- a "mean cost of stay / DRG" is computed using (national) data;
- mean costs are used to fix payments,
  e.g., a fixed DRG specific payment for each stay.

The design and development of PCS began in the late sixties (at Yale);
Today PCS are implemented in US and several European countries;
Switzerland is gradually introducing casemix funding.

There are several PCS (depending on country, version, etc.).
The typical number of groups ranges from a few hundreds to $\sim$ 1000.

Basic statistical problems

For each DRG:

- Compute a "mean cost of stay";
- Compare mean cost among hospitals (or over different periods of time);
- Explain mean cost using available covariates (e.g., insurance type).

  Note: Length of stay (LOS) is often used as a proxy of cost of stay.

Analysis is made difficult because:

- cost and length of stay distributions are asymmetric;
- distributions have varying shapes and variances;
- outliers make usual statistics unstable.

Common remedy for outliers used by practitioners:

- remove outliers outside a certain interval from mean or MED;
  bounds depend on a measure of dispersion
  e.g, standard deviation, MAD, interquartile range;
- compute mean of remaining data, i.e. a *robust mean.*

  $\longrightarrow$ Refinement of basic reimbursement principle:

  use robust mean to fix the DRG specific payment for one "typical" stay;
  "atypical" stays (outliers) are inspected and reimboursed separately.

  Note: Procedures cannot be "subjective".

Our approach:

- Use the concepts of (parametric) robust statistics;
- Remain "close" to practitioner's rule:
  a. fit a parametric model using a very robust method,
  b. reject outliers with respect to the fitted model,
  c. use a classical estimation method with remaining data.
- Interpret the mean of the rubust model as an estimate of the population mean after removal of extreme values.

2 Asymmetric models with 2 parameters

Location (τ) - scale (σ)

$$y_i = \tau + \sigma e_i, \qquad e_i \sim F_0$$

Examples

- y_i is the log of a Lognormal variable, $F_0 = \Phi$;
- y_i is the log of a Weibull variable, F_0 is a standard "log-Weibull".

Interest in

$$\mu = E(\exp(y_i)).$$

Shape (τ) - scale (σ)

$$y_i \sim F(\cdot\ ;\ \tau, \sigma)$$

Examples

- $y_i \sim$ Gamma(τ, σ);
- $y_i \sim$ Pareto(τ, σ) first and second kind.

Interest in

$$\mu = E(y_i).$$

Important differences with respect to the symmetric case:

- Scale is a main component of mean, not a nuisance parameter, e.g.,

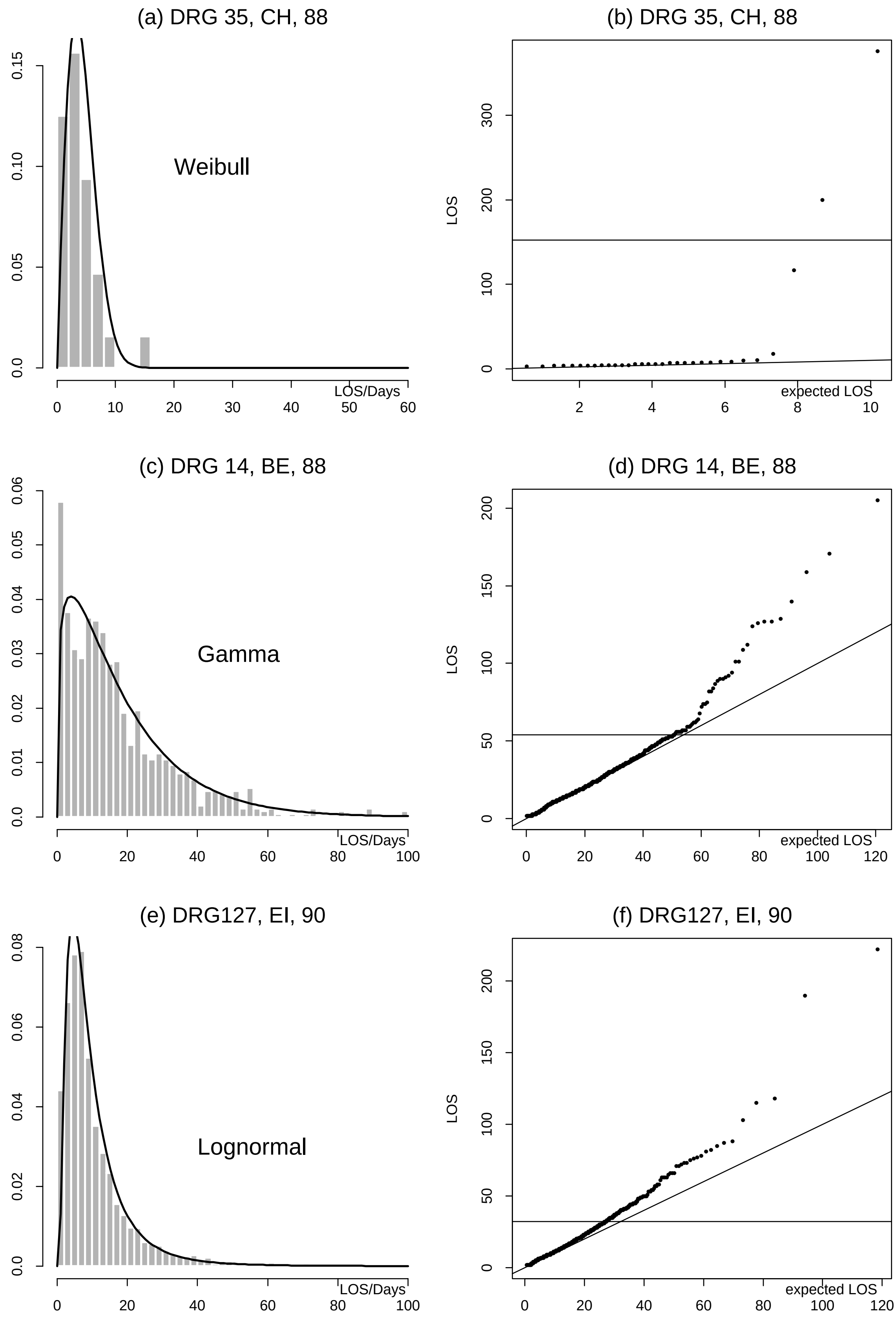
$$\text{Lognormal mean} = \exp(\tau + \sigma^2/2).$$

- Genuine shape-scale models + estimators are not shape equivariant. There is no parameter free standard F_0.
- The parameter estimates are correlated.

In practice

these models provide adequate descriptions for most DRGs;
typical exception: mixed (bimodal) distributions
$\rightarrow$ DRG requires a re-definition.

LOS distributions and fitted models



3 Robust estimates: without covariates

We use two step estimates:

Step 1: compute a high breakdown point estimate;
Step 2: compute efficient estimate maintaining the high breakdown point.

[inspired from MM-estimate; see Appendix]

Initial estimates

D-estimator i.e. MED & MAD like (Marazzi+Ruffieux, CS&DA, 1999)

Let

$$m(\text{data}) = \text{robust measure of "location" of data},$$
$$s(\text{data}) = \text{robust measure of scale of data},$$

$$m(\tau, \sigma) = \text{measure } m \text{ of location of model},$$
$$s(\tau, \sigma) = \text{measure } s \text{ of scale of model},$$

and solve

$$m(\tau, \sigma) = m(\text{data}), \quad s(\tau, \sigma) = s(\text{data}).$$

Take m and s with a high breakdown point (bdp).

Examples

m = median,	s = median absolute deviation
m = 0.4-symmetric trimmed mean,	s = 0.4-trimmed absolute deviation

- Any pair of location-scale robust estimates can be used.
- OK for location-scale and shape-scale.
- Approximations of $m(\tau, \sigma), s(\tau, \sigma)$ possible for shape-scale case.

- MED & MAD have 50% bdp but inefficient.
- MED and MAD have discontinuous influence function.

Corrected S-estimator for location-scale (Marazzi+Yohai, JSPI, 2004)

(a) Compute usual S-estimate (Rousseeuw+Yohai, 1984)

$$T^* = \arg\min_{\tau} S_k(\tau),$$

where $S_k(\tau)$ is the solution of

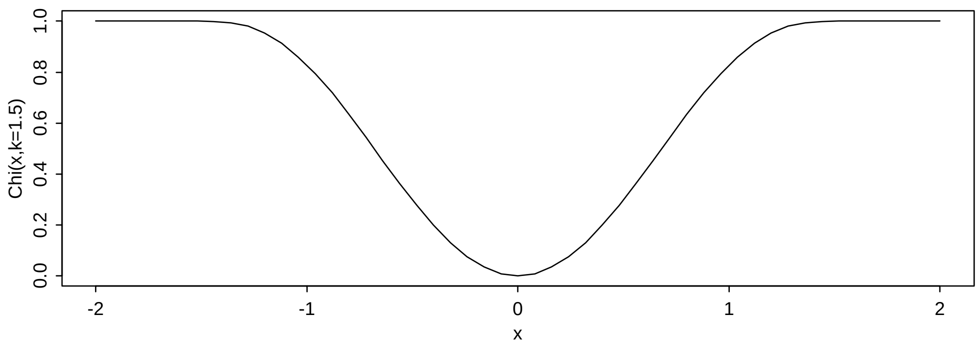
$$\frac{1}{n-1}\sum \rho_k((y_i - \tau)/S) = 0.5,$$

and let $S^* = S_k(T^*)$;

(b) Correct for bias: $T = T^* - S^* a$,

where

$$\rho_k(z) = \begin{cases} 3(z/k)^2 - 3(z/k)^4 + (z/k)^6 & \text{if } |z| \le k, \\ 1 & \text{if } |z| > k, \end{cases}$$



and

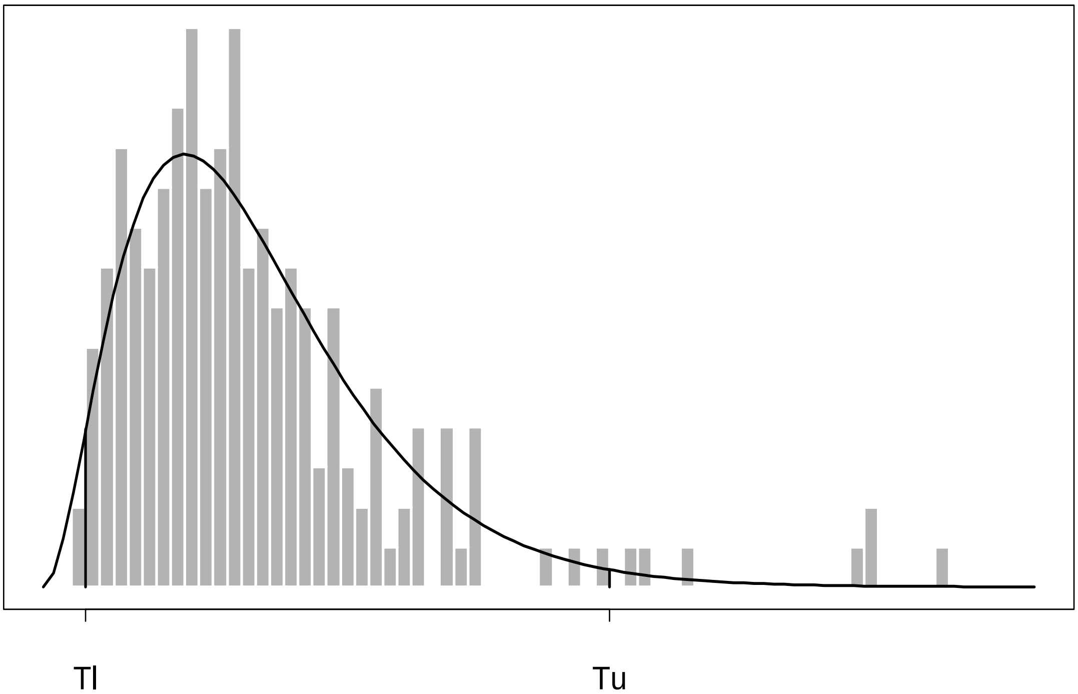
$$k: \qquad \int \rho_k(z-a) f_0(z) dz = 0.5,$$

$$a = \arg\min_t \int \rho_k(z-t) f_0(z) dz,$$

e.g., $k = 1.548, a = 0$ (Gauss),
$k = 1.718, a = -0.134$ (log-Weibull),
[a is the asymptotic value of T^* at $f_0 \Leftarrow$ Rousseeuw 1987, p. 140].

- Extension to regression OK: correct intercept.
- Most robust estimates (LMS-, LTS-, τ-) can be corrected in this way.
- Extension to shape-scale OK: but k and a depend on estimated shape (Marazzi+Barbati, Estadistica, 2003).
- Continuous influence function;
- 50% bdp but inefficient.

Truncated mean (Marazzi+Ruffieux, CS& DA, 1999)



1. Adjust a model using D or S;
2. Choose a number u, e.g., $u = 99\%$ and compute *cut off*

   $$t_u = \text{upper } u\text{-quantile of the adjusted model};$$

   Determine t_l such that the mean of the truncated model equals the mean of the complete model;
3. *Truncated mean* (TM): mean of data between t_l and t_u.

- The bdp of TM is 50%.
- Good efficiency (e.g., 90%; depending on u and on model).
- Low bias and variance (w.r.t. M-estimators) for severe contamination.
- The model is used just to compute t_l and t_u;
  $\Rightarrow$ TM does not strongly depend on the model.
- Can be extended to regression (truncated least squares).

Truncated maximum likelihood estimator
(Marazzi+Yohai, JSPI, 2004)

Assume (location-scale model)

$$y_i = \tau + \sigma e_i, \qquad e_i \sim F_0, \qquad \text{i.e.} \quad y_i \sim F_{\tau,\sigma}.$$

Let $\rho(z) = \ln f_0(z)$ be the log-likelihood function.

The *TML-estimator* is computed in three steps:

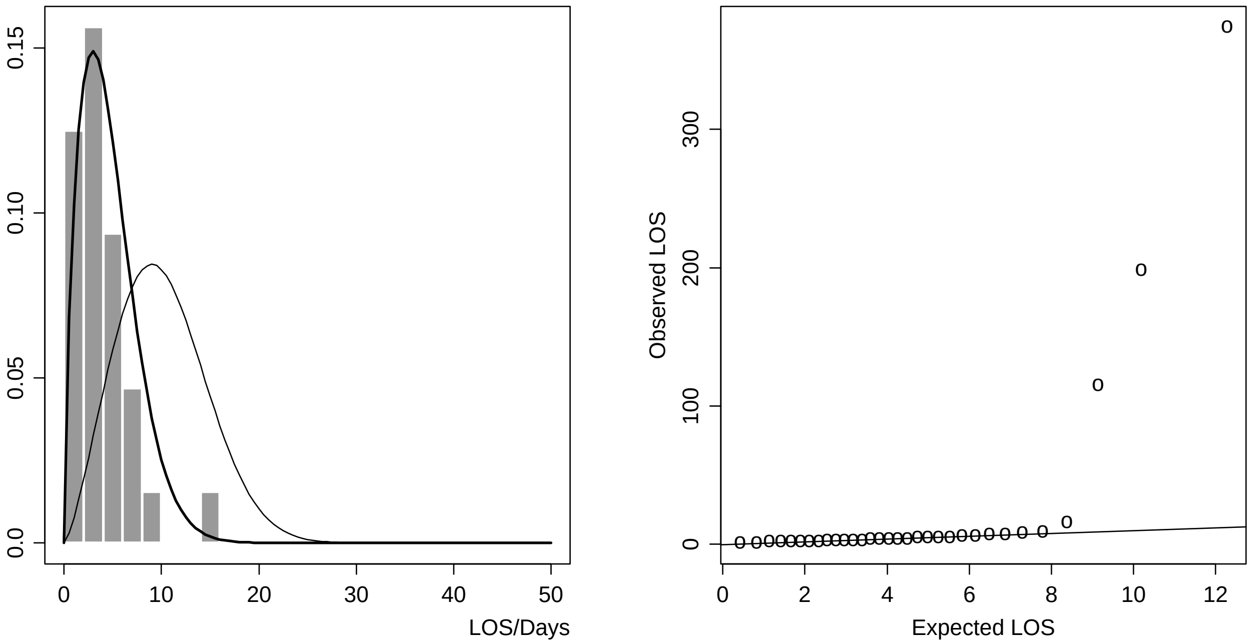
1. Initial $(T^{(0)}, S^{(0)})$, e.g., D or S;
2. Compute
$$r_i = (y_i - T^{(0)})/S^{(0)}$$
$$\rho_i = \rho(r_i)$$
and remove observations with small likelihood, i.e.
$$\rho_i < \text{a given cut-off}$$
3. Compute ML on retained observations.

- $T^{(0)}$ is automatically Fisher consistent (asymptotically).
  $S^{(0)}$ must be corrected for (asymptotic) consistency.
- Cut-off depends on desired efficiency level.
- The cut-off can be made adaptive $\rightarrow$ *ATML-estimator*:
  - adapted to discrepancy between the residuals cdf and F_0
  - if discrepancy $\rightarrow 0$, then cut-off $\rightarrow \infty$.
- The breakdown point of TML- and ATML-estimators is not smaller than the breakdown point of the initial estimator: it can attain 50%.
- Influence functions are available and can be used to estimate the variance of TML- and ATML-estimators.
- At the model, the influence function of the ATML-estimator coincides with the influence function of the ML-estimator.
  Therefore, the ATML-estimator is fully efficient at the model.
- Quantitative results including TM, TML, ATML suggest that ATML starting with S is the best.
- Extension to regression OK.

Example: modeling a length of stay distribution

Frequency distribution (Freq.) of LOS in days of patients hospitalized at CHUV during 1986 for certain "disorders of the nervous system".

LOS	1	2	3	4	5	6	7	8	9	16	115	198	374
Freq.	2	6	5	5	4	2	2	1	1	1	1	1	1



Thin line: Weibull distribution estimated by maximum likelihood
Thick line: Weibull distribution estimated by MED & MAD

mean=25.5 days, ML=18.77 days, TM=4.00 days, TML=3.96 days

4 Tests for comparing robust means

(Marazzi, CS&DA, 2002; Marazzi+Barbati, Estadistica, 2003)

- Compare DRG specific "cost means" among different hospitals;
  Compare DRG specific "cost means" over different periods of time.

- Cost distributions for same DRG
  - are asymmetric,
  - contain outliers,
  - may have different shapes,
  - may have different scales.

- Commonly used procedures for testing <u>means</u>
  of asymmetric distributions are often inappropriate:
  - variants of t-test: not robust
  - rank tests: not appropriate when distributions differ in shape
  - transformations to symmetry + "usual" robust estimates:
    - · not always possible,
    - · null shift in transformed scale $\not\equiv$ equality of means,
    - · scale is not a nuisance parameter.

Our approach:

Use existing robust estimates of mean, e.g., TM, TML, ATML.

Compare *robust population means*

Use the bootstrap to compute the finite sample null distribution of a studentized test statistic.

One sample problem (generalization to multi-sample straightforward)

- Data $y = (y_1, \ldots, y_n)$ i.i.d $\sim F$, let $\mu(F) = \int y dF(y)$
  Parametric model F_θ, $\theta = (\tau, \sigma)$, $\mu(F_\theta)$
  Robust estimator (e.g. TML) $\hat{\theta}(F_{n,y})$, $\hat{\mu}(F_{n,y}) = \mu(F_{\hat{\theta}(F_{n,y})})$
  Define parameter of interest = population value of its estimate

  $$\hat{\mu}(F) \qquad \text{\textit{“robust population mean”}}$$

  interpreted as the population mean after “removal” of extreme values.
- Test of $\mathcal{H}_0 : \quad \hat{\mu}(F) = \mu_0$.
- Studentized test statistic

  $$t(F_{n,y}) = \frac{h(\hat{\mu}(F_{n,y})) - h(\mu_0)}{\sqrt{\text{Estim. variance of } h(\hat{\mu}(F_{n,y}))}},$$

  h: variance stabilizing transformation, often $\ln(\cdot)$.
- Null model: estimate $\hat{F}_0$ of F satisfying the constraint $\hat{\mu}(\hat{F}_0) = \mu_0$.
- Simulation

  $$\hat{F}_0 \quad \longrightarrow \quad y^*, \quad t(F_{n,y^*}) \quad \text{many times;}$$

  P-value estimate

  $$P(t(F_{n,y^*}) > t_{obs} \quad | \quad \hat{F}_0).$$

- Various null models are possible:
  - nonparametric ($\rightarrow$ exponential tilting),
  - parametric ($\rightarrow$ constrained robust estimates),
  - semiparametric.

  The most simple null model for shape/scale problems assumes that: the possible distributions of Y are just rescaled versions of one another; then:
  - Let $\hat{\mu} = \hat{\mu}(F_{n,y})$
  - set $\hat{y} = (\mu_0/\hat{\mu})y$,
  - set $\hat{F}_0 = F_{n,\hat{y}}$, then $\hat{\mu}(\hat{F}_0) = \mu_0$.

- Simulation of a robust estimator from (a modified version) of the empirical distribution is non-robust: an exceeding frequency of outliers in many simulated samples may affect the tails of the null distribution. Remedy: Salibian-Barrera+Zamar, AS, 2002; Marazzi+Barbati, 2002.

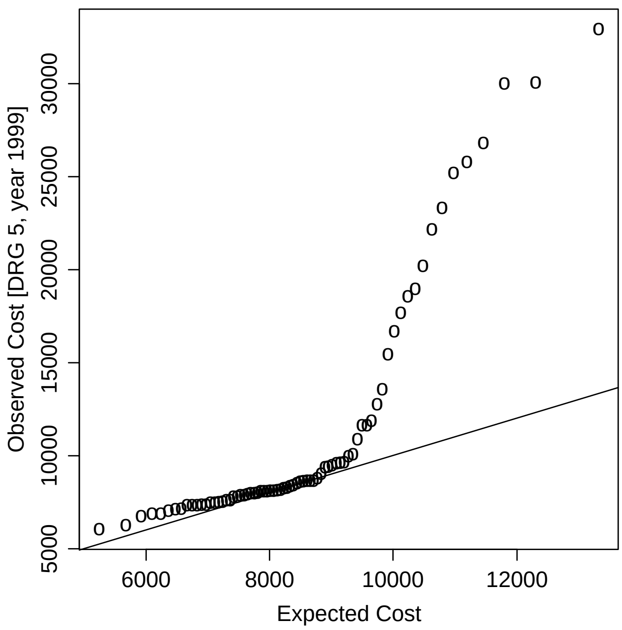
Example: Cardiovascular surgery, CHUV, two-sample test

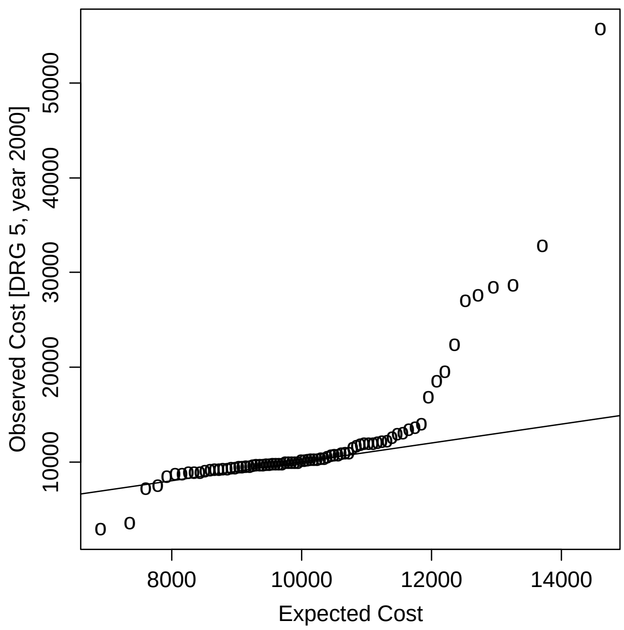
- 1999, 74 stays
  mean cost = 11'211 CHF
  TM = 8'273 CHF (Lognormal model)

- 2000, 77 stays
  mean cost = 12'303 CHF
  TM = 10'163 CHF (Lognormal model)

Test	Full data set	Outliers removed
Pooled t	0.33	0.00
Cressie-Whitford[1]	0.15	0.00
Guo-Luh[2]	0.20	0.00
TM/bootstrap	0.001	0.00

[1] based on skewness adjusted statistic
[2] based on transf. for skewness and symmetric 5%-trimmed mean





Example: Cardiovascular surgery, CHUV, two-sample test, emergencies removed

- 1999, 65 stays

  mean cost = 9'269 CHF

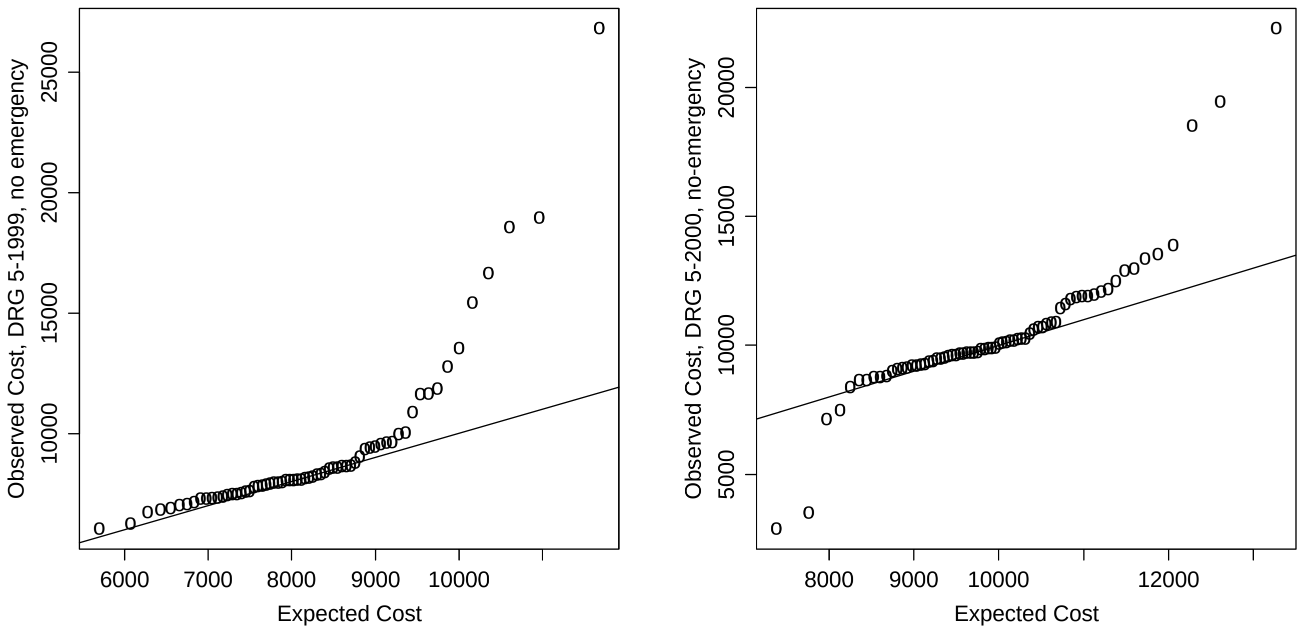
  TM = 8'087 CHF (Lognormal model)

- 2000, 70 stays

  mean cost = 10'441 CHF

  TM = 10'058 CHF (Lognormal model)

Test	Emergencies removed
Pooled t	0.03
Cressie-Whitford[1]	0.03
Guo-Luh[2]	0.01
TM/bootstrap	0.00



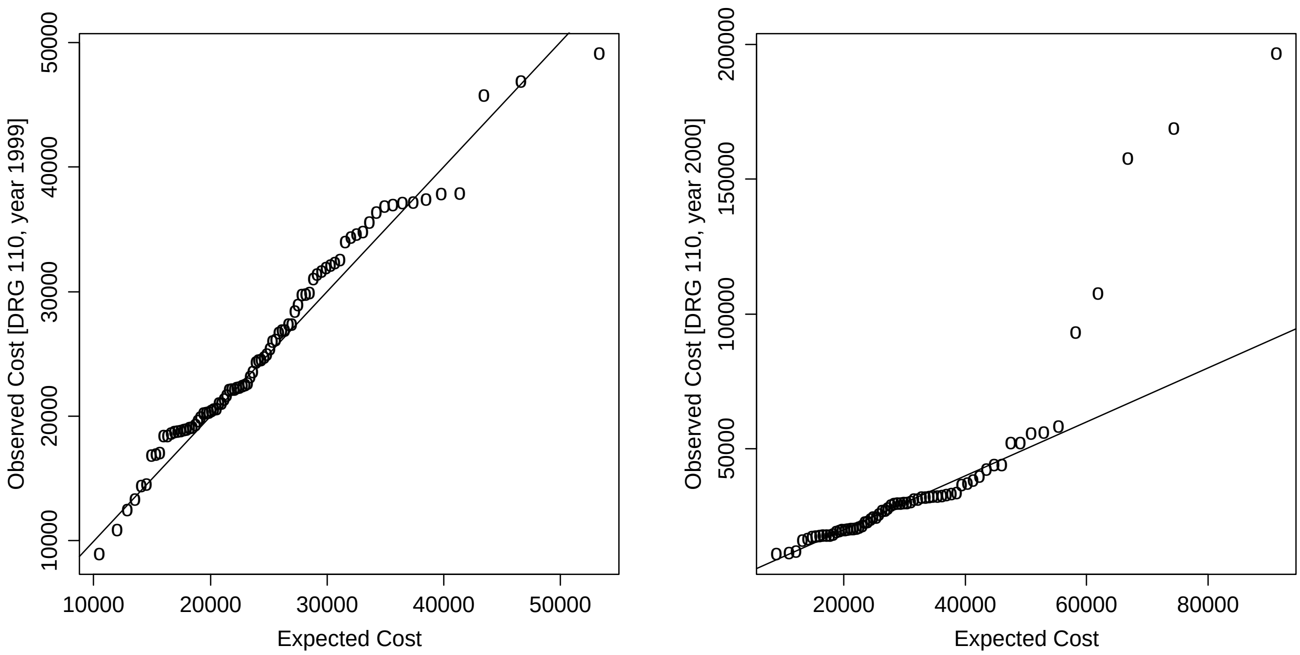
Example: Heart surgery, CHUV, two-sample test

- 1999, 83 stays
  mean cost = 25'508 CHF
  TM = 25'694 CHF (Lognormal model)

- 2000, 66 stays
  mean cost = 37'115 CHF
  TM = 27'706 CHF (Lognormal model)

Test	Full data set	Outliers removed
Pooled t	0.0030	0.14
Cressie-Whitford[1]	0.0003	0.07
Guo-Luh[2]	0.0020	0.07
TM/bootstrap	0.0800	0.09

[1] based on skewness adjusted statistic
[2] based on transf. for skewness and symmetric 5%-trimmed mean



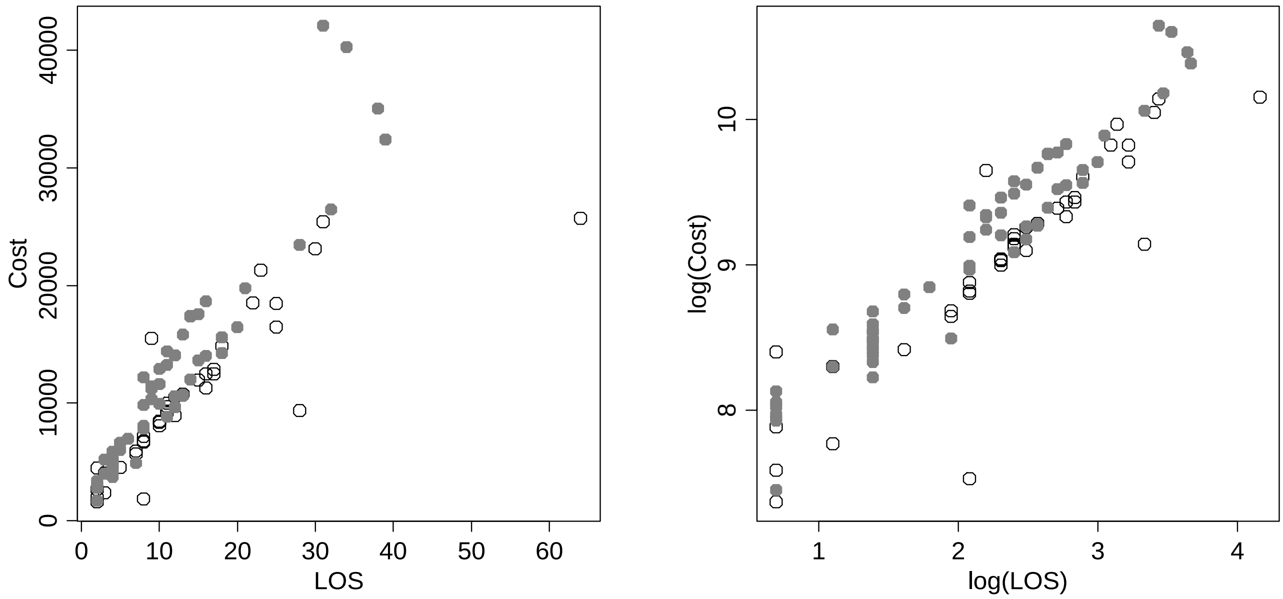
5 Truncated maximum likelihood for regression

Motivation

Analysis of hospital cost of stay:
y = CHF; x = length of stay + other covariates

Goal: estimate $E(y|\text{covariate values})$

Sample of 100 stays at CHUV, 1999, for "medical back problems"



bullet = emergency admission
circle = planned admission

Location-scale model for regression

$$y_i = \mathbf{x}_i^{\mathrm{T}}\theta + \sigma e_i, \quad e_i \sim F_0, \quad \text{i.e.} \quad y_i \sim F_{\tau,\sigma}, \quad \tau = \mathbf{x}_i^{\mathrm{T}}\theta$$

Let $\rho(z) = -\ln f_0(z)$,

$$s_1(z) = [\partial \ln f_{\tau,\sigma}/\partial\tau]_{\tau=0,\sigma=1}, \quad s_2(z) = [\partial \ln f_{\tau,\sigma}/\partial\sigma]_{\tau=0,\sigma=1} + 1.$$

The truncated maximum likelihood estimate of regression

1. Initial $(\mathbf{T}^{(0)}, S^{(0)})$, e.g., S;
2. $r_i = (y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{T}^{(0)})/S^{(0)}$ and $\rho_i = \rho(r_i)$;
   cut-off value t_n on the log-likelihood scale (see below);

$$
\begin{aligned}
w_i &= 0 \qquad \text{if likelihood of } y_i \text{ is small: } -\rho_i < t_n, \\
w_i &= 1 \qquad \text{if likelihood of } y_i \text{ is large: } -\rho_i > t_n.
\end{aligned}
$$

   $\tilde{n} = \sum w_i$ number of retained observations.
3. Compute ML on retained observations: solve

$$
\begin{aligned}
\frac{1}{\tilde{n}} \sum_1^n w_i s_1((y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{T})/S) &= 0, \\
\frac{1}{\tilde{n}} \sum_1^n w_i s_2((y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{T})/S) &= 1.
\end{aligned}
$$

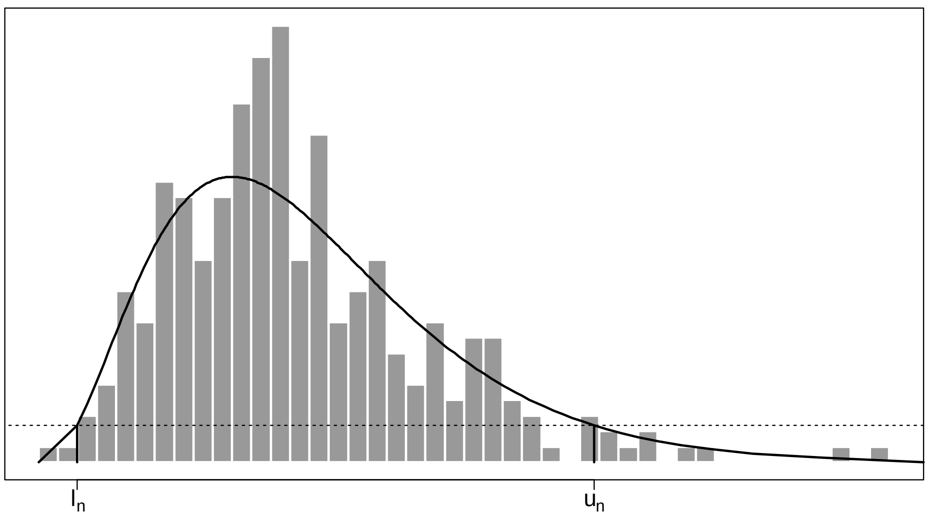
4. Correct the scale estimate for asymptotic consistency: $S := S/b$, where

$$
\frac{1}{F_0(u_n) - F_0(l_n)} \int_{u_n}^{l_n} s_2(z/b) f_0(z) dz = 1,
$$

   and $\rho(u_n) = t_n, \quad \rho(l_n) = t_n$

- T is automatically Fisher consistent:

$$
\int_{\rho(z)<t_n} s_1(z) f_0(z) dz = f_0(u_n) - f_0(l_n) = 0.
$$



Cut-off value

Let F_0^+ be the model cdf for $\rho(e_i)$ (e.g., model cdf of e_i^2 in Gaussian case)

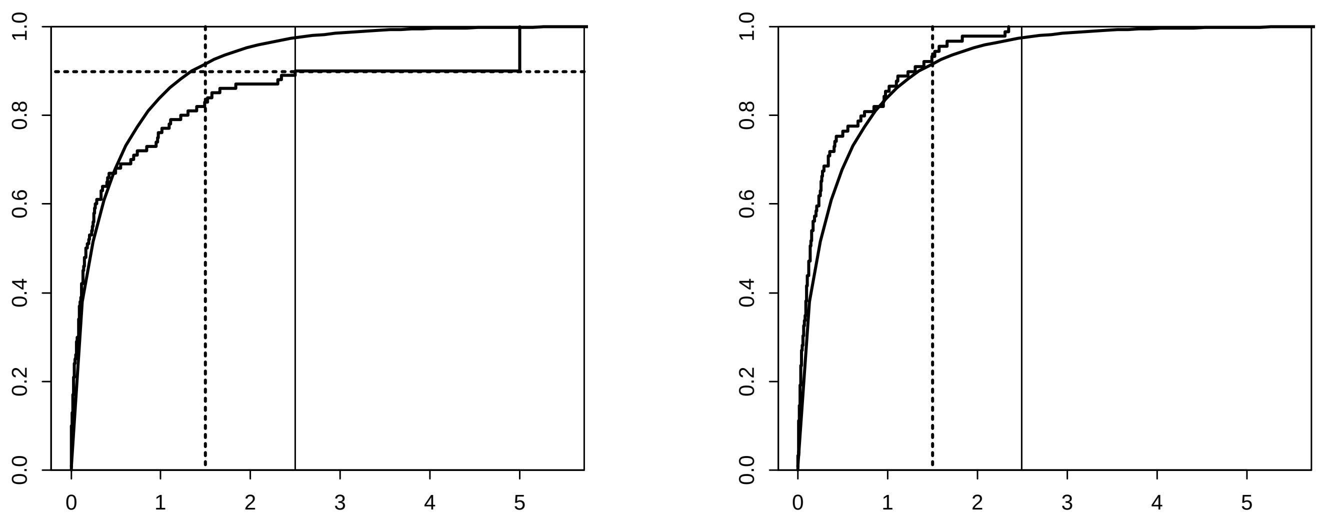
Non adaptive cut-off

e.g., $t_n = F_0^{+^{-1}}(0.99) \Longrightarrow$ Truncated ML or *TML-estimator*

Adaptive cut-off $\qquad \Longrightarrow$ Adaptively TML or *ATML-estimator*

Let F_n^+ be the empirical cdf of $\rho(r_i)$ and

$$\tilde{F}_n^+(z) = \begin{cases} F_n^+(z) & \text{if } z \le t, \\ 1 & \text{otherwise.} \end{cases}$$



Look for the largest $t > 0$ such that

$$\tilde{F}_n^+(z) \ge F_0^+(z) \quad \text{for all } z \ge \eta, \qquad \text{e.g., } \eta{=}F_0^{+^{-1}}(0.99).$$

It turns out that $t_n = F_n^{+^{-1}}(\alpha_n)$, where $\alpha_n = \min[\inf_{z \ge \eta} F_n^+(z)/F_0^+(z), 1]$
α_n is the "proportion of retained observations";
it can be computed by inspecting $F_n^+(z)/F_0^+(z)$ for $z = \rho_1, \ldots, \rho_n$.

Results

• The breakdown point of TML- and ATML-estimators is not smaller than the breakdown point of the initial estimator: it can attain 50%.

• Influence functions are available and can be used to estimate the variance of TML- and ATML-estimators.

• At the model, the influence function of the ATML-estimator coincides with the influence function of the ML-estimator. Therefore, the ATML-estimator is fully efficient at the model.

Example: regression of log(cost) on 6 covariates

- Sample of 100 stays at CHUV, 1999, for "medical back problems".
- Tentative model

$$y = \theta_0 + \theta_1 x_1 + \ldots + \theta_6 x_6 + e, \qquad e \sim \text{log-Weibull},$$

$$\begin{aligned} y &= \log(\text{Cost}), \\ x_1 &= \log(\text{LOS}), \\ x_2 &= \text{admission type } (0 = \text{planned},\ 1 = \text{emergency}), \\ x_3 &= \text{insurance type } (0 = \text{regular},\ 1 = \text{private}), \\ x_4 &= \text{age (years)}, \\ x_5 &= \text{sex } (0 = \text{female},\ 1 = \text{male}), \\ x_6 &= \text{discharge destination } (1 = \text{home},\ 0 = \text{another health institution}). \end{aligned}$$

Full model

	TML			ML			LS		
j	$\hat\theta_j$	st.err.	t	$\hat\theta_j$	st.err.	t	$\hat\theta_j$	st.err.	t
0	7.10	0.086	82.59	7.19	0.137	52.30	7.25	0.162	44.77
1	0.89	0.017	53.00	0.81	0.026	30.93	0.82	0.031	26.75
2	0.31	0.030	10.30	0.17	0.047	3.58	0.24	0.055	4.28
3	-0.06	0.049	-1.17	0.13	0.074	1.81	0.08	0.087	0.95
4	-0.00	0.001	-1.21	0.00	0.001	1.04	-0.00	0.001	-0.85
5	0.03	0.030	1.17	0.18	0.047	3.79	0.07	0.055	1.21
6	-0.07	0.040	-1.69	-0.06	0.065	-0.96	-0.11	0.076	-1.50
	scale estimate: 0.136			scale estimate: 0.208			scale estimate: 0.246		

– Covariates 3,4,5, and 6 are non-significant

- Remove the non-significant effects of variables x_3, x_4, x_5, and x_6:

Reduced model

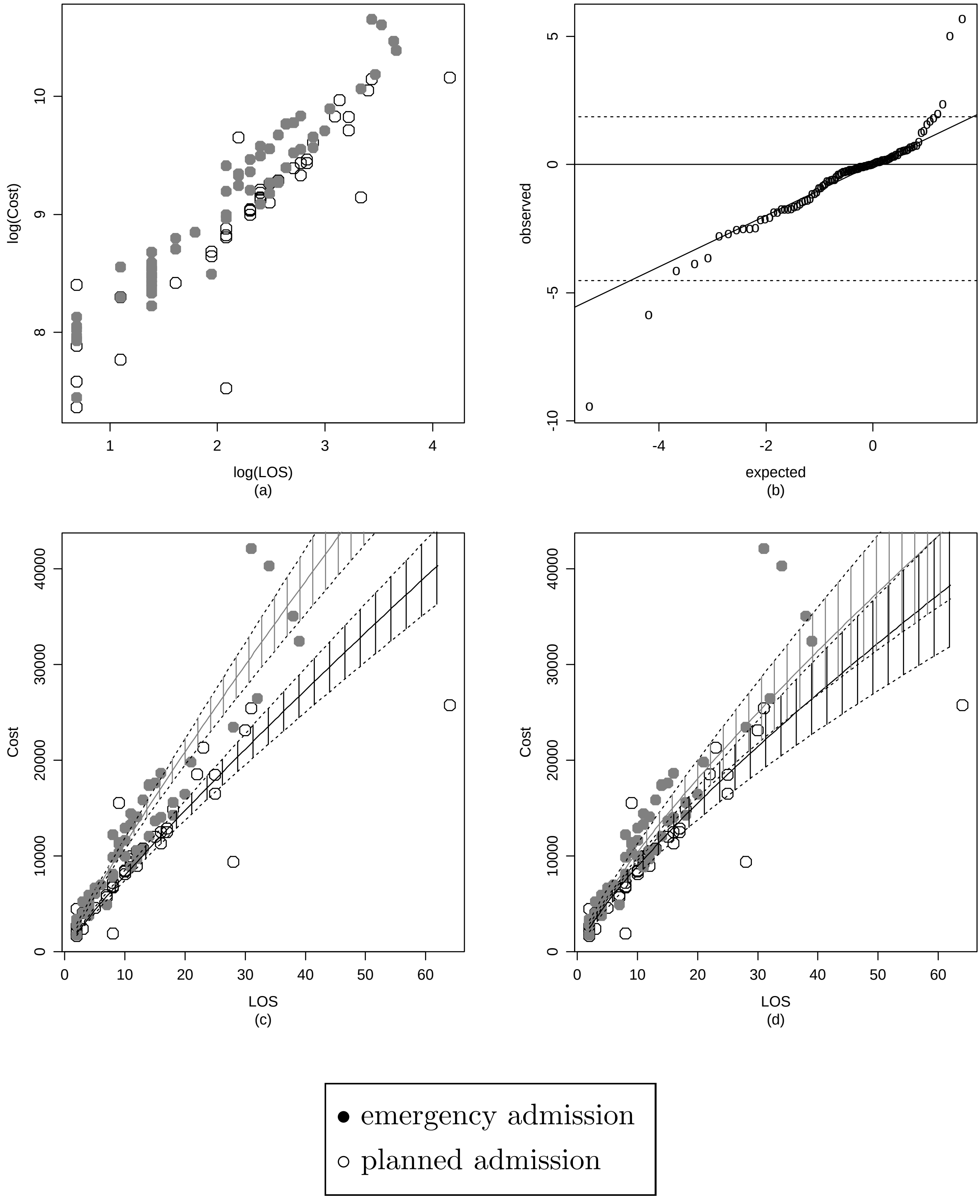
	TML			ML			LS		
j	$\hat{\theta}_j$	st.err.	t	$\hat{\theta}_j$	st.err.	t	$\hat{\theta}_j$	st.err.	t
0	7.00	0.049	144.02	7.35	0.077	94.91	7.12	0.081	88.32
1	0.89	0.017	51.49	0.80	0.029	27.72	0.82	0.030	27.28
2	0.35	0.029	11.96	0.15	0.049	3.14	0.26	0.051	5.17
	scale estimate: 0.139			scale estimate: 0.236			scale estimate: 0.247		

  - The classical estimate of scale (0.236) is markedly larger than the robust estimate (0.139).
  - Removing outliers, the classical estimate becomes 0.134 which is very close to the robust estimate 0.139.
  - The robust coefficient of "admission type" (0.35) is markedly larger than the classical estimate (0.15).

- Use the reduced model to estimate

$$E(\exp(y) \mid x_1, x_2) = \exp(\theta_0 + \theta_1 x_1 + \theta_2 x_2)\Gamma(1 + \sigma),$$

  i.e., the expected cost for given x_2 and x_2.
  - Robust conditional cost estimates of emergency cases are markedly higher that the classical ones, especially for large LOS values.
  - The robust confidence intervals are shorter than the classical confidence intervals.



log(Cost)
log(LOS)
(a)
observed
expected
(b)
Cost
LOS
(c)
Cost
LOS
(d)
emergency admission
planned admission

6 Robust response transformations in regression

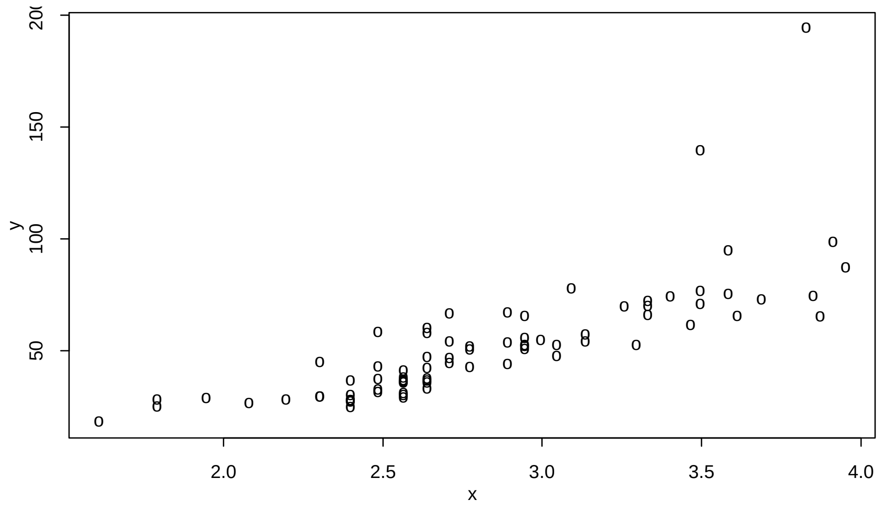
Joint work with Victor Yohai

Motivation

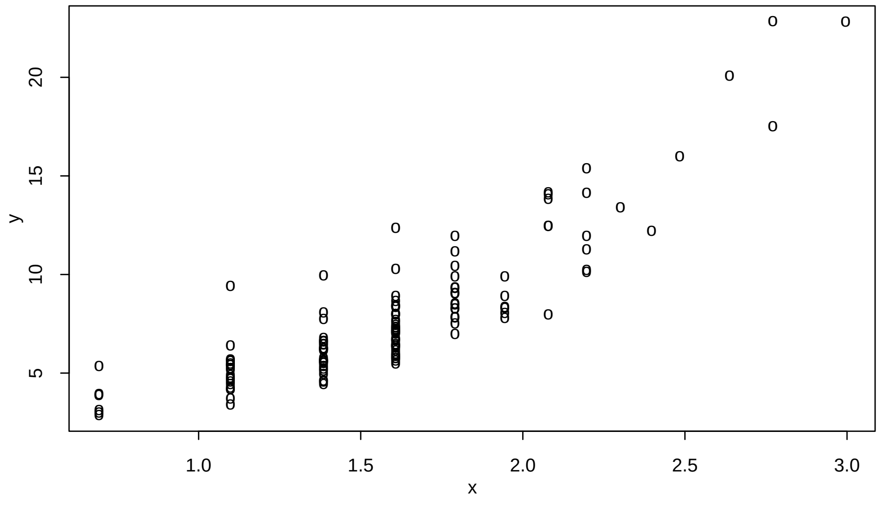
Analysis of hospital cost of stay:
y = CHF/1000; x = log(length of stay) + other covariates

Goal: estimate $E(y|\text{covariate values})$
with increasing model flexibility wrt location-scale models

78 stays, valve cardiac surgery, with major cc, CHUV Lausanne 2000



135 stays, cholecystectomy without cc, CHUV Lausanne 2000



We observe:
- asymmetrically distributed responses
- heteroschedasticity
- non linearity
- outliers

The Box-Cox regression model

Consider the family of Box-Cox transformations of the response:

$$y^{(\lambda)} = \begin{cases} (y^\lambda - 1)/\lambda & \text{if } \lambda \neq 0, \\ \ln(y) & \text{if } \lambda = 0. \end{cases}$$

Assume that the observations $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ satisfy

$$y_i^{(\lambda_0)} = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + h(\mathbf{x}_i) u_i, \quad \mathbf{x} \in \mathbb{R}^p, \quad \boldsymbol{\beta}_0 \in \mathbb{R}^p,$$

where λ_0 and $\boldsymbol{\beta}_0$ are unknown

$$u_i \text{ i.i.d. } \sim F,$$
$$u_i \text{ independent of } \mathbf{x}_i.$$

Classical setup: F is the normal distribution,
homoscedasticity, $h(\mathbf{x}) = \text{constant}$.

Under these assumptions, the classical estimates of λ_0 and $\boldsymbol{\beta}_0$ are based on the ML (maximum likelihood) criterion.

Unfortunately:

(near) normality & homoscedasticity are hard to attain simultaneously with a single transformation and

– the ML estimate is not consistent under non-normal errors,
– the ML estimate is not consistent under heteroscedastic errors.

In addition, the ML estimate is non-robust.

We propose robust approaches that do not require neither normality,
nor homoscedasticity.

For simplicity, consider the power transformation model:

$$y_i^{\lambda_0} = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + u_i, \quad i = 1, \ldots, n.$$

For any λ define

- $\tilde{\boldsymbol{\beta}}_n(\lambda)$: the S-estimate when $z_i = y_i^{\lambda}$,
- $\hat{\boldsymbol{\beta}}_n(\lambda)$: the MM-estimate when $z_i = y_i^{\lambda}$,
- $\tilde{s}_n(\lambda) = \tilde{\sigma}_n(\tilde{\boldsymbol{\beta}}_n(\lambda))$ the associated scale estimate,

as well as the corresponding asymptotic values

- $\tilde{\boldsymbol{\beta}}(\lambda)$,
- $\tilde{\boldsymbol{\beta}}(\lambda)$,
- $\tilde{s}(\lambda)$.

If the error distribution is symmetric and unimodal,

$$\hat{\boldsymbol{\beta}}(\lambda_0) = \hat{\boldsymbol{\beta}}_0,$$

i.e., the MM-estimate $\hat{\boldsymbol{\beta}}_n(\lambda)$ is Fisher consistent when $\lambda = \lambda_0$.

Robust transformations based on residual autocorrelation

The simple regression case

$$y_i^{\lambda_0} = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + u_i, \quad i = 1, \ldots, n,$$

$$\mathbf{x}_i^{\mathrm{T}} = (1, x_i), \quad \boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02})^{\mathrm{T}}.$$

For a given λ let

- $\hat{\boldsymbol{\beta}}_n(\lambda)$ be the MM-estimates of $\boldsymbol{\beta}_0$ based on (x_i, y_i^{λ}),
- $\tilde{s}_n(\lambda)$ be the associated scale estimate,
- $r_i(\lambda) = \left[y_i^{\lambda} - \mathbf{x}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_n(\lambda) \right] \Big/ \tilde{s}_n(\lambda)$.

• Observe that,
 if $\lambda = \lambda_0$, the residual conditional mean does not depend on x.

• Proposal (a), Marazzi and Yohai (ICORS 2003 + Draft 2004)

– a measure of functional dependency between the residuals and the x_i-s, more precisely, a *robust autocorrelation* measure of the residuals sorted according to the x_i-s:

$$\gamma_n(\lambda) = \frac{1}{n} \sum_{i=1}^{n-1} \psi\left[r_{j_i}(\lambda) \right] \psi\left[r_{j_{i+1}}(\lambda) \right],$$

 where $j_1, \ldots, j_n$ such that $x_{j_1} < \ldots < x_{j_n}$,
 and ψ is Huber's function with tuning constant 1.34 (e.g.);

– the *RAC-estimator* of λ_0 :

$$\hat{\lambda}_n = \operatorname{argmin} \gamma_n(\lambda).$$

Main result

◦ Under not too restrictive assumptions satisfied by MM-estimates but <u>no homoscedasticity assumption</u>

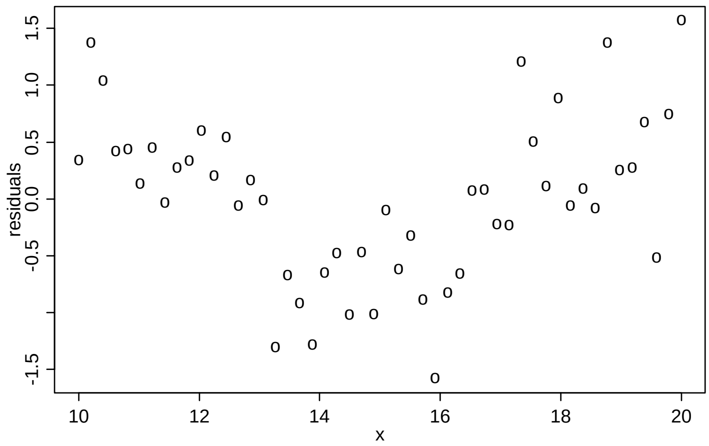
$$\hat{\lambda}_n \to \lambda_0, \qquad \text{in probability.}$$

Remark

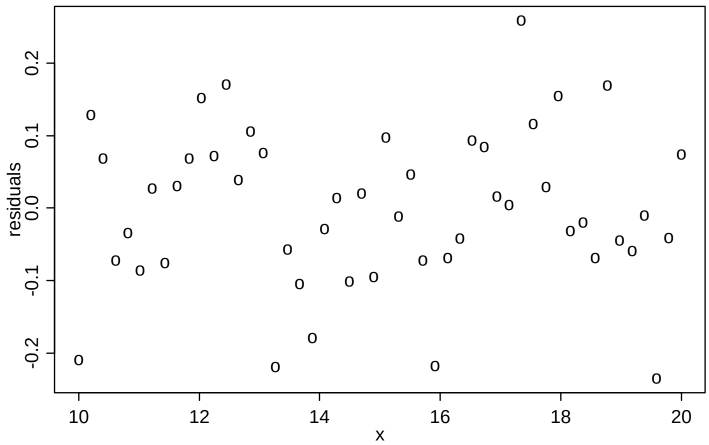
For the case with tied x-values, it is possible to modify the definition of $\gamma_n(\lambda)$.

Illustration with $\lambda_0 = 0.5$

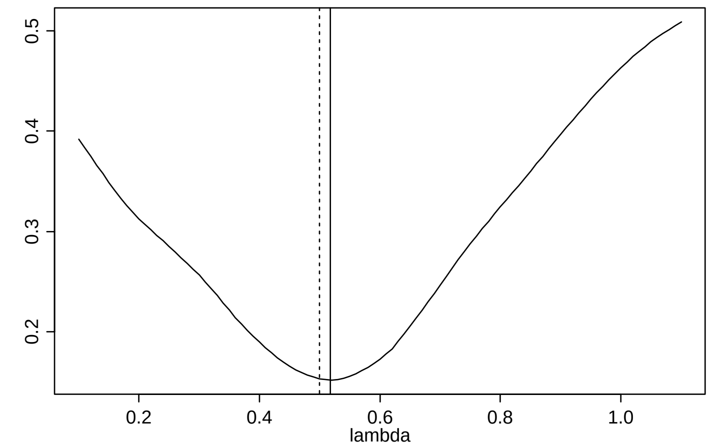
If $\lambda = 1.0$,
$\Rightarrow \gamma_n(\lambda)$ is large



If $\lambda = 0.5$,
$\Rightarrow \gamma_n(\lambda)$ is small



Function $\gamma_n(\lambda)$



The multiple regression case

$$y_i^{\lambda_0} = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + u_i, \quad i = 1, \ldots, n.$$

Proposal (b)

Use Proposal (a) with residuals sorted according to the fitted values $\mathbf{x}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_n(\lambda)$. Simulation results show that the performace is not satisfactory.

Proposal (c)

Suppose that μ is any value of the transformation parameter.

Let $j_1, \ldots, j_n$ such that $\mathbf{x}_{j_1}^{\mathrm{T}} \hat{\boldsymbol{\beta}}_n(\mu) < \ldots < \mathbf{x}_{j_n}^{\mathrm{T}} \hat{\boldsymbol{\beta}}_n(\mu)$.

Note that $j_1, \ldots, j_n$ depend on μ.

Let

$$\gamma_n(\lambda, \mu) = \frac{1}{n} \sum_{i=1}^{n-1} \psi\left[r_{j_i}(\lambda)\right] \psi\left[r_{j_{i+1}}(\lambda)\right],$$

and define the *RAC-estimate* by

$$\hat{\lambda}_n = \arg\min_{\lambda} \sup_{\mu} \gamma_n(\lambda, \mu).$$

This proposal provides satisfactory simulation results
but a consistency proof is not available.

The M-conditional expectation

Goal: predict y^λ – for a fixed λ – using a function $M(\mathbf{x})$

- Classical measure of the degree of approximation: *mean square error*

$$S^2(M,\lambda) = E\left(y^\lambda - M(\mathbf{x})\right)^2$$

  minimized when $M(\mathbf{x}) = E(y^\lambda|\mathbf{x})$: *conditional expectation* of y^λ given $\mathbf{x}$.

- Replace $(\cdot)^2$ by $\hat{\rho}(\cdot)$, the ρ-function for the MM-estimates.

  Define the *M-mean error* (MME)

$$S_{\hat{\rho}}(M,\lambda) = E\left[\hat{\rho}\left(\frac{y^\lambda - M(\mathbf{x})}{\tilde{s}(\lambda)}\right)\right],$$

  where $\tilde{s}(\lambda)$ is the scale associated to $\hat{\beta}(\lambda)$
  [required for the sake of equivariance under scale transformations].

  Define the *M-conditional expectation*

$$M_0(\mathbf{x},\lambda) = \arg\min_a E\left[\hat{\rho}\left(\frac{y^\lambda - a}{\tilde{s}(\lambda)}\right)\Big|\mathbf{x}\right]$$

  For any $M(\mathbf{x})$: $\quad S_{\hat{\rho}}(M_0,\lambda) \le S_{\hat{\rho}}(M,\lambda)$.

Observe that:

∘ For given $\mathbf{x}$, $M_0(\mathbf{x},\lambda)$ is a M-estimate of location.

∘ Since $y_i^{\lambda_0} = (\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + u_i) \Longrightarrow y_i = (\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + u_i)^{1/\lambda_0}$

$$M_0(\mathbf{x}_i,\lambda) = \arg\min_a E\left[\hat{\rho}\left(\frac{(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + u_i)^{\lambda/\lambda_0} - a}{\tilde{s}(\lambda)}\right)\Big|\mathbf{x}_i\right].$$

∘ If
  – $(\mathbf{x}_1,u_1),\ldots,(\mathbf{x}_n,u_n)$ is a random sample of $(\mathbf{x},u)$,
  – λ_0, $\boldsymbol{\beta}_0$ are known,

  we can define a consistent estimate of $M_0(\mathbf{x}_i,\lambda)$ by

$$M_{0n}(\mathbf{x}_i,\lambda,\lambda_0,\boldsymbol{\beta}_0) = \arg\min_a \frac{1}{n}\sum_{j=1}^n \hat{\rho}\left(\frac{(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + u_j)^{\lambda/\lambda_0} - a}{\tilde{s}(\lambda)}\right).$$

The M-conditional expectation estimate of λ_0

Model: $y_i^{\lambda_0} = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + u_i, \quad i = 1, \ldots, n,$
λ_0, $\boldsymbol{\beta}_0$ unknown.

Suppose that we consider μ as the true value λ_0 and we want to predict y^λ. To predict y^λ, we can use

$$M_{0n}(\mathbf{x}_i, \lambda, \lambda_0, \boldsymbol{\beta}_0) = \arg\min_a \frac{1}{n} \sum_{j=1}^{n} \hat{\rho} \left(\frac{(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + u_j)^{\lambda/\lambda_0} - a}{\tilde{s}(\lambda)} \right)$$

replacing

- λ_0 by μ and $\boldsymbol{\beta}_0$ by $\hat{\boldsymbol{\beta}}_n(\mu)$,
- the errors u_j by the residuals $r_j(\mu) = y_j^\mu - \mathbf{x}_j^{\mathrm{T}}\hat{\boldsymbol{\beta}}_n(\mu)$.

∘ We obtain the prediction function

$$M_n(\mathbf{x}_i, \lambda, \mu) = \arg\min_a \frac{1}{n} \sum_{j=1}^{n} \hat{\rho} \left(\frac{(\mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}_n(\mu) + r_j(\mu))^{\lambda/\mu} - a}{\tilde{s}(\lambda)} \right)$$

[robust extension of the *smearing estimate* (Duan, 1983), $\lambda = 1$, $\hat{\rho} = (\cdot)^2$].

∘ We estimate the MME of $M_n(\mathbf{x}_i^{\mathrm{T}}, \lambda, \mu)$ by

$$S_n(\lambda, \mu) = \frac{1}{n} \sum_{i=1}^{n} \hat{\rho} \left(\frac{y_i^\lambda - M_n(\mathbf{x}_i^{\mathrm{T}}, \lambda, \mu)}{\tilde{s}(\lambda)} \right).$$

∘ Let

$$m_n(\lambda) = \arg\min_\mu S_n(\lambda, \mu)$$

and define the *M-conditional expectation estimate (MCE-estimate)* $\hat{\lambda}_n$ of λ_0 by

$$m_n(\hat{\lambda}_n) = \hat{\lambda}_n.$$

Some results

Let

$$M(\mathbf{x}, \lambda, \mu) = \lim_{n\to\infty} M_n(\mathbf{x}, \lambda, \mu) \quad \text{a.s.}$$
$$S(\lambda, \mu) = \lim_{n\to\infty} S_n(\lambda, \mu) \quad \text{a.s.}$$
$$m(\lambda) = \arg\min_{\mu} S(\lambda, \mu).$$

- If the errors are i.i.d., homoscedastic + some regularity assumptions:
  1. $M(\mathbf{x}, \lambda, \lambda_0)$ is the M-conditional expectation of y^λ for given $\mathbf{x}$.
  2. The unique solution of $m(\lambda) = \lambda$ is λ_0.
  3. $\hat{\lambda}_n$ "is" (strongly) consistent.
- If u is symmetric, even heteroscedastic, $\lambda = \lambda_0$ is still a solution of $m(\lambda) = \lambda$, however, unicity not yet proved.
- Empirical results show that the asymptotic distribution is normal, however, a proof is still missing.
- Influence function estimates of the covariance matrix can be derived and used for inference.

Computation of the CME-estimate

The CME-estimate can be computed by solving the equation

$$\left.\frac{\partial S_n(\lambda, \mu)}{\partial \mu}\right|_{\mu=\lambda} = 0.$$

An efficient algorithm (based on resampling) has been worked out.

Monte Carlo results

Compare the following estimates:

RAC:	robust autocorrelation estimate
MCE:	M-conditional expectation estimate
F :	Foster et al. (JASA, 2001) based on minimisation of a distance between a parametric and a non parametric estimate of F; uses LS coeff.estimates
RF :	same as F, but based on 50%-bdp MM coeff.estimates
BI :	Bounded influence (Carrol and Ruppert, 1988)
ML :	Maximum likelihood

Results provide average values over 1000 samples of

bias = $\text{mean}(\hat{\lambda} - \lambda_0) \times 1000$, $\quad \sqrt{\text{mse}} = (\text{mean}(\hat{\lambda} - \lambda_0)^2)^{0.5} \times 1000$

Simple regression

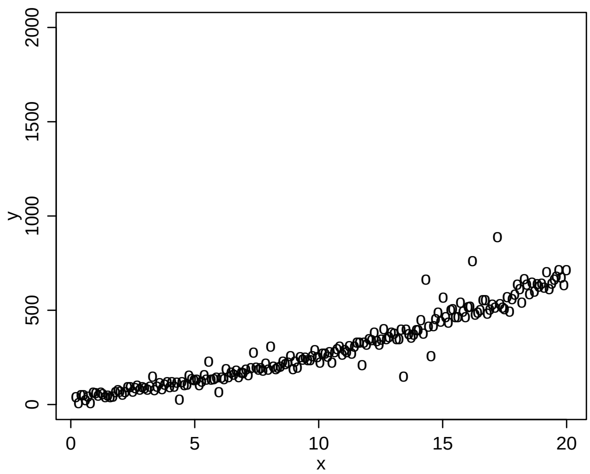
Simulate (x_i, y_i) according to

$$y_i^{0.5} = 10 + 2x_i + h(x_i)u_i, \quad i = 1, \ldots, 200,$$

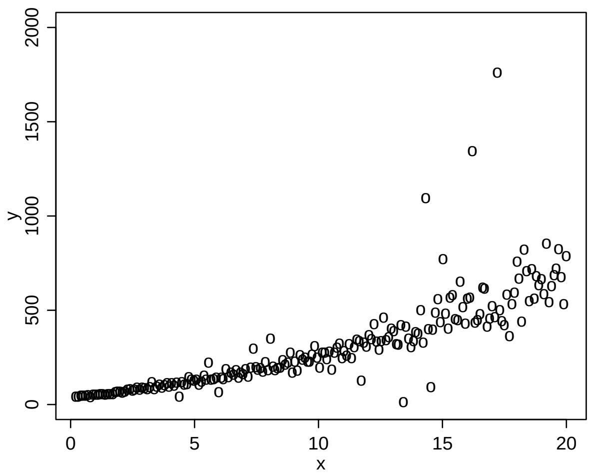
where

x_i: equally spaced on $[0, 20]$,
u_i: Uniform, Gauss, t_3, t_6, 10%-contaminated Gauss, Exponential
$E(u_i) = 0$, MAD$(u_i) = 1/3$.

Homoscedastic $h(x) = 3$



Heteroscedastic $h(x) = x/2$



Simple regression: homoscedastic case

		RAC	MCE	F	RF	BI	ML
Unif	bias	0	1	0	0	-2	0
	$\sqrt{\text{mse}}$	22	21	22	21	23	18
Gauss	bias	-1	0	0	0	-3	0
	$\sqrt{\text{mse}}$	28	26	27	26	26	23
t6	bias	0	0	1	1	-4	0
	$\sqrt{\text{mse}}$	29	27	29	28	26	29
t3	bias	-2	-1	-1	-1	-6	1
	$\sqrt{\text{mse}}$	31	27	30	29	27	54
CntG	bias	0	0	-1	-1	-4	0
	$\sqrt{\text{mse}}$	30	27	31	30	27	54
Exp	bias	-2	-1	-1	-1	2	-26
	$\sqrt{\text{mse}}$	28	25	24	22	26	44

Comments

short tails (Unif, Gauss)	:	performances are "similar" (ML slighly better)
long tails (t_3, CntG)	:	ML nonrobust; RAC, MCE, BI, RF, F robust
exponential	:	ML bad
	:	F and RF slightly better than RAC, MCE and BI

Simple regression: heteroschedastic case

		RAC	MCE	F	RF	BI	ML
Unif	bias	1	7	-34	-34	-120	-136
	$\sqrt{\text{mse}}$	45	61	62	54	124	139
Gauss	bias	2	4	-33	-32	-142	-192
	$\sqrt{\text{mse}}$	46	50	62	54	146	195
t6	bias	5	5	-30	-30	-149	-222
	$\sqrt{\text{mse}}$	49	49	63	55	153	226
t3	bias	1	-1	-36	-34	-158	-234
	$\sqrt{\text{mse}}$	47	44	66	56	162	252
CntG	bias	3	2	-36	-33	-146	-216
	$\sqrt{\text{mse}}$	48	46	68	58	150	242
Exp	bias	22	63	-56	-54	-171	-313
	$\sqrt{\text{mse}}$	70	80	87	74	179	319

Comments

symmetric : ML, BI, F, RF strongly biased
: RAC and MCE not biased
: RAC and MCE are robust
exponential : all estimates are strongly biased

Simulated coverages (%) of standard normal 95%-confidence intervals based on MCE

Homoscedastic case

	Unif	Gauss	t6	t3	CntG	Exp
Intercept	95.6	94.3	95.2	94.2	93.0	50.0
Slope	94.2	95.1	94.6	95.2	93.2	94.0
λ_0	94.0	94.8	95.0	95.2	93.8	94.3

Heteroscedastic case

	Unif	Gauss	t6	t3	CntG	Exp
Intercept	94.2	94.6	95.4	95.2	95.6	44.9
Slope	85.5	92.1	93.5	94.8	92.9	91.2
λ_0	84.4	92.4	93.2	94.9	92.4	70.2

Multiple regression: homoscedastic case

Simulate $(\mathbf{x}_i, y_i)$ according to

$$y_i^{0.5} = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + u_i, \quad i = 1, \ldots, 200,$$

$$\mathbf{x}_i = (1, x_{1i}, \ldots, x_{pi}), \quad \boldsymbol{\beta}_0 = (10, 1, \ldots, 1),$$

x_{ki}: uniform on $[0, 20]$,
u_i: Gauss, 10%-contaminated Gauss, $E(u_i) = 0$, $\mathrm{MAD}(u_i) = 1$.

		RAC	MCE	F	RF	BI	ML
Gauss $p = 4$	ave	-1	-1	0	0	-2	-1
	$\sqrt{\text{mse}}$	35	31	39	39	33	30
CntG $p = 4$	ave	-1	-2	-1	-1	-3	1
	$\sqrt{\text{mse}}$	39	32	43	42	34	66
Gauss $p = 8$	ave	-1	-1	0	0	-2	-1
	$\sqrt{\text{mse}}$	33	27	36	37	29	27
CntG $p = 8$	ave	0	0	3	3	2	0
	$\sqrt{\text{mse}}$	37	28	40	39	30	54

Comments

bias OK for all estimates
ML best for Gauss and worst for CntG
RAC and MCE outperform F, RF, and BI
MCE has the smallest mse

Multiple regression: heteroschedastic case

Simulate $(\mathbf{x}_i, y_i)$ according to

$$y_i^{0.5} = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + h(\mathbf{x}_i)u_i, \quad i = 1, \ldots, 200,$$

$$\mathbf{x}_i = (1, x_{1i}, \ldots, x_{pi}), \quad \boldsymbol{\beta}_0 = (10, 1, \ldots, 1),$$

$h(\mathbf{x}_i) = \sum_k x_{ki}/(6p)$,
x_{ki}: uniform on $[0, 20]$, $k = 1, \ldots, p$,
u_i: Gauss, 10%-contaminated Gauss, $E(u_i) = 0$, $\mathrm{MAD}(u_i) = 1$,

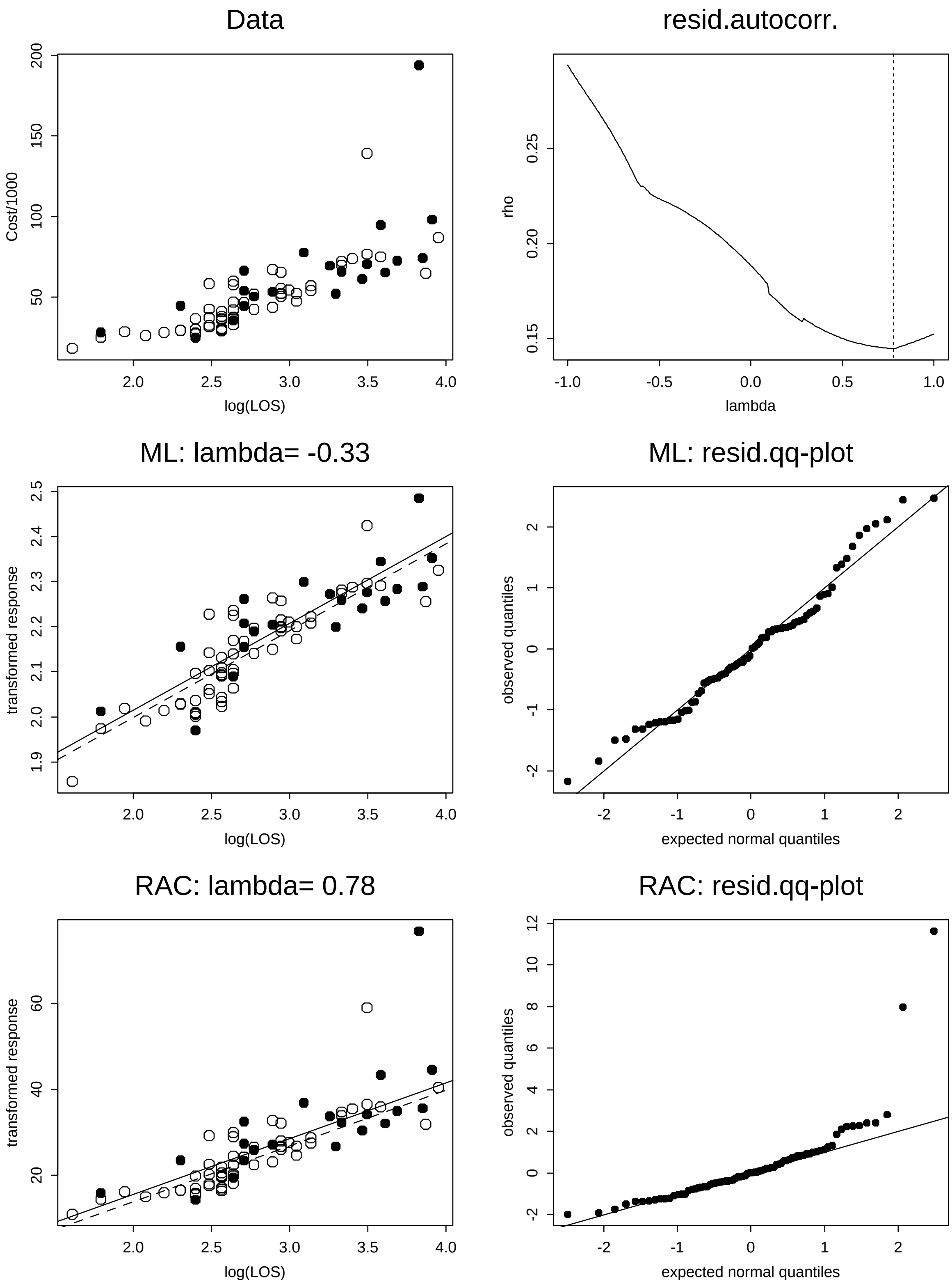
		RAC	MCE	F	RF	BI	ML
Gauss $p=4$	ave	-3	0	-28	-26	-86	-101
	$\sqrt{\text{mse}}$	59	54	72	69	100	112
CntG $p=4$	ave	-3	-2	-32	-27	-89	-137
	$\sqrt{\text{mse}}$	62	52	79	73	104	184
Gauss $p=8$	ave	-3	-2	-16	-13	-40	-50
	$\sqrt{\text{mse}}$	53	46	62	62	60	67
CntG $p=8$	ave	-2	0	-17	-11	-44	-98
	$\sqrt{\text{mse}}$	59	47	69	63	68	143

Comments

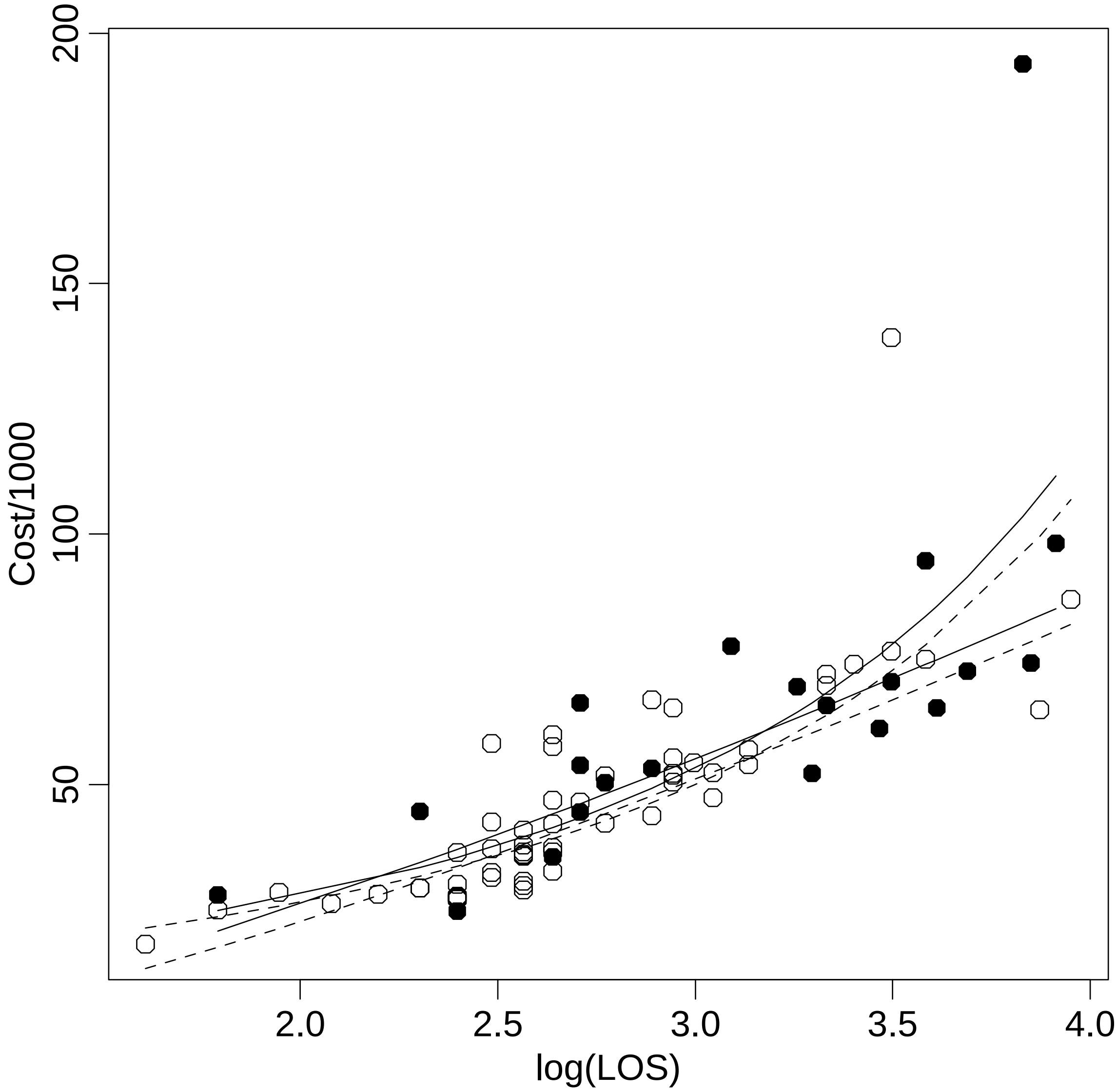
ML and BI are very bad
F and RF are biased
RAC and MCE provide satisfactory results

Examples

Valve cardiac surgery, with major cc
78 stays, CHUV Lausanne, 2000

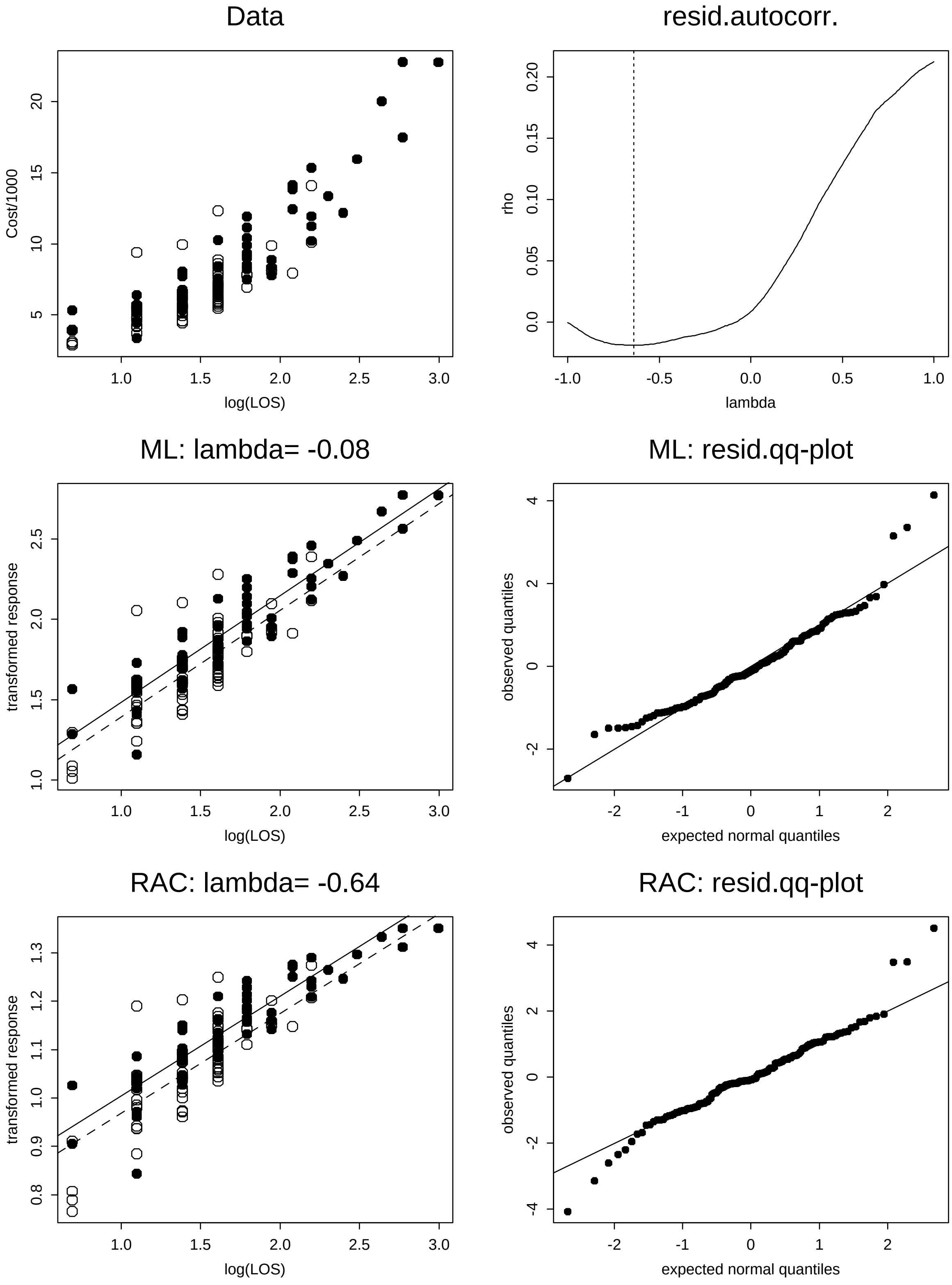


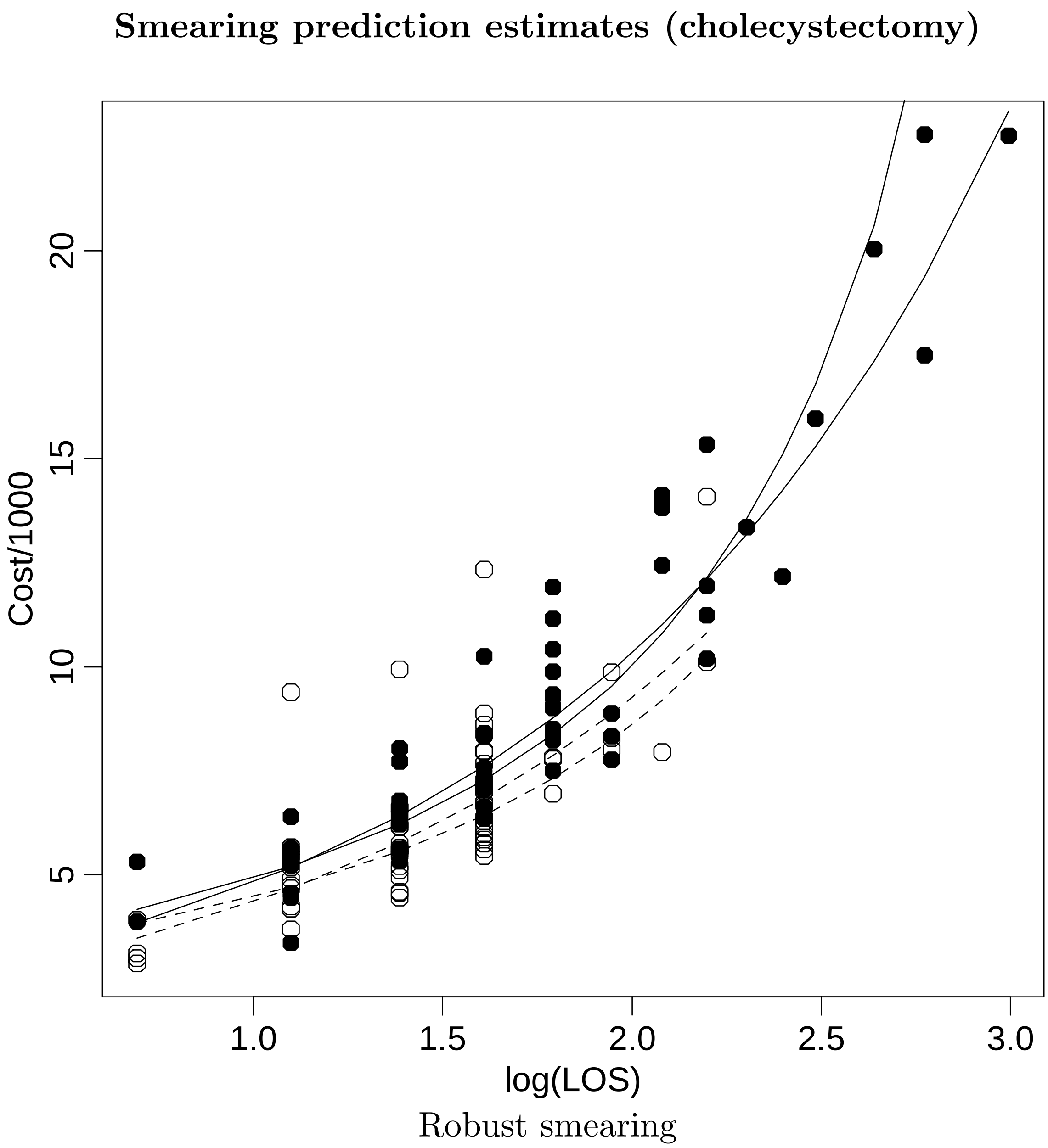
Smearing prediction estimates (valve cardiac surgery)



Robust smearing

Cholecystectomy without cc
135 stays, CHUV Lausanne, 2000




Smearing prediction estimates (cholecystectomy)
Cost/1000
5
10
15
20
1.0
1.5
2.0
2.5
3.0
log(LOS)
Robust smearing

Final remarks

- The analysis of hospital costs requires the development of new robust procedures for asymmetric responses.

- These methods are useful for other applications.

- Many open problems:

  - automatic model selection,
  - extensions of model families,
  - robust comparison (evaluation) of PCS,
  - improve DRG specific estimates using "pooled estimates"
  - ...

- Splus/R functions and documentation for

  - robust means M, TM, TML, ATML based on Lognormal, Weibull, and Gamma models,
  - TML and ATML regression based on Lognormal and Weibull errors,
  - one- and two- sample tests based on TM (Lognormal, Gamma, Weibull),

  are available at

  `www.hospvd.ch/iumsp`

  Some procedures are available in Splus6.

  Slides and references are available at

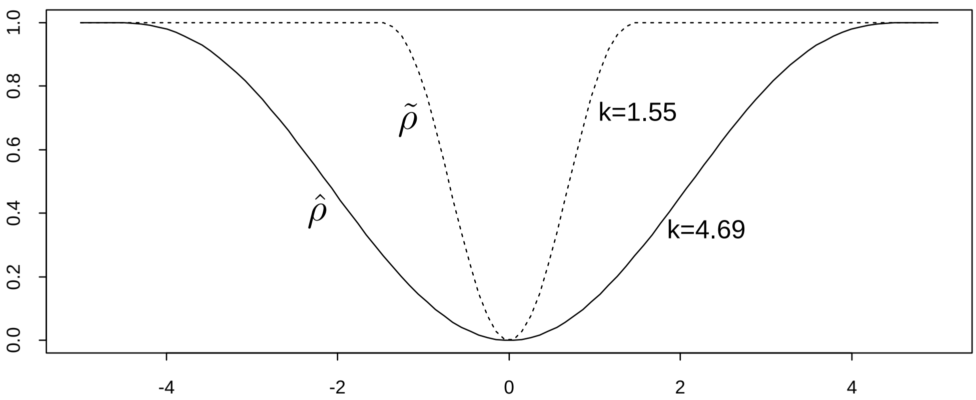
  `www.hospvd.ch/iumsp`

Appendix

S- and MM-estimates

Let $\tilde{\rho}$ and $\hat{\rho}$ be real functions stisfying $\hat{\rho} \leq \tilde{\rho}$ and other conditions, e.g., $\tilde{\rho} = \rho_{1.55}$, $\hat{\rho} = \rho_{4.69}$ in Tukey's biweight family:

$$\rho_k(z) = \begin{cases} 3(z/k)^2 - 3(z/k)^4 + (z/k)^6 & \text{if } |z| \leq k, \\ 1 & \text{if } |z| > k, \end{cases}$$



Suppose that we have a sample $(\mathbf{x}_1, z_1), \ldots, (\mathbf{x}_n, z_n)$ such that

$$z_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + u_i, \quad i = 1, \ldots, n, \quad \mathbf{x}_i \in \mathbb{R}^p, \quad \boldsymbol{\beta} \in \mathbb{R}^p.$$

<u>S-estimate</u> of $\boldsymbol{\beta}$ (Rousseeuw+Yohai, 1984):

$$\tilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \tilde{\sigma}_n(\boldsymbol{\beta}),$$

where $\tilde{\sigma}_n(\boldsymbol{\beta})$ solves

$$\frac{1}{n} \sum \tilde{\rho} \left(\frac{z_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}}{\tilde{\sigma}_n(\boldsymbol{\beta})} \right) = 0.5.$$

Associated scale estimate : $\tilde{\sigma}_n(\tilde{\boldsymbol{\beta}}_n)$.

○ With $\tilde{\rho} = \rho_{1.55}$, the breakdown point (bdp) of the S-estimate is 50%; the efficiency wrt ML for normal errors is very low (29%).

<u>MM-estimate</u> of $\boldsymbol{\beta}$ (Yohai, 1987):

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \hat{\rho} \left(\frac{z_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}}{\tilde{\sigma}_n(\tilde{\boldsymbol{\beta}}_n)} \right).$$

○ With $\hat{\rho} = \rho_{4.96}$ the bdp is 50% and the efficiency is very high: 95%.

The estimator of Foster et al. (JASA, 2001)

Model

$$y_i^{(\lambda_0)} = x_i^{\mathrm{T}} \beta_0 + u_i, \quad i = 1, \ldots, n,$$

where β is a vector of p parameters, x_i a vector of p covariates,
u_i i.i.d. $\sim F$, u_i independent of x_i,
$E(u_i) = 0$, $\mathrm{Var}(u_i) = \sigma^2$ (a constant).

For a given λ let $\beta_n(\lambda)$ an estimate (e.g., LS) of β based on $(x_i, y_i^{(\lambda)})$.
For $i = 1, \ldots, n$, an estimate of

$$\begin{aligned} P(y_i < t) &= P(y_i^{(\lambda_0)} - x_i^{\mathrm{T}} \beta_0 < t^{(\lambda_0)} - x_i^{\mathrm{T}} \beta_0) \\ &= F(t^{(\lambda_0)} - x_i^{\mathrm{T}} \beta_0) \end{aligned}$$

is

$$F_{n,i}(\lambda, t) = \frac{1}{n} \sum_{j=1}^{n} I(r_j(\lambda) < t^{(\lambda)} - x_i^{\mathrm{T}} \beta_n(\lambda)),$$

where I is the indicator function and

$$r_i(\lambda) = y_i^{(\lambda)} - x_i^{\mathrm{T}} \beta_n(\lambda).$$

A measure of error is

$$q_n(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\infty \left[I(y_i < t) - F_{n,i}(\lambda, t) \right]^2 dW(t),$$

where W is a positive (differentiable) strictly increasing bounded weight function (e.g., the cdf of a normal deviate with mean $\sum y_i/n$ and variance $\sum (y_i - \bar{y})^2/(n-1)$).

Define

$$\tilde{\lambda}_n = \operatorname{argmin}\, q_n(\lambda).$$

References (some available at `www.hospvd.ch/iumsp`)

Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A., 1986. Robust statistics: An approach based on the influence function. Wiley, New York.

Marazzi A., Ruffieux C., 1996. Implementing M-estimators of the gamma distribution. In: H. Rieder (Ed.), Robust Statistics, Data Analysis, and Computer Intensive Methods, In Honor of Peter Huber's 60th Birthday, Lecture Notes in Statistics, 109, Springer Verlag, Heidelberg.

Marazzi A., Paccaud F., Ruffieux C., Beguin C., 1998. Fitting the distributions of length of stay by parametric models. Medical Care 36(6) 915-927.

Marazzi A., Ruffieux C., 1999. The truncated mean of an asymmetric distribution. Computational Statistics and Data Analysis, 32(1), 79-100.

Marazzi A., 2002. Bootstrap tests for robust means of asymmetric distributions with unequal shapes. Computational Statistics & Data Analysis. 39(4), 503-528.

Marazzi A., Barbati G., 2003. Robust parametric means of asymmetric distributions: estimation and testing. Estadistica, Vol 54/162-3, pp.47-72.

Marazzi A.,Yohai V.J., 2004. Adaptively truncated maximum likelihood regression with asymmetric errors. Journal of Statistical Planning and Inference, Vol 122/1-2 pp 271-291.

Marazzi A.,Yohai V.J., 2004. Robust Box-Cox Transformations for Simple Regression". In: Theory and Applications of Recent Robust Methods, edited by M.Hubert, G. Pison, A. Struyf and S. Van Aelst, Series: Statistics for Industry and Technology, Birkhauser, Basel.

Marazzi A.,Yohai V.J., 2004. Robust Box-Cox transformations based on minimum residual autocorrelation. Draft. Submitted.

Rousseeuw P. J., Leroy A.M, 1987. Robust regression and outlier detection. Wiley, New York.

Salibian-Barrera M., Zamar R. H., 2002. Bootstrapping robust estimates of regression. The Annals of Statistics, 30 (2) 556-582.

Victoria-Feser M.P., Ronchetti E., 1994. Robust methods for personal-income distribution models. The Canadian Journal of Statistics, 22, 247-258.

Victoria-Feser M.P., Ronchetti E., 1997. Robust estimation for grouped data. Journal of the American Statistical Association, 92(437), 333-340.

Victoria-Feser M.P., 2000. Robust methods for the analysis of income distribution, inequality and poverity. International Statistical Review, 68, 3 277-293.