

Robust Regression and Scale Based on Infinitesimal Neighborhoods:

M- vs. AL-Estimators

Matthias Kohl and Helmut Rieder



UNIVERSITÄT
BAYREUTH

Mathematics VII

Fourth International Conference on Statistical Data Analysis
Based on the L1-Norm and Related Methods

Université de Neuchâtel

August 9th 2002

E-mail: matthias.kohl@uni-bayreuth.de



I Linear regression and scale

- Ideal Model:

$$y = x^T \beta + \sigma u \quad x \sim K(dx), \quad u \sim F(du) \text{ sto. independent}$$

- Assumptions on F and K :

(F1) F is symmetric

(F2) $\mathcal{I}_F^{\text{loc}} < \infty$ (finite Fisher information of location)

i.e., F has abs. continuous density f and
 $\mathcal{I}_F^{\text{loc}} = \int (f'/f)^2 f d\lambda < \infty$ (Huber 81)

(F3) $\mathcal{I}_F^{\text{sc}} < \infty$ (finite Fisher information of scale)

i.e., $u \mapsto u f(u) =: u f$ abs. continuous and
 $\mathcal{I}_F^{\text{sc}} = \int (1 + u(f'/f))^2 f d\lambda < \infty$ (Rieder, Ruckdeschel 02)

(K) $\text{rk}(K) = k$ with $K := \int x x^T K(dx) \in \mathbb{R}^k \times k$



I-1

- Model distributions: $\theta = (\beta^T, \sigma)^T \in \mathbb{R}^k \times (0, \infty)$

$$P_\theta(dx, dy) = \frac{1}{\sigma} f\left(\frac{y - x^T \beta}{\sigma}\right) \lambda(dy) K(dx)$$

model invariant under $g_\theta(x, y) := (x, x^T \beta + \sigma y)$;

i.e., $P_\theta = g_\theta(P_{\theta_0})$, where $\theta_0 := (0, 1)^T$.

- Scores:

$$\Delta_\theta(x, y) = \frac{1}{\sigma} \begin{pmatrix} x \Lambda_f\left(\frac{y - x^T \beta}{\sigma}\right) \\ \frac{y - x^T \beta}{\sigma} \Lambda_f\left(\frac{y - x^T \beta}{\sigma}\right) - 1 \end{pmatrix} = \frac{1}{\sigma} \Lambda_{\theta_0} \circ g_\theta^{-1}(x, y)$$

where $\Lambda_f(u) := -(f'/f)(u)$ and $u = \frac{y - x^T \beta}{\sigma}$.

- Fisher information: $\mathcal{I}_\theta = \frac{1}{\sigma^2} \mathcal{I}_{\theta_0}$, where by (F1)

$$\mathcal{I}_{\theta_0} = \begin{pmatrix} \mathcal{I}_f^{\text{loc}} K & 0 \\ 0 & \mathcal{I}_f^{\text{scal}} \end{pmatrix}$$



I-2

- Proposition:
The parametric family

$$\mathcal{P} := \{P_\theta \mid \theta = (\beta^T, \sigma)^T \in \mathbb{R}^k \times (0, \infty)\}$$

is L_2 -differentiable at θ with L_2 -derivative Δ_θ .

[Swensen (80), Ruckdeschel (01), Rieder (94)]

- Influence Curves: to given θ the set of all $\eta_\theta \in L_2^{k+1}(P_\theta)$ s.t.

$$\mathbb{E} \eta_\theta = 0, \quad \mathbb{E} \eta_\theta \Delta_\theta^T = \mathbb{I}_{k+1}$$

- As. Linear Estimators: Class of all sequences of estimators S_n for θ s.t.

$$\sqrt{n}(S_n - \theta) = \frac{1}{n} \sum_{j=1}^n \eta_\theta((x^T, y)_j^T) + o_{P_\theta}(n^0)$$



I-3

- Correspondence between the set of ICs at θ and the set of ICs at $\theta_0 = (0, 1)^\tau$:

$$\eta_\theta(x, y) = \sigma \eta_{\theta_0} \left(x, \frac{y - x^\tau \beta}{\sigma} \right) = \sigma \eta_{\theta_0} \circ g_\theta^{-1}(x, y)$$

- unconditional, or errors-in-variables, r/\sqrt{n} -neighborhoods about P_{θ_0} at sample size n :

$$Q_n = \left(1 - \frac{r}{\sqrt{n}} \right) P_{\theta_0} + \frac{r}{\sqrt{n}} H$$

In the sequel, expectation is always taken under P_{θ_0} .



I-4

Schematic overview of the talk

	AL	M	BM
total bound	II	III	—
separate bounds	IV	V	VI

- total bound: simultaneous estimation of β and σ
- separate bounds: sequential estimation of β and σ



I-6

Motivation

Following Huber (81, p.135) the simultaneous M-estimate of location and scale is any pair of estimators (T_n, S_n) determined by two equations of form (ψ generalizes Δ_f)

$$(4.3) \quad \sum_{i=1}^n \psi \left(\frac{u_i - T_n}{S_n} \right) = 0$$

$$(4.4) \quad \sum_{i=1}^n \left[\frac{u_i - T_n}{S_n} \psi \left(\frac{u_i - T_n}{S_n} \right) - 1 \right] = 0 \quad (\text{M})$$

where he without argument generalizes (4.4) to his

$$(4.6) \quad \sum_{i=1}^n \chi \left(\frac{u_i - T_n}{S_n} \right) = 0 \quad (\text{AL})$$

χ an arbitrary function, independent of ψ

II AL-estimators: total bound

- MSE problem for $r \in [0, \infty)$, $\eta \in L_2^{k+1}(P_{\theta_0})$:

$$\text{E}|\eta|^2 + r^2 \sup_{P_{\theta_0}} |\eta|^2 = \min! \quad \text{E}\eta = 0, \quad \text{E}\eta \Delta_{\theta_0}^\tau = \mathbb{I}_{k+1}$$

- Solution by Riederer (94) using (F1):

$$\tilde{\eta}(x, u) = \begin{pmatrix} \tilde{\eta}_{\text{reg}}(x, u) \\ \tilde{\eta}_{\text{sc}}(x, u) \end{pmatrix} = \begin{pmatrix} A_{\text{reg}} x \Delta_f(u) \\ A_{\text{sc}}(u \Delta_f(u) - a_{\text{sc}}) \end{pmatrix} w(x, u)$$

with

$$w(x, u) = \min \left\{ 1, \frac{b}{|Y|} \right\}$$

and

$$|Y| := [A_{\text{reg}} x^2 \Delta_f(u)^2 + A_{\text{sc}}^2 (u \Delta_f(u) - a_{\text{sc}})^2]^{1/2}$$



I-5

II-1

determine a_{sc} , A_{rg} , A_{sc} , b by the implicit equations

$$0 = E(u \Delta_f(u) - a_{sc}) w(x, u)$$

$$\mathbb{I}_k = A_{rg} E x x^T \Delta_f(u)^2 w(x, u)$$

$$1 = A_{sc} E(u \Delta_f(u) - a_{sc})^2 w(x, u)$$

and

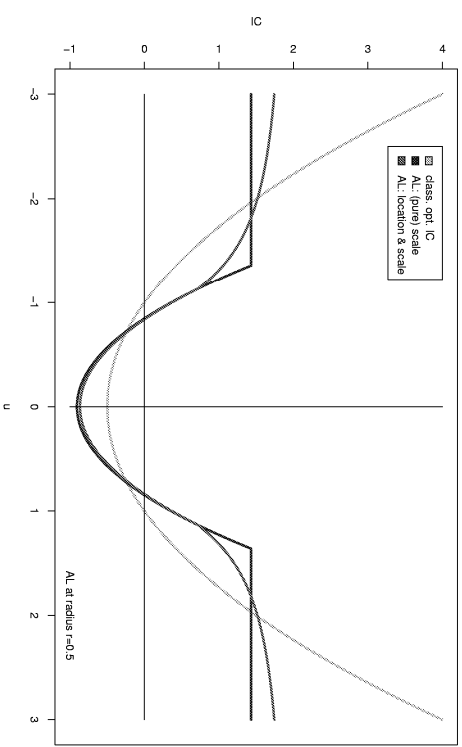
$$r^2 b = E(|Y| - b)_+$$

• Proposition:

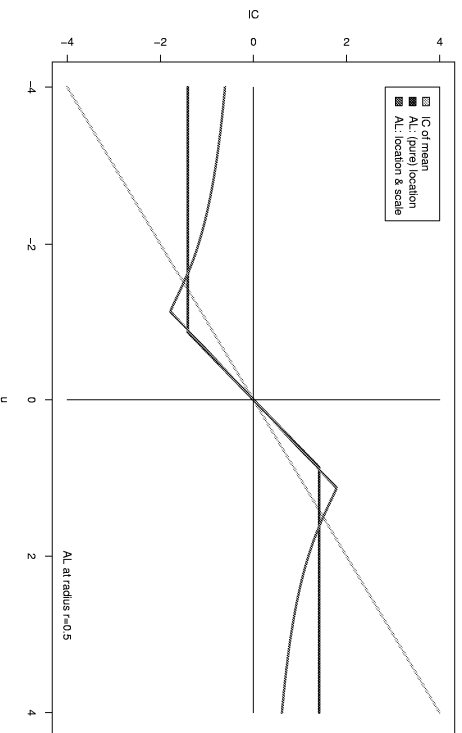
$$\max \text{MSE}(\tilde{\eta}, r) = E|\tilde{\eta}|^2 + r^2 \sup_{P_{\theta_0}} |\tilde{\eta}|^2 = \text{tr } A_{rg} + A_{sc}$$



Influence curves for scale: $F=N(0,1)$



Influence curves for location: $F=N(0,1)$



III M-estimators: total bound

• Ideal error distribution: $F = \mathcal{N}(0, 1)$

• ICs for M-estimators: $\rho \in L_2^{k+1}(P_{\theta_0})$

$$\rho(x, u) = \begin{pmatrix} K^{-1} x \psi(x, u) \\ \gamma(u \psi(x, u) - 1) \end{pmatrix} \quad \gamma \in \mathbb{R} \text{ fixed}$$

where $E \rho = 0$, $E \rho \Delta_{\theta_0}^T = \mathbb{I}_{k+1}$

• Sufficient conditions for $\psi \in L_2(P_{\theta_0})$: ($E_{\bullet} = \int dF$)

(MS1) $\psi(x, \cdot)$ is odd in u a.e. $K(dx)$

(MS2) $E_{\bullet} u \psi(x, u) = 1$ a.e. $K(dx)$

(MS3) $E u^3 \psi(x, u) = 1 + \gamma^{-1}$



- MSE problem for $r \in [0, \infty)$, $\gamma \in \mathbb{R}$ fixed:

$$E|\rho|^2 + r^2 \sup_{P_{\theta_0}} |\rho|^2 = \min! \quad \text{(MS1)-(MS3)}$$

- Solution for fixed γ :

(by Lagrange arguments as in Rieder (94))

$$\tilde{\psi}(x, u) = \frac{\alpha_1(x)u + \alpha_3u^3}{q(x, u)^2} w(x, u) + \gamma^2 \frac{u}{q(x, u)^2}$$

where

$$w(x, u) = \min \left\{ 1, \frac{b(x, u)}{|g(x, u)|} \right\} \quad q(x, u)^2 = |K^{-1}x|^2 + \gamma^2 u^2$$

$$g(x, u) = \frac{\alpha_1(x)u + \alpha_3u^3}{q(x, u)} \quad b(x, u) = \left[b_0^2 - \gamma^2 \frac{|K^{-1}x|^2}{q(x, u)^2} \right]^{1/2}$$



Results for location and scale: $F = \mathcal{N}(0, 1)$

r	MSE _{AL,0}	MSE _{M,0}	γ_{opt}	MSE _{M,0} /MSE _{AL,0}
0.1	1.648	1.648	0.575	1.000
0.25	2.011	2.013	0.712	1.001
0.5	2.820	2.829	0.940	1.003
0.75	3.859	3.877	1.158	1.0047
0.825	4.216	4.236	1.220	1.0048
1.0	5.137	5.159	1.356	1.0043
1.25	6.679	6.693	1.510	1.002
1.5	8.506	8.508	1.575	1.000
1.75	10.637	10.641	1.581	1.000
2.0	13.092	13.096	1.584	1.000
∞	∞	∞	1.589	1.000



determine $\alpha_1(x)$, α_3 and b_0 by the implicit equations

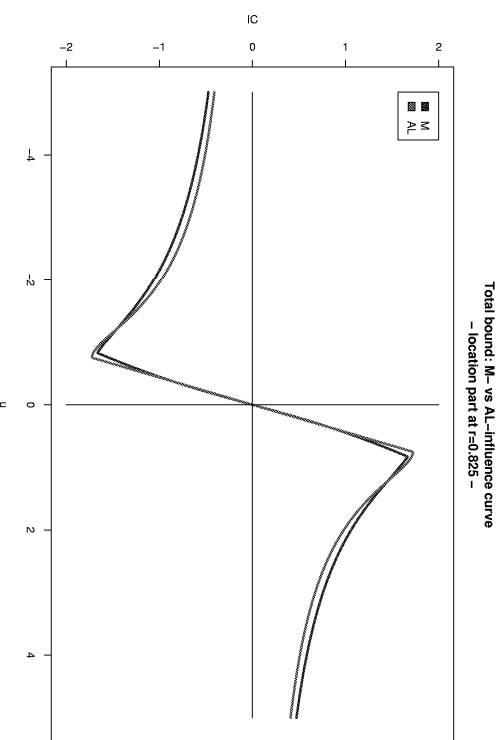
$$1 = E_{\bullet} u \tilde{\psi}(x, u)$$

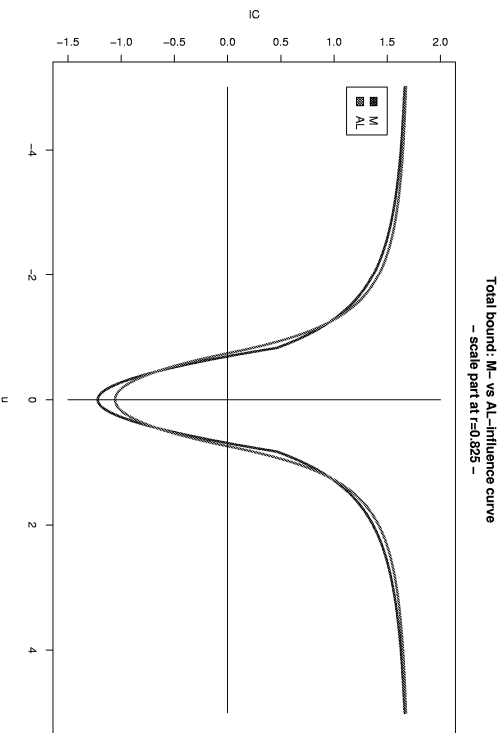
$$1 + \gamma^{-1} = E u^3 \tilde{\psi}(x, u)$$

and

$$r^2 = E \left(\frac{|g(x, u)|}{b(x, u)} - 1 \right)_+$$

- Additionally outer optimization in γ necessary! (done numerically)





IV AL-estimators: separate bounds

- MSE problem for $r \in [0, \infty)$, $\eta_{reg} \in L_2^k(P_{\theta_0})$, $\eta_{sc} \in L_2(P_{\theta_0})$:

$$\mathbb{E} |\eta_{reg}|^2 + \mathbb{E} |\eta_{sc}|^2 + r^2 (\sup P_{\theta_0} |\eta_{reg}|^2 + \sup P_{\theta_0} |\eta_{sc}|^2) = \min!$$

under

$$\begin{aligned} \mathbb{E} \eta_{reg} &= 0 & \mathbb{E} \eta_{reg} \Lambda_f x^T &= \mathbb{I}_k & \mathbb{E} \eta_{reg} u \Lambda_f &= 0 \\ \mathbb{E} \eta_{sc} &= 0 & \mathbb{E} \eta_{sc} u \Lambda_f &= 1 & \mathbb{E} \eta_{sc} \Lambda_f x^T &= 0 \end{aligned}$$

- MSE problem falls apart into two separate MSE problems by robust adaptivity, which follows by **(F1)**



- Problem 1 (regression part):** $\eta_{reg} \in L_2^k(P_{\theta_0})$
 $\mathbb{E} |\eta_{reg}|^2 + r^2 \sup P_{\theta_0} |\eta_{reg}|^2 = \min!$
 under $\mathbb{E} \eta_{reg} = 0$ $\mathbb{E} \eta_{reg} \Lambda_f x^T = \mathbb{I}_k$
- Solution for η_{reg} by Riederer (94) using (F1):**

$$\tilde{\eta}_{reg}(x, u) = A_{reg} x \Lambda_f(u) \min \left\{ 1, \frac{b_{reg}}{|A_{reg} x \Lambda_f(u)|} \right\}$$

 where
$$\mathbb{I}_k = A_{reg} \mathbb{E} \Lambda_f(u)^2 x x^T \min \left\{ 1, \frac{b_{reg}}{|A_{reg} x \Lambda_f(u)|} \right\}$$

$$\text{and } r^2 b_{reg} = \mathbb{E} (|A_{reg} x \Lambda_f(u)| - b_{reg})_+$$

- Problem 2 (scale part):** $\eta_{sc} \in L_2(P_{\theta_0})$
 $\mathbb{E} |\eta_{sc}|^2 + r^2 \sup P_{\theta_0} |\eta_{sc}|^2 = \min!$
 under $\mathbb{E} \eta_{sc} = 0$ $\mathbb{E} \eta_{sc} u \Lambda_f = 1$

- Solution for η_{sc} by Riederer (94):** $c_{sc} := \frac{b_{sc}}{A_{sc}}$

$$\tilde{\eta}_{sc}(u) = A_{sc}(u \Lambda_f(u) - \alpha_{sc}^2) \min \left\{ 1, \frac{c_{sc}}{|u \Lambda_f(u) - \alpha_{sc}^2|} \right\}$$

where

$$0 = \mathbb{E}(u \Lambda_f(u) - \alpha_{sc}^2) \min \left\{ 1, \frac{c_{sc}}{|u \Lambda_f(u) - \alpha_{sc}^2|} \right\}$$

$$1 = A_{sc} \mathbb{E} |u \Lambda_f(u) - \alpha_{sc}^2| \min \{|u \Lambda_f(u) - \alpha_{sc}^2|, c_{sc}\}$$

and

$$r^2 c_{sc} = \mathbb{E} (|u \Lambda_f(u) - \alpha_{sc}^2| - c_{sc})_+$$



V M-estimators: separate bounds

- Ideal error distribution: $F = \mathcal{N}(0, 1)$
- ICS for M-estimators: $\rho \in L_2^{k+1}(P_{\theta_0})$

$$\rho(x; u) = \begin{pmatrix} \mathcal{K}^{-1} x \psi(x, u) \\ \gamma(u \psi(x, u) - 1) \end{pmatrix} \quad \gamma \in \mathbb{R} \text{ fixed}$$

where $E\rho = 0$, $E\rho \Delta_{\theta_0}^T = \mathbb{I}_{k+1}$

- Sufficient conditions for $\psi \in L_2(P_{\theta_0})$: ($E\bullet = \int dF$)
 - (MS1) $\psi(x, \cdot)$ is odd in u a.e. $K(dx)$
 - (MS2) $E\bullet u \psi(x, u) = 1$ a.e. $K(dx)$
 - (MS3) $E u^3 \psi(x, u) = 1 + \gamma^{-1}$



- Solution for fixed γ : (by Lagrange arguments)

$$\tilde{\psi}(x, u) = g(x, u)u(x, u) \quad w(x, u) = \min \left\{ 1, \frac{b_{rg,sc}(x, u)}{|g(x, u)|} \right\}$$

where

$$g(x, u) = \frac{\alpha_1(x)u + \alpha_3 u^3}{q^2(x, u)} \quad b_{rg,sc} = \min \left\{ \frac{b_{rg}}{|\tilde{x}|}, \frac{1 + b_{sc}}{|u|} \right\}$$

$$q(x, u) = \sqrt{|\tilde{x}|^2 + \gamma^2 u^2} \quad \tilde{x} = \mathcal{K}^{-1}x$$

determine $\alpha_1(x)$, α_3 , b_{rg} and b_{sc} by the implicit equations

$$E\bullet u \tilde{\psi} = 1 \quad E u^3 \tilde{\psi} = 1 + \gamma^{-1}$$

$$\begin{aligned} r^2 b_{rg} &= E \frac{1}{|\tilde{x}|} E\bullet \left[\alpha_1(x)u + \alpha_3 u^3 - \frac{b_{rg} q^2}{|\tilde{x}|} \right] + \mathbf{1} \left\{ \frac{b_{rg}}{|\tilde{x}|} < \frac{1 + b_{sc}}{|u|} \right\} \\ r^2 \gamma^2 b_{sc} &= E E\bullet \left[\alpha_1(x) + \alpha_3 u^2 - \frac{(1 + b_{sc}) q^2}{u^2} \right] + \mathbf{1} \left\{ \frac{b_{rg}}{|\tilde{x}|} > \frac{1 + b_{sc}}{|u|} \right\} \end{aligned}$$

- Additionally outer optimization in γ necessary!



- MSE problem for $r \in [0, \infty)$:

$$E|\rho|^2 + r^2 \omega_{M,2}^2(\rho) = \min! \quad \text{(MS1)-(MS3)}$$

where

$$\omega_{M,2}^2(\rho) = \sup P_{\theta_0} |\mathcal{K}^{-1}x|^2 \psi^2 + \gamma^2 \sup P_{\theta_0} |u\psi - 1|^2$$

- Robust adaptive, if

$$E x u^2 \psi(x, u) = 0$$

which follows by (F1) and (MS1)

- But: This problem doesn't fall apart into two separate problems, since the two components of ρ are coupled via ψ .



Results for location and scale: $F = \mathcal{N}(0, 1)$

r	$MSE_{AL,2}$	$MSE_{M,2}$	γ_{opt}	$MSE_{M,2}/MSE_{AL,2}$
0.1	1.651	1.660	0.580	1.005
0.25	2.030	2.073	0.725	1.021
0.5	2.900	3.056	0.953	1.054
0.75	4.048	4.406	1.140	1.088
1.0	5.480	6.152	1.276	1.123
1.25	7.224	8.317	1.358	1.151
1.5	9.305	10.941	1.360	1.176
1.75	11.733	14.040	1.361	1.197
2.0	14.515	17.615	1.362	1.214
∞	∞	∞	1.395	1.295



VI BM-estimators: separate bounds

- Only location and scale with $F = \mathcal{N}(0,1)$
- Bednarski and Müller (01) propose (suboptimal) estimators for location and scale by separately solving

$$E\psi^2 = \min! \quad \text{under (Ms1), } |\psi| \leq b_{loc}$$
 and

$$E|u\psi - 1|^2 = \min! \quad \text{under (Ms1)-(Ms3), } |u\psi - 1| \leq b_{sc}$$
- Solution by Bednarski and Müller (01):

$$\tilde{\psi} = \text{sign}(u) \min \left\{ \alpha|u|, b_{loc}, \frac{1+b_{sc}}{|u|} \right\}$$

$$1 = E \min \{ \alpha u^2, b_{loc}|u|, (1+b_{sc}) \}$$
 solves both problems, if γ in (Ms3) is determined by

$$\gamma = [E \min \{ \alpha u^4, b_{loc}|u|^3, (1+b_{sc})u^2 \} - 1]^{-1}$$



VI-1

Results for location and scale: $F = \mathcal{N}(0,1)$

r	$b_{M,L}$	$b_{M,S}$	$MSE_{M,2}$	$MSE_{BM,2}$	γ_M	γ_{BM}
0.1	2.119	4.457	1.6595	1.6595	0.5796	0.5784
0.25	1.701	2.599	2.0727	2.0728	0.7248	0.7229
0.5	1.489	1.636	3.0559	3.0563	0.9534	0.9520
0.75	1.426	1.265	4.4062	4.4065	1.1402	1.1397
1.0	1.408	1.087	6.1517	6.1519	1.2758	1.2756
1.25	1.405	1.000	8.3174	8.3175	1.3582	1.3582
1.5	1.403	1.000	10.9412	10.9412	1.3598	1.3598
1.75	1.401	1.000	14.0402	14.0402	1.3610	1.3610
2.0	1.400	1.000	17.6147	17.6148	1.3618	1.3619
∞	1.360	1.000	∞	∞	1.3945	1.3945

But: confer table V-4 as for the suboptimality of M, BM with respect to ALI



VI-3

- No outer optimization in $\gamma \implies$ only suboptimal solution
- For computation of the MSE we provide the optimal bounds b_{loc} and b_{sc} of the M-estimators
- Generalization to regression and scale possible, if
 - the ideal regressor distribution K has finite support
 - (Ms3) is replaced by

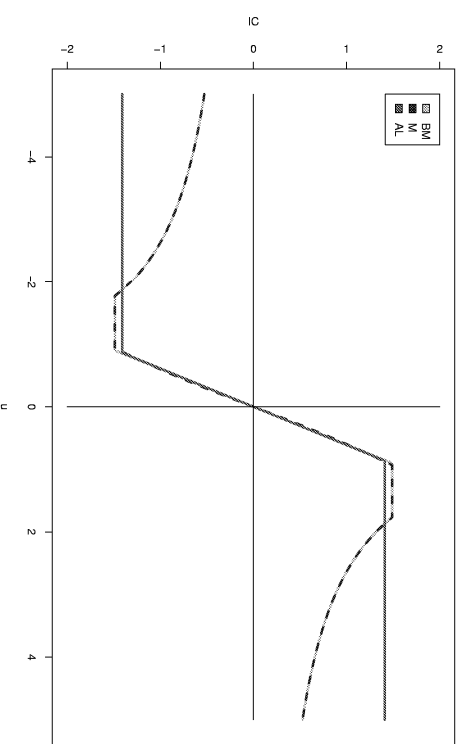
$$(Ms3_{\bullet}) E_{\bullet} u^3 \psi(x, u) = 1 + \gamma x^{-1} \quad \text{a.e. } K(dx)$$

Then: solve a location and scale problem at each design point x



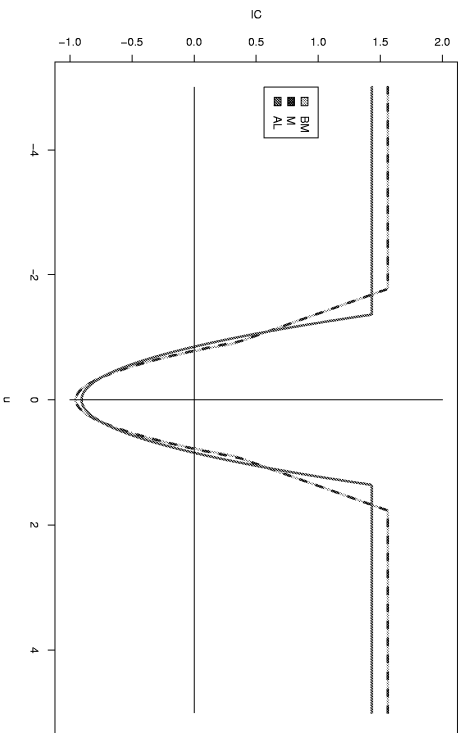
VI-2

Separate bounds: BM- vs. M- vs. ALI-influence curve
– location part at $r=0.5$ –



VI-4

Separate bounds: BM- vs. M- vs. AL- influence curve
 - scale part at $\tau=0.5$ -



VI-5

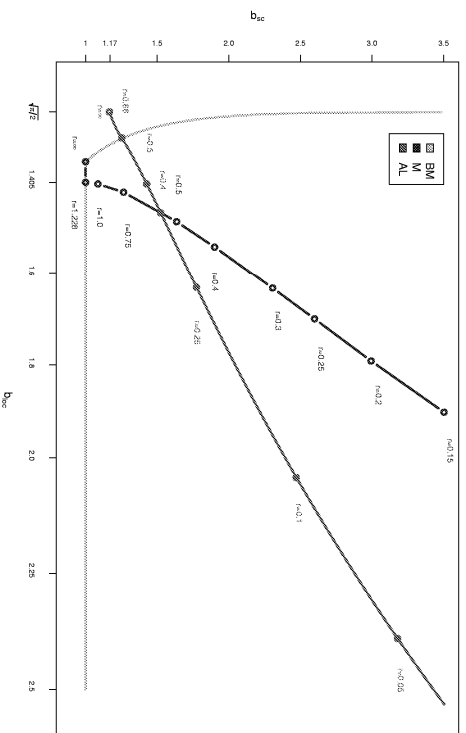
VII References

- [1] Bednarski, T., Müller, Ch. (2001): Optimal bounded influence regression and scale M-estimators in the context of experimental design. *Statistics* 35 (2001), pp. 349-369.
- [2] Huber, P.J. (1981): *Robust Statistics*. Wiley, New York.
- [3] Rieder, H. (1994): *Robust Asymptotic Statistics*. Springer, New York.
- [4] Rieder, H., Kohl, M., Ruckdeschel, P. (2001): The Costs of Not Knowing the Radius. Discussion Paper. SFB 373, Berlin.
- [5] Rieder, H., Ruckdeschel, P. (2002): Fisher Information for Parametric Group Models. Manuscript
- [6] Ruckdeschel, P. (2001): *Ansätze zur Robustifizierung des Kalman-Filters*. Dissertation, University of Bayreuth.
- [7] Swensen, A.R. (1980): *Asymptotic Inference for a class of stochastic processes*. Dissertation, University of California, Berkeley, CA.



Ref.

(Optimal) bounds for AL-, M- and BM-estimators



VI-6