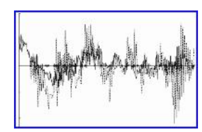
# ROBUST ESTIMATION FOR TIME SERIES MODELS
## Based On Infinitesimal Neighborhoods

Helmut Rieder

University of Bayreuth; EPF Lausanne, 12 June 2003

1. **Elementary robustness for time series** (up to 1987)
2. **LAN time series models**
3. **Influence curves and asy. linear (AL) estimators**
4. **Neighborhoods of transition probabilities**
5. **Asymptotic maximum mean square error over nbds**
6. **Optimally robust influence curves**
   - minimax MSE
   - robust adaptivity (with M. Kohl)
7. **Estimator construction** (with P. Ruckdeschel)
8. **Unknown neighborhood radius** (with P. Ruckdeschel, M. Kohl)





# 1 Elementary Concepts

References: Martin[85], Martin/Yohai[86], Dutter/Stockinger[87]

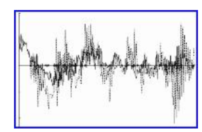
ARMA$(p,q)$**:** $\quad \boldsymbol{\phi(B)(X_t - \mu) = \xi(B)V_t}, \quad t = \ldots, -1, 0, 1, \ldots$
Innovations $V_t$ i.i.d.$\sim F$, $\mu_F = 0$, $\sigma_F^2 < \infty$, $\mathcal{I}_F^{\mathrm{loc}} < \infty$.
Stationarity and invertibility assumptions on $\phi$ and $\xi$:
$\quad \phi(z)\xi(z) \neq 0 \quad (|z| \leq 1)$; no common zeros
Parameter $\theta = (\phi, \xi, \mu)$; $\sigma_F^2$ assumed known (mostly).

Quick derivation of the **scores $\boldsymbol{\Lambda_{\theta,t}}$ of transition probabilities:**
Write $X_t = \mu + \psi\, V_t$ with $\psi = \xi/\phi$; then, setting $\psi_1 = (\psi - 1)/B$,

$$\begin{aligned} & P_\theta\big(X_t \in dx_t \big| \bar{X}_{t-1} = \bar{x}_{t-1}\big) \\ &\qquad = P_\theta\big(\mu + V_t + \psi_1 V_{t-1} \in dx_t \big| \bar{X}_{t-1} = \bar{x}_{t-1}\big) \\ &\qquad = P_\theta\big(\mu + V_t + \psi_1 \psi^{-1}(X_{t-1} - \mu) \in dx_t \big| \bar{X}_{t-1} = \bar{x}_{t-1}\big) \\ &\qquad = f\big(x_t - \mu - \psi_1\psi^{-1}(x_{t-1} - \mu)\big)\, dx_t \end{aligned}$$

where $\psi_1\psi^{-1} = (1 - \psi^{-1})/B$ and $\psi^{-1}(x_t - \mu) = v_t$.





Hence, by the chain rule, $\frac{\partial}{\partial\theta}\log f\big(x_t-\mu-\psi_1\psi^{-1}(x_{t-1}-\mu)\big)$ equals

$$\Lambda_F(v_t)\,\frac{\partial}{\partial\theta}\Big\{\mu+(1-\psi^{-1})(x_t-\mu)\Big\}\qquad\text{where }\ \Lambda_F=-\frac{\dot f}{f}$$

But $\psi^{-1}(B)\mu=\psi^{-1}(1)\mu$, hence $\partial/\partial\mu\,\{\ldots\}=\phi(1)/\xi(1)$. Moreover,

$$\frac{\partial}{\partial\phi_k}\psi^{-1}(B)(x_t-\mu)=B^k\xi^{-1}(x_t-\mu)=B^k\phi^{-1}v_t$$
$$\frac{\partial}{\partial\xi_k}\psi^{-1}(B)(x_t-\mu)=-\phi\xi^{-2}B^k(x_t-\mu)=-B^k\xi^{-1}v_t$$

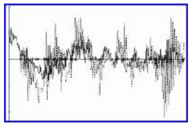
Thus $\quad \boldsymbol{\Lambda_{\theta,t}=\Lambda_F^{\mathrm{loc}}(V_t)\,\tilde H_{\theta,t}},\quad \boldsymbol{\tilde H'_{\theta,t}=(H'_{\theta,t};\nu)},\ \boldsymbol{\nu=\frac{\phi(1)}{\xi(1)}}$

where $\quad \boldsymbol{H'_{\theta,t}=\big(-B\phi^{-1},\ldots,-B^p\phi^{-1};B\xi^{-1},\ldots,B^q\xi^{-1}\big)V_t}$.

Fisher information at $\theta$ is

$$\boldsymbol{\mathcal{I}_\theta=\mathcal{I}_F^{\mathrm{loc}}\begin{pmatrix}\mathcal{K}_{\theta,F}&0\\0&\nu^2\end{pmatrix},\qquad \mathcal{K}_{\theta,F}=\mathrm{Cov}_\theta\,H_{\theta,1}}$$





**$\boldsymbol{M}$-estimates:** $\sum_{t=p+1}^{T} \varrho(v_t) = \min_\theta !$, resp. $\sum_t^T \psi(v_t)\frac{\partial v_t}{\partial\theta} = 0$
where $\psi = \dot\varrho$ and $v_t(\theta) = \xi^{-1}\phi(x_t - \mu)$ ($x_t := \mu$ for $t \le 0$):

$$\frac{\partial v_t}{\partial \mu} = -\frac{\phi(1)}{\xi(1)} = -\nu$$

$$\frac{\partial v_t}{\partial \xi} = -\xi^{-2}(B,\dots,B^p)'\phi(x_t-\mu) = -\xi^{-1}(B,\dots,B^p)'v_t$$

$$\frac{\partial v_t}{\partial \phi} = \xi^{-1}(B,\dots,B^q)'(x_t-\mu) = \phi^{-1}(B,\dots,B^q)'v_t$$

Thus the influence curve at $\theta$ is just $\boldsymbol{\eta_{\theta,t} = A_{\theta,t}\psi(V_t)\tilde H_{\theta,t}}$, where

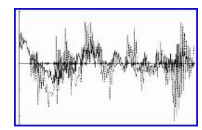
$$A_{\theta,t}^{-1} = \mathrm{E}_\theta\ \psi(V_t)\tilde H_{\theta,t}\Lambda'_{\theta,t} = (\mathrm{E}_F\ \psi\Lambda_F)\mathcal{I}_F^{-1}\mathcal{I}_\theta \qquad \text{(Fisher consistency)}$$

By the CLT for martingale differences (Billingsley[68]), under mild conditions on $\psi$, the $M$-estimate will under $P_\theta$ be asy. normal with mean zero and covariance

$$\mathrm{E}_\theta\ \eta_{\theta,t}\eta'_{\theta,t} = \mathcal{I}_F^{\mathrm{loc}}\frac{\mathrm{E}_F\ \psi^2}{(\mathrm{E}_F\ \psi\Lambda_F)^2}\mathcal{I}_\theta^{-1} = \frac{\mathrm{E}_F\ \psi^2}{(\mathrm{E}_F\ \psi\Lambda_F)^2}\begin{pmatrix}\mathcal{K}_{\theta,F} & 0\\ 0 & \nu^2\end{pmatrix}^{-1}$$

In front of $\mathcal{I}_\theta^{-1}$ appears the relative variance of the $M$-estimate of location based on $\psi$: it is $\ge 1$ (Cramèr–Rao), $= 1$ iff $\psi \propto \Lambda_F$ (MLE).





**LSE:** $\psi(v) = v$, $\mathrm{E}_F\, \psi \Lambda_F = -1$; asy. covariance

$$\mathcal{I}_F^{\mathrm{loc}} \sigma_F^2\, \mathcal{I}_\theta^{-1} = \sigma_F^2 \begin{pmatrix} \mathcal{K}_{\theta,F} & 0 \\ 0 & \nu^2 \end{pmatrix}^{-1}$$

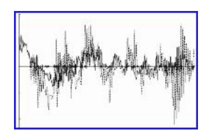
The relative variance $\mathcal{I}_F^{\mathrm{loc}} \sigma_F^2$ may become arbitrarily large if $F$ varies arbitrarily little—in total variation, even under symmetry and moment constraints—resulting in an arbitrarily small ARE near a normal $F$ (where ARE=1).

However, $\sigma_F^2$ cancels out in $\sigma_F^2\, \mathcal{K}_{\theta,F}^{-1}$ since this is $= \mathrm{Cov}_\theta^{-1}$ of

$$\frac{H_{\theta,t}}{\sigma_F} = \left(-B\phi^{-1}, \ldots, -B^p\phi^{-1}; B\xi^{-1}, \ldots, -B^q\xi^{-1}\right)' \frac{V_t}{\sigma_F}$$

Therefore, subject to $\mu_F = 0$ and $\sigma_F^2 < \infty$, LSE is asy. distribution free for $(\phi, \xi)$; but, in view of the entry $\sigma_F^2/\nu^2$, not so for $\mu$; Whittle[52]. Innovation outliers at time $t$ provide leverage points at time $t+1$ for the estimation of $(\phi, \xi)$.





$M$**-estimates with estimated scale:** $\sum_t^T \varrho(\frac{v_t(\theta)}{\tilde\sigma}) = \min_\theta !$ employing a consistent estimator $\tilde\sigma$ of the innovation scale $\sigma_F$.
Then the asy. covariance is:

$$\frac{\mathrm{E}_F\,\psi_{\sigma_F}^2}{\left(\mathrm{E}_F\,\psi_{\sigma_F}\Lambda_F\right)^2}\begin{pmatrix}\mathcal{K}_{\theta,F} & 0\\ 0 & \nu^2\end{pmatrix}^{-1} \qquad \text{where} \quad \psi_{\sigma_F}(v) = \psi(\frac{v}{\sigma_F}).$$

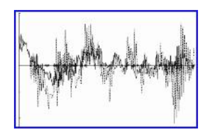
But, with $F_1 = \mathcal{L}_F(\frac{V_j}{\sigma_F})$, since $\Lambda_F(v) = \sigma_F^{-1}\Lambda_{F_1}(\frac{v}{\sigma_F})$,

$$\frac{\mathrm{E}_F\,\psi_{\sigma_F}^2}{\left(\mathrm{E}_F\,\psi_{\sigma_F}\Lambda_F\right)^2} = \frac{\mathrm{E}_{F_1}\,\psi^2}{\left(\mathrm{E}_{F_1}\,\psi\Lambda_{F_1}\right)^2}\,\sigma_F^2$$

Thus—subject to a given $F_1$—$M$-estimators enjoy the same asy. scale invariance of their $(\phi,\xi)$-component w.r.t. rescaling of $F_1$ by $\sigma_F$ (↔Dutter/Stockinger[87,p29])!

**Outlier Models** Martin/Yohai[85]
IO heavy-tailed innovation law $F$ (symmetric)
AO $Y_t = X_t + Z_t W_t$, SO $Y_t = (1-Z_t)X_t + Z_t W_t$
where the i.i.d. 0–1 switching process $Z_t$, the ideal process $X_t$, and the contaminating process $W_t$ are sto. independent (SO); resp. $(Z_t)$ and $(X_t, W_t)$ sto. independent (AO).
Arcs $\mu_\gamma = (1-\gamma)\mu_{\mathrm{id}} + \mu_{\mathrm{cont}}$ (implicit in IC-definition below)





**Influence Curves as Gateaux–Derivatives**

Martin/Yohai[85,86], Künsch[84]

$M$-estimators: $\sum_t^T \psi_t(y_t,\ldots,y_1;\hat\theta) = 0$, where "$\psi_t \to$ some $\psi$".

$M$-functionals: $\int \psi\big(y; M(\mu)\big)\, d\mu = 0$

IC(Hampel) at $\mu_{\mathrm{id}}$ (stationary, ergodic process–law):

$$\mathrm{ICH}(M,\mu_{\mathrm{id}};y) = \lim_{\gamma\downarrow 0} \frac{M(\mu_\gamma) - M(\mu_{\mathrm{id}})}{\gamma}$$

employing $\mu_{\mathrm{cont}} = 1_y$, Dirac in any $y \in \mathbb{R}^\infty$, "if limit exists", resp.

$$M(\mu_\gamma) = M(\mu_{\mathrm{id}}) + \gamma \int \mathrm{IC}(M,\mu_{\mathrm{id}};y)\,\mu_{\mathrm{cont}}(dy) + \mathrm{o}(\gamma)$$

employing $\mu_{\mathrm{cont}}$, too, the law of any stationary and ergodic process.

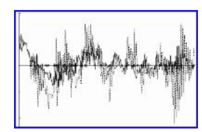
**Unpleasant Aspects** of the elementary approach:

(1) $\mathrm{ICH} \neq \mathrm{IC}$;

(2) $\sup_y \big|\mathrm{IC(H)}(M,\mu_{\mathrm{id}};y)\big| > \sup_{\mu_{\mathrm{cont}}} \big|\int \mathrm{IC}(M,\mu_{\mathrm{id}};y)\,\mu_{\mathrm{cont}}(dy)\big|$;
the second sup extending over all stationary and ergodic laws $\mu_{\mathrm{cont}}$;

(3) nonuniqueness of IC.





**and Restrictions** (Künsch[84])

(4) to $\mathrm{AR}(p)$-models, and to functions

(5) $\psi(x_m, \ldots, x_1)$, $\mathrm{IC}(x_m, \ldots, x_1)$ of a fixed number $m > p$ of variables,

(6) conditional centering (E.0) below only formal (uniqueness).

(7) Arbitrary bound $b_\theta$ for the minimum-trace problem:

$$\int |\mathrm{IC}|^2 \, d\mu_\theta^{(m)} = \min ! \qquad \text{s.t. } |\mathrm{IC}| \le b_\theta$$

and

$$\int \mathrm{IC} \; \mu_\theta^{(m)}(dx_m | x_{m-1}, \ldots, x_1) = 0 \qquad \text{(E.0)}$$

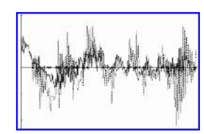
$$\int \mathrm{IC} \, d\mu_\theta^{(m)} = \mathrm{I}_p \quad \text{(Fisher consistency)}$$

with scores $\Lambda_\theta^{(m)} = \sum_{j=p+1}^{m} \Lambda_F(v_j)(x_{j-1}, \ldots, x_{j-p})'$. Optimal IC

$$\mathrm{IC}^* = A(\theta)\Lambda_\theta^{(m)} \min\left\{1, \frac{b}{|A\Lambda_\theta^{(m)}|}\right\} \qquad \text{(Hampel–Krasker)}$$

provided there exists $A = A_{p\times p}(\theta)$ such that $\mathrm{IC}^*$ has conditional expectation zero (not taken up by some function $a_\theta$ of the past)
and provided that it is Fisher consistent.





# 2 LAN Time Series Models

- LAN, $L_2$-differentiability (w.r.t. $k$-dim. parameter):

$$\log dP^{(n)}_{\theta+h/\sqrt{n}}/dP^{(n)}_{\theta} = \frac{h'}{\sqrt{n}}\sum_{j=1}^{n}\Lambda_{\theta,n,j}(x_{1:j}) - \frac{1}{2}h'\mathcal{I}_\theta h + \mathrm{o}_{P_\theta}(n^0)$$

  where $P^{(n)}_\theta = \mathcal{L}_\theta(X_{1:n})$, $X_{1:n} = (X_n, X_{n-1}, \ldots X_1)$

- triangular arrays of scores, and stationary approximations

$$\boldsymbol{\Lambda_{\theta,n,j} \approx \Lambda_{\theta,j} = \Lambda_f(V_j)H_\theta(\bar{X}_{j-1})} \qquad \text{(product form)}$$

  with $V_j$ the innovation at time $j$, $(V_j)$ i.i.d. $\sim F$; the extended past $\boldsymbol{\bar{X}_{j-1} = (X_{j-1}, X_{j-2}, \ldots X_1, X_0, \ldots)}$, sto. indep. of $\boldsymbol{V_j}$

- $(\Lambda_{\theta,j})$ is a stationary, ergodic MD sequence, under $P_\theta$. In our examples, the process $(X_t)_{t\in\mathbb{Z}}$ itself is stationary, ergodic, $\forall\, P_\theta$.

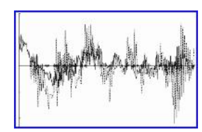
- location/scale models with finite Fisher info. $\mathcal{I}_f = \int \Lambda_f^2\, dF$:

$$\boldsymbol{\Lambda_f^{\mathrm{loc}}(v) = -[\dot{f}/f](v), \qquad \Lambda_f^{\mathrm{scal}}(v) = v\Lambda_f^{\mathrm{loc}}(v) - 1}$$

  Fisher information of the model of process laws $\mathcal{L}_\theta\{(X_t)_{t\in\mathbb{Z}}\}$ is

$$\boldsymbol{\mathcal{I}_\theta = \mathrm{E}_\theta[\Lambda_{\theta,1}\Lambda'_{\theta,1}] = \mathcal{I}_f\,\mathcal{K}_{\theta,F}, \quad \mathcal{K}_\theta = \mathrm{E}_\theta[H_{\theta,1}H'_{\theta,1}]}$$





# List of Models

**References** (more general, incomplete): Jeganathan[82,88], Kreiss[84], Wang[94], Hallin et al.[95,99] (m-variate ARMA, FARIMA), Koul/Schick[96,97], Basawa/Hwang[97] (bilinear), Drost/Klaassen/Werker[97], Drost/Klaassen[98], Newey[90], Robinson[88], . . .
**Aim** (mostly): Adaptive (w.r.t. unknown innovation density), fully efficient estimation

- ARMA [Kreiss[84], LAN independently by Staab[84] (weaker conds., incl. regression trend)]

- TAR [Tong/Lim[80], Chan et al.[85]]

$$X_t = \sum_{j=1}^{k} (\mu_j + \rho_j X_{t-1})\, \mathrm{I}_{A_j}(X_{t-1}) + V_t,$$

$$A_1, \ldots, A_k \text{ a partition of } \mathbb{R}$$

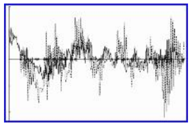
Chan's suff. conds. for stationarity, ergodicity

$$\theta' = (\rho_1, \ldots, \rho_k; \mu_1, \ldots, \mu_k)$$

$$\Lambda_{\theta,t} = \Lambda_f^{\mathrm{loc}}(V_t) H_{\theta,t}$$

$$H'_{\theta,t} = (\ldots, X_{t-1}\, \mathrm{I}_{A_j}(X_{t-1}), \ldots; \ldots, \mathrm{I}_{A_j}(X_{t-1}), \ldots)$$





- ARCH [Engle[82], Nelson[90]]

$$
\begin{aligned}
X_t &= \sigma\big(1+\alpha_1 X_{t-1}^2+\ldots+\alpha_p X_{t-p}^2\big)^{1/2}\,V_t\\
\theta' &= (\alpha_1,\ldots,\alpha_p,\sigma)\\
0 &> \mathrm{E}\{\log V_t^2+\log\sigma^2+\log\max\alpha_j\}\quad \text{(statio., ergod.)}\\
\Lambda_{\theta,t} &= \Lambda_f^{\mathrm{scal}}(V_t)H_{\theta,t}\\
H'_{\theta,t} &= \Big(\ldots,\frac{X_{t-j}^2}{2(1+\alpha_1X_{t-1}^2+\ldots+\alpha_pX_{t-p}^2)},\ldots;\sigma^{-1}\Big)
\end{aligned}
$$

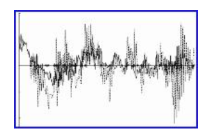
- GARCH [Bollerslev[86], Nelson[91], Drost/Klaassen[98]]

$$
\begin{aligned}
X_t &= \sqrt{h_t}\,\xi_t,\quad \xi_t=\sigma V_t \qquad\qquad \text{GARCH(1,1)}\\
h_t &= 1+\beta h_{t-1}+\alpha X_{t-1}^2 = 1+h_{t-1}(\beta+\alpha\xi_{t-1}^2)\\
\theta' &= (\alpha,\beta,\sigma);\quad \mathrm{E}\,\log(\beta+\alpha\xi^2)<0\quad \text{(statio., ergod.)}
\end{aligned}
$$

$$
\Lambda_{\theta,t}=\Lambda_F^{\mathrm{scal}}\tilde H_{\theta,t},\qquad \tilde H_{\theta,t}=\begin{pmatrix}\frac12 h_t^{-1}(\theta)H_t(\theta)\\ \sigma^{-1}\end{pmatrix}
$$

$$
H_t(\theta)=\sum_{i\ge0}\beta^i h_{t-1-i}(\theta)\begin{pmatrix}\xi_{t-1-i}^2\\ 1\end{pmatrix},\qquad h_t(\theta)=\sum_{i\ge0}\prod_{k=1}^i(\beta+\alpha\xi_{t-k}^2)
$$





# 3 Influence Curves and Asy. Linear Estimators — w.r.t. Ideal Model

- **Asymptotic linear estimators** of $D\theta$ (some $D = D_{p\times k}$)

$$\sqrt{n}(S_n - D\theta) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \psi_\theta(\bar{x}_j) + \mathrm{o}_{P_\theta}(n^0)$$

- **Influence curves**

$$\begin{array}{llrcl} \psi_\theta \in L_2^p(P_\theta) & \text{(i)} & \mathrm{E}_\theta[\psi_\theta(\bar{x}_1)\big|\bar{x}_0\,] & = & 0 \\ \text{Fisher consistency} & \text{(ii)} & \mathrm{E}_\theta[\psi_{\theta,1}\Lambda'_{\theta,1}] & = & D \end{array}$$

**Rem:** (i) will also be enforced by optimization criterion.

With ideal $(X_t)_{t\in\mathbb{Z}}$, the MDS $\psi_{\theta,t} = \psi_\theta(\bar{X}_t)$ is stationary, ergodic.

$\boldsymbol{\psi_\theta(\bar{x}_j)}$ **=influence of** $\boldsymbol{X_j}\big|$**past** $\boldsymbol{\bar{X}_{j-1} = (X_{j-1},\ldots,X_0,\ldots)}$

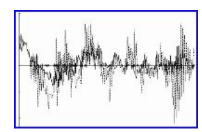
- **CLT** (Billingsley[68], McLeish[74])

$$\mathcal{L}_{P_\theta}\sqrt{n}(S_n - D\theta) \longrightarrow \mathcal{N}(0, \mathrm{Cov}_\theta(\psi_{\theta,1})), \quad \mathrm{Cov}_\theta(\psi_{\theta,1}) = \mathrm{E}_\theta[\psi_{\theta,1}\psi'_{\theta,1}]$$

- **Fisher consistency** equivalent to **Hàjek–regularity**

$$\mathcal{L}_{P_{\theta+h/\sqrt{n}}}\sqrt{n}(S_n - D\theta) \longrightarrow \mathcal{N}(Dh, \mathrm{Cov}_\theta(\psi_{\theta,1}))$$


*Helmut Rieder*

- **Cramér–Rao bound**

$$\mathrm{Cov}_\theta(\psi_{\theta,1}) \succeq \ D\mathcal{I}_\theta^{-1}D' \ = \mathrm{Cov}_\theta(\hat{\psi}_{\theta,1})$$

achieved iff $\psi_{\theta,1} = \hat{\psi}_{\theta,1} = D\mathcal{I}_\theta^{-1}\Lambda_{\theta,1} =$

$$= D\mathcal{I}_F^{-1}\Lambda_f(V_1)[\mathrm{Cov}_\theta\, H_{\theta,1}]^{-1}H_{\theta,1}$$

- **Examples:** LSE, M(LE), GM, RA, TRA; Bustos, Fraiman, Martin, Yohai[82-84]

# 4 Nbds of Transition Probabilities

- joint distribution of $\bar{X}_n = (X_n, X_{n-1}, \ldots X_1, X_0, X_{-1} \ldots)$

$$Q^{(n)}(dx_n, \ldots, \bar{x}_0) = \prod_{j=1}^n Q^{(n,j)}(dx_j|\bar{x}_{j-1})Q^{(n,0)}(d\bar{x}_0)$$

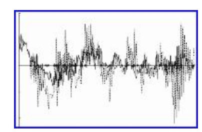
- **Full neighborhoods**, at time $n$ with **radius curves** $\boldsymbol{\varepsilon^{(n,j)}}$

$$\begin{aligned} \mathrm{dist}(Q^{(n,j)}, P_\theta^{(n,j)}) &\leq \frac{r}{\sqrt{n}}\varepsilon^{(n,j)}(\bar{x}_{j-1}); \quad && j = 1, \ldots, n \\ \mathrm{dist}(Q^{(n,0)}, P_\theta^{(n,0)}) &\leq \frac{r}{\sqrt{n}}\varepsilon^{(n,0)}; && \varepsilon^{(n,0)} = \mathrm{const} \end{aligned}$$

- Neighborhood **submodel**: $Q_n^{(n,0)} = P_\theta^{(n,0)}$ and

$$\boldsymbol{Q_n^{(n,j)}(dx_j|\bar{x}_{j-1}) = (1 + \tfrac{r}{\sqrt{n}}q(\bar{x}_j)P_\theta^{(n,j)}(dx_j|\bar{x}_{j-1})}$$





where $\boldsymbol{q} \in \boldsymbol{L_2(P_\theta)}$, $\mathbf{E}_\theta[\boldsymbol{q}(\bar{\boldsymbol{x}}_1)|\bar{\boldsymbol{x}}_0] = \mathbf{0}$ and, matching the nbd–type,

$$\begin{array}{lrcl} \textbf{(c)} & \boldsymbol{q_1 = q(\bar{x}_1)} & \geq & \boldsymbol{-\varepsilon(\bar{x}_0)} \\ \textbf{(v)} & \mathbf{E}_\theta[|\boldsymbol{q_1}|\,|\boldsymbol{\bar{x}_0}] & \leq & \mathbf{2}\,\boldsymbol{\varepsilon(\bar{x}_0)} \\ \textbf{(h)} & \mathbf{E}_\theta[\boldsymbol{q_1^2}\,|\boldsymbol{\bar{x}_0}] & \leq & \mathbf{8}\,\boldsymbol{\varepsilon^2(\bar{x}_0)} \end{array}$$

with radius curve $\boldsymbol{\varepsilon^{(n,j)} = \varepsilon}$ $(j = 1, \dots, n)$

- Analogy to error–free–variables, or conditional, neighborhoods in regression (here past $\hat{=}$ regressor); Huber[82], Bickel[84], Ri[94]
- **Generalized IO's**: perturbation of trans.pr. depend on the past; to include $Q^{(n)}$(AO,SO), let $\varepsilon^{(n,j)}(\bar{x}_{j-1}) = \sqrt{n}\,\mathrm{dist}(Q^{(n,j)}, P_\theta^{(n,j)})$
- LAN: $\log dQ_n^{(n)}/dP_\theta^{(n)} = \frac{r}{\sqrt{n}} \sum_1^n q(\bar{x}_j) - \frac{r^2}{2}\,\mathrm{E}_\theta\, q_1^2 + \mathrm{o}_{P_\theta^{(n)}}(n^0)$

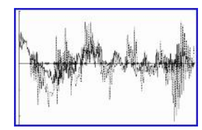
  CLT: $\mathcal{L}_{Q_n^{(n)}}\, \sqrt{n}(S_n - D\theta) \longrightarrow \mathcal{N}\big(\mathrm{E}_\theta\, \psi_{\theta,1} q_1, \mathrm{Cov}_\theta(\psi_{\theta,1})\big)$

# 5 Asymptotic Maximum MSE

$$\underset{\boldsymbol{\theta}}{\mathbf{MSE}}(\boldsymbol{\psi_{\theta,1}}) = \mathbf{tr}\,\mathbf{Cov}_{\boldsymbol{\theta}}(\boldsymbol{\psi_{\theta,1}}) + \boldsymbol{r^2}|\boldsymbol{b_{*,\varepsilon}}(\boldsymbol{\psi_{\theta,1}})|^{\mathbf{2}}$$

where $* = c, v, h$ (contamination, total varation, Hellinger). We fix $\theta$.





**Proposition 6.1 (bias evaluation)** Ri[94, §7.3])

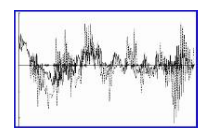
- standardized max. bias terms for **fixed radius curve $\varepsilon$**

$$
\begin{array}{llll}
\textbf{(c)/(v)} & b_{c,\varepsilon}(\psi) & = (\mathrm{E}\,\varepsilon \sup_\bullet \psi) \vee (-\,\mathrm{E}\,\varepsilon \inf_\bullet \psi) & k=1 \\
 & & \approx \mathrm{E}\,\varepsilon \sup_\bullet |\psi| & k>1 \\
 & & = \mathrm{E}\,\varepsilon(\bar x_0) \sup_{x_1} |\psi(x_1,\bar x_0)| & \\
 & b_{v,\varepsilon}(\psi) & = \mathrm{E}\,\varepsilon\,[\sup_\bullet \psi - \inf_\bullet \psi] & k=1 \\
 & & \approx 2\, b_{c,\varepsilon}(\psi) & k>1 \\
\textbf{(h)} & b_{h,\varepsilon}(\psi) & = \sqrt{8}\ \mathrm{E}\Big[\varepsilon(\bar x_0)\, \mathrm{E}\big[|\psi(x_1,\bar x_0)|^2 \big|\bar x_0\big]^{\frac12}\Big] &
\end{array}
$$

- standardized bias terms for **varying $\boldsymbol{\varepsilon}$ s.t. $\|\boldsymbol{\varepsilon}\|_{\boldsymbol{\alpha}} \le 1$**
  $\boldsymbol{\alpha = 1 : \mathrm{E}\,\varepsilon \le 1;} \qquad \boldsymbol{\alpha = 2 : \mathrm{E}\,\varepsilon^2 \le 1} \qquad \boldsymbol{\alpha = \infty : \varepsilon \le 1 \equiv \varepsilon_1}$

$$
\begin{array}{lllll}
\textbf{(c)/(v)} & b_{c,1}(\psi) & = & \sup|\psi| & \\
 & b^2_{c,2}(\psi) & = & \mathrm{E}\sup_{x_1}|\psi(x_1,\bar x_0)|^2 & \\
 & b_{v,\alpha}(\psi) & \approx & 2\, b_{c,\alpha}(\psi) & \alpha = 1,2 \\
\textbf{(h)} & b^2_{h,1}(\psi) & = & 8\ \sup_{|e|=1} \sup_\bullet \mathrm{E}_\bullet (e'\psi)^2 & \\
 & & \approx & 8\ \sup_{\bar x_0} \mathrm{E}\big[|\psi|^2\big|\bar x_0\big] & \\
 & b^2_{h,2}(\psi) & = & 8\ \max\mathrm{ev}\,\mathrm{Cov}_\theta(\psi_{\theta,1}) &
\end{array}
$$





# 6 Optimally Robust ICs

Ri[94, §7.4]

**Theorem 7.1 (MSE–solutions)** $\Lambda := \Lambda_f(V_1)H(\bar{X}_0)$ from LAN

$\boxed{\boldsymbol{* = c}}$ $\quad \boldsymbol{\tilde{\psi}_{c,\alpha} = AH(\Lambda_f - \vartheta(\bar{X}_0)) \min\{1, \frac{r_\alpha(\bar{X}_0)}{|\Lambda_f - \vartheta(\bar{X}_0)|}\}}$

$\boldsymbol{\alpha = \varepsilon}$**:**

**(i)** $0 = \mathrm{E}_\bullet(\Lambda_f - \vartheta) \min\{1, \frac{r_\varepsilon}{|\Lambda_f - \vartheta|}\}$

**(ii)** $D = A\,\mathrm{E}\,HH'\,\mathrm{E}_\bullet(\Lambda_f - \vartheta)^2 \min\{1, \frac{r_\varepsilon}{|\Lambda_f - \vartheta|}\},$

**(iii)** $\mathrm{E}_\bullet(|\Lambda_f - \vartheta| - r_\varepsilon)_+ = \frac{r^2\varepsilon(\bar{x}_0)}{|AH|}\,\mathrm{E}[|AH|r_\varepsilon],$

$r_\varepsilon = 0$ if $\mathrm{E}_\bullet[|\Lambda_f - \vartheta|] < \frac{r^2\varepsilon(\bar{x}_0)}{|AH|}\,\mathrm{E}[|AH|r_\varepsilon]$

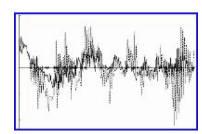
$\boldsymbol{\alpha = 1}$**:** $r_1(\bar{x}_0) = \frac{b}{|AH|}$ [Hampel–Krasker]

**(i)** $0 = \mathrm{E}_\bullet(\Lambda_f - \vartheta) \min\{1, \frac{r_1}{|\Lambda_f - \vartheta|}\}, \qquad \vartheta = \vartheta(H)$

**(ii)** $D = A\,\mathrm{E}\,HH'\,\mathrm{E}_\bullet(\Lambda_f - \vartheta)^2 \min\{1, \frac{r_1}{|\Lambda_f - \vartheta|}\},$

**(iii)** $r^2 b = \mathrm{E}(|AH||\Lambda_f - \vartheta| - b)_+$

least favorable $\quad \tilde{\varepsilon}_{c,1} \propto |AH|\,\mathrm{E}_\bullet(|\Lambda_f - \vartheta| - \frac{b}{|AH|})_+$



$\boldsymbol{\alpha} = \mathbf{2}$: $r_2(\bar{x}_0) \equiv d$ (const) [Huber]

**(i)** $0 = \mathrm{E}(\Lambda_f - a)\min\{1, \frac{d}{|\Lambda_f - a|}\}, \quad \vartheta \equiv a$ (const)

**(ii)** $A = \rho D \mathcal{K}^{-1},\ \mathcal{K} = \mathrm{E}\, HH',\ \rho = 1/\,\mathrm{E}(\Lambda_f - a)^2 \min\{1, \frac{d}{|\Lambda_f - a|}\}$

**(iii)** $r^2 d = \mathrm{E}(|\Lambda_f - a| - d)_+$

least favorable $\tilde{\varepsilon}_{c,2} \propto |\mathcal{K}^{-1} H|$

$\boxed{* = h}$ $\quad \tilde{\boldsymbol{\psi}}_{\boldsymbol{h},\boldsymbol{\alpha}} = \boldsymbol{A H \Lambda_f h_\alpha(\bar{X}_0)}$ (weighted min$L_2$)

$\boldsymbol{\alpha} = \boldsymbol{\varepsilon}$: $h_\varepsilon(\bar{x}_0) = \bigl(1 - \beta\frac{\varepsilon(\bar{x}_0)}{|AH|}\bigr)_+$

$$D = \mathcal{I}_f A\, \mathrm{E}\, HH' h_\varepsilon(\bar{x}_0), \quad \beta = 8\, r^2\, \mathrm{E}\bigl[\varepsilon(\bar{x}_0)(|AH| - \beta\varepsilon(\bar{x}_0))_+\bigr]$$

$\boldsymbol{\alpha} = \mathbf{1}$: $h_1 = \min\{1, \frac{b}{|AH|}\}$

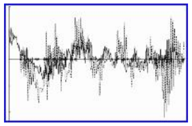
$$D = \mathcal{I}_f A\, \mathrm{E}\, HH' h_1, \quad 8\, r^2 b = \mathrm{E}(|AH| - b)_+$$

least favorable $\tilde{\varepsilon}_{h,1} \propto (|AH| - b)_+$

$\boldsymbol{\alpha} = \mathbf{2}$: $h_2(\bar{x}_0) = \mathrm{const}$ [MLE]

least favorable $\tilde{\varepsilon}_{h,2} = \tilde{\varepsilon}_{c,2} \propto |\mathcal{K}^{-1} H|$





Interpretation of Lagrange-multiplier matrix $A$ (which is to ensure Fisher consistency) in terms of statistical risk by the following extension of the classical Cramér–Rao bound
(in which $A = \mathcal{I}^{-1}$, and squared error loss $\rightsquigarrow$ trace).

**Proposition 7.2** In all models—except $* = h$, $\alpha = 2$:

$$\mathbf{MSE}(\tilde{\boldsymbol{\psi}}_{*,\boldsymbol{\alpha}}) = \mathbf{tr}\,\boldsymbol{A}.$$

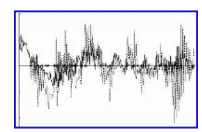
---

The following lemma (for $\gamma(v) = v^2$) provides the solution to the true MSE problems (that is, with bias squared) in Theorems 7.4.11(b), 7.4.12(b), 7.4.16(b), and 7.4.18(b) of Ri[94, §7.4].

**Lemma 7.3** For $X$ a real vector space, $A \subset X$ a convex subset, let $f, g \colon A \to \mathbb{R}$ and $\gamma \colon [0, \infty) \to \mathbb{R}$ be three convex functions, where $g, \gamma \geq 0$, and $\gamma$ increasing. Let $\beta_0 \in [0, \infty)$ and $z_0 \in A$. Assume $\gamma$ is differentiable at $g(z_0)$, and put $\boldsymbol{\beta_1 = \beta_0\, \gamma'(g(z_0))}$.

Then $z_0$ minimizes the Lagrangian $\boldsymbol{L_0 = f + \beta_0\,(\gamma \circ g)}$ over $A$ iff $z_0$ minimizes the Lagrangian $\boldsymbol{L_1 = f + \beta_1\, g}$ over $A$.





# 6.1 Robust Adaptivity

**Minmax MSE problem:** $\mathbf{E}\,|\boldsymbol{\eta}|^2 + \boldsymbol{r}^2\,\boldsymbol{B}^2(\boldsymbol{\eta}) = \mathbf{min}\,!$ ($B$ :=maxbias)

**1.** subject to $\eta \in L_2^k(P)$, $\mathrm{E}\,\eta = 0$ (or condl. $\mathrm{E}_\bullet\,\eta = 0$), and $\mathrm{E}\,\eta\Lambda^\tau = \mathbb{I}_k$,

**2.** s.t. same side conditions but, in addition, $\mathbf{E}\,\boldsymbol{\eta g} = \mathbf{0}\ \forall\,\boldsymbol{g} \in \boldsymbol{\mathcal{G}_2}$

($\mathcal{G}_2$ :=tangent set belonging to the nuisance parameter).

Then robust adaptivity says: **minmaxMSE1 = minmaxMSE2** which happens iff solution to 1st problem already is $\perp\,\mathcal{G}_2$ (ext. Stein).

**Example:** $\mathrm{ARMA}(p,q)$: $\phi(B)(X_t - \mu) = \xi(B)V_t$

**i.** Robust estimation of $\boldsymbol{\mu}$ with nuisance parameter $(\phi,\xi)$ is adaptive.

**ii.** Classical adaptivity of the estimation of $(\boldsymbol{\phi},\boldsymbol{\xi})$ with nuisance $\mu$ extends to robust estimation in case $* = c, \alpha = 2$ [Huber]. In case $* = c, \alpha = 1$ [Hampel–Krasker], adaptivity holds if $F$ is symmetric.

In fact, $\mathrm{E}\,H = 0$ and $H$, $\Lambda_f(V)$ are stochastically independent, so

$$\mathrm{E}\,AH(\Lambda_f - a)^2 \min\{1, \tfrac{d}{|\Lambda_f - a|}\} = 0 \qquad (\alpha = 2)$$

But

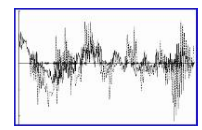
$$\mathrm{E}\,AH(\Lambda_f - \vartheta(H))^2 \min\{1, \tfrac{b/|AH|}{|\Lambda_f - \vartheta(H)|}\} = 0 \qquad (\alpha = 1)$$

where $\vartheta(H) = \vartheta(-H)$, requires $\mathcal{L}_F(H)$, resp. $F$, to be symmetric.
For asymmetric $F$, nonadaptivity remains to be evaluated numerically.



*Helmut Rieder*



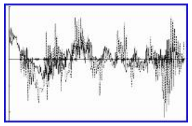


## Robust Nonadaptivity in Linear Regression numerical results

for $F = \mathcal{N}(0,1)$ (sym.), $K = 0.75\,\mathrm{I}_{\{-1\}} + 0.25\,\mathrm{I}_{\{3\}}$ (asym.)

| $r$ | MSE ( $\mu$ known) | MSE ( $\mu$ unknown) | relMSE |
|---|---|---|---|
| 0.0 | 0.333 | 0.333 | 1.000 |
| 0.1 | 0.371 | 0.373 | 1.005 |
| 0.25 | 0.471 | 0.492 | 1.045 |
| 0.5 | 0.684 | 0.824 | 1.205 |
| 0.75 | 0.956 | 1.333 | 1.394 |
| 1.0 | 1.297 | 2.030 | 1.565 |
| 1.25 | 1.714 | 2.919 | 1.703 |
| 1.5 | 2.210 | 4.002 | 1.811 |
| 1.75 | 2.790 | 5.280 | 1.892 |
| 2.0 | 3.453 | 6.754 | 1.956 |
| 2.5 | 5.036 | 10.290 | 2.043 |
| 5.0 | 18.144 | 39.745 | 2.191 |





## 6.2 Estimation of location $\mu_X$ in ARMA

- $D = (0, \ldots, 0, 1)$, $\Lambda'_{\theta,1} = \Lambda_f(v_1)(H_{\theta,1}, \nu)'$, $\mathcal{I}_\theta = \mathcal{I}_f \left( \begin{smallmatrix} \sigma_F^2 \mathcal{K}, & 0 \\ 0, & \nu^2 \end{smallmatrix} \right)$
- $\hat{\psi}_{\theta,1} = D\mathcal{I}_\theta^{-1} \Lambda_{\theta,1} = (\nu \mathcal{I}_f)^{-1} \Lambda_f$ classical
- $\boldsymbol{\alpha = \infty}$: $\varepsilon_1 \equiv 1$; $* = v$ $(* = c, h$ likewise)
  - $\tilde{\psi}_{\theta,1} = c_1 \vee \kappa \Lambda_f \wedge c_2$ robust
  - $\mathrm{E}[\tilde{\psi}_{\theta,1} \nu \Lambda_f] = 1$
  - $\mathrm{E}(\kappa \Lambda_f - c_2)_+ = \mathrm{E}(c_1 - \kappa \Lambda_f)_+ = r^2 (c_2 - c_1)$
- $\boldsymbol{\alpha = 1}$: since $\mathrm{E}\,\varepsilon(c_2 - c_1) \le c_2 - c_1$ for $\mathrm{E}\,\varepsilon \le 1$, the constant contamination $\boldsymbol{\varepsilon_1 \equiv 1}$ of transition distributions turns out **least favorable** among all $\varepsilon$ s.t. $\mathrm{E}\,\varepsilon \le 1$, when estimating $\mu$.

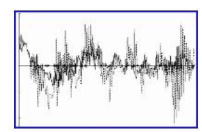
## 6.3 Deviation from independence in ARMA(0,0)≡i.i.d. location

Same minmax asy. MSE whether errors under distortion are kept i.i.d. or distortion of ideal transition probability (i.e., marginal) is allowed as defined (depending on the past, introducing dependence).

Compare estimation over Hellinger balls ($* = h$, $\alpha = 2$, conditional as well as errors-in-variables), where classical scores is optimal, but classical risk increases by bias $b_h^2 = 8 \max \mathrm{ev} \, \mathrm{Cov}$.

Robustness shifted (from IC) to construction problem.





# 7 Estimator Construction

**Algorithm** to determine the optimal ICs: one-parameter (monotony), multiparameter case (fixed point algorithm); $\mathrm{C}^{++}$, $\mathrm{S}^{+}$, R

**Finite-sample IC** (approximate) finite-sample inversion from observations to innnovations

**One-step M-estimates:** $\boldsymbol{S_n = \hat\theta_n + \frac{1}{n}\sum_{j=1}^{n}\psi_{n,\hat\theta_n}(\bar x_j)}$ based on suitable MDA approximations $\psi_{n,\theta}$ of $\psi_\theta$ and an initial estimator $\hat\theta_n$ which is $\sqrt{n}$-consistent, uniformly on shrinking nbds. Then

$$\sqrt{n}\,(S_n-\theta)=\sqrt{n}\,(\hat\theta_n-\theta)+\frac{1}{\sqrt{n}}\sum\psi_\theta(\bar x_j)+\sqrt{n}\,\frac{1}{n}\sum(\psi_{n,\theta_n}-\psi_\theta)(\bar x_j)$$

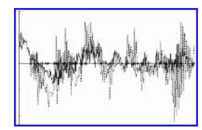
May use LeCam's discretization: $\sqrt{n}\,(\hat\theta_n-\theta)\to h$. Then

$$\frac{1}{n}\sum(\psi_{n,\hat\theta_n}-\psi_\theta)(\bar x_j)\approx\frac{1}{n}\sum\mathrm{E}_\theta^{n,j}(\psi_{n,\hat\theta_n}-\psi_\theta)$$
$$=-\frac{1}{n}\sum\int\psi_{n,\hat\theta_n}\,[P_{\hat\theta_n}^{n,j}-P_\theta^{n,j}](dx_j|\bar x_{j-1})\approx-\,\mathrm{E}_\theta\,\psi_{\theta,1}\Lambda_{\theta,1}\frac{h}{\sqrt{n}}=-\frac{h}{\sqrt{n}}$$

by $L_2$-diff. of trans. probs and the CLT; done by Staab[84] for ARMA.

**Minimum distance estimation** (not yet settled)
special results in literature: Koul[89], Beran[95], . . . (nonrobust);
need a general result on uniform tightness of the empirical process
(of estimated residuals), over infinitesimal transition nbds.





**Extension of a basic ufo. convergence result** Ri[94; Prop. 6.2.1] **from independence to dependence:**

Given filtrations $\mathcal{F}_{n,0} \subset \mathcal{F}_{n,1} \subset \ldots \subset \mathcal{F}_{n,n} = \mathcal{A}_n$ and measurable functions $\psi_{n,j} \colon (\Omega_n, \mathcal{F}_{n,j}) \to (\mathbb{R}^k, \mathbb{B}^k)$. For any pm. $Q_n$ on $\mathcal{A}_n$, denote by $Q_n^{n,j}(F|\mathcal{F}_{n,j-1})$, $F \in \mathcal{F}_{n,j}$, the (regular) transition probability from $\mathcal{F}_{n,j-1}$ to $\mathcal{F}_{n,j}$, and by $\mathrm{E}_{Q_n}^{n,j}$, $\mathrm{Cov}_{Q_n}^{n,j}$ the corresponding conditional expectation, resp. condl. covariance.

**Proposition 7.1 (ufo. weak LLN and CLT for MDA)**

Let $\max_j \sup_{\Omega_n} |\psi_{n,j}| = \sqrt[4]{n}$, and consider pm's $Q_n$, $P_{\theta,n}$ on $\mathcal{A}_n$ such that $\exists\, M < \infty \;\; \forall\, j = 1, \ldots, n \;\; \forall\, n$,

$$d_v\big(Q_n^{n,j}, P_{\theta,n}^{n,j}\big) \le M/\sqrt{n} \qquad \text{on } (\Omega_n, \mathcal{F}_{n,j-1})$$

(LLN) Then, $\frac{1}{n} \sum_{j=1}^n \big[\, \psi_{n,j} - \mathrm{E}_{P_{\theta,n}}^{n,j}\, \psi_{n,j} \,\big] = \mathrm{o}_{Q_n}(n^0)$

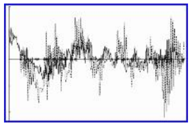
(CLT) If, moreover, $\exists$ some $k \times k$ matrix $C > 0$ such that

$$\frac{1}{n} \sum_{j=1}^n \mathrm{Cov}_{P_{\theta,n}}^{n,j}\, \psi_{n,j} = C + \mathrm{o}_{Q_n}(n^0) \qquad (\dagger)$$

then,
$$\mathcal{L}_{Q_n}\Big(\frac{1}{\sqrt{n}} \sum_{j=1}^n \big[\, \psi_{n,j} - \mathrm{E}_{Q_n}^{n,j}\, \psi_{n,j} \,\big]\Big) \longrightarrow \mathcal{N}(0, C)$$

$\boldsymbol{* = v,\ \alpha = 1}$**:** $\mathrm{E}_\theta\, \varepsilon^{n,j} \le 1$ and $\varepsilon^{n,j} \le$ some $M < \infty$.





# Failure of Construction

**Model AR(1):** $x_j = \theta x_{j-1} + v_j$, with i.i.d. $v_j \sim \mathcal{N}(0,1)$.

**Estimate:** $S_n = \theta + \frac{1}{n}\sum_{j=1}^n \eta_j(\bar{x}_j)$, one-step M based on exact initial estimate with $\eta_j$ of Hampel–Krasker form (w.l.o.g. $A = 1$)

$$\eta_j(\bar{x}_j) = (x_j - \theta x_{j-1})\, x_{j-1} \min\{1, \tfrac{b}{|(x_j - \theta x_{j-1})\, x_{j-1}|}\}$$

**Proposition 7.2** Let $0 < \theta < 1/2$ and $r \in (0,\infty)$. Then there is a sequence of measures $Q_n \in U_{c,\varepsilon_1}(P_\theta; r/\sqrt{n}\,)$ with $\varepsilon_1 \equiv 1$ such that

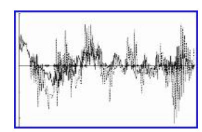
$$\mathrm{E}\,\sqrt{n}\,(S_n - \theta) \longrightarrow rb, \qquad \mathrm{Var}\,\sqrt{n}\,(S_n - \theta) \longrightarrow b^2$$

under $Q_n$, hence: $\mathrm{E}\, n(S_n - \theta)^2 \to b^2(1 + r^2) > \underset{\text{submod}}{\mathrm{MSE}}(\eta)$.

**Idea of Proof** Innovation $v_j$ at time $j = 1,\dots,n$ may be distorted so much (employing exponentially increasing contamination points) that, on an event $D_n$ with probability tending to $1$,
(i) $\eta_l(\bar{x}_l) = \pm b\ \forall\, l = j+1,\dots,n$, and (ii) $\eta_j(\bar{x}_j) = b$ (using $\operatorname{sign} x_{j-1}$).
(ii) generates the asy. bias: $\frac{1}{\sqrt{n}}\, n \frac{r}{\sqrt{n}}\, b = rb$, while the oscillation (i), for ideal $v_l$, accounts for the asy. variance $\frac{1}{n} n(1 - \frac{r}{\sqrt{n}})\, b^2 \to b^2$
(though MD-property is lost under contamination).

**Construction** for nbd-submodels defined by **$\boldsymbol{\rho}$-mixing**.





# 8 Unknown Neighborhood Radius

## 8.1 List of Ideal Models

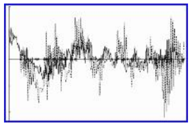
- Location: $y = \theta + u$, $u \sim \mathcal{N}_k(0, \mathbb{I}_k)$, $P_0 = \mathcal{N}_k(0, \mathbb{I}_k) = P$
- Scale ($k = 1$): $y = \sigma u$, $u \sim \mathcal{N}(0,1) = P_1 = P$
- Regression ($k \geq 1$): $y = x\,\theta + u$;  $x$, $u$ sto. indep.

$$u \sim \mathcal{N}(0,1)\ ,\ x \sim K(dx)$$
$$P = P_0(dx, du) = K(dx)\,\mathcal{N}(0,1)(du)$$

  For $\alpha = 2$, coincidence of results with 1-dim. location. For $\alpha = 1$, convergence of results to those for 1-dim. location as $k \to \infty$.

- ARMA($p, q$)-models (with shift) are covered, setting $K = \mathcal{L}(H)$. Ideal innovations i.i.d. $\sim \mathcal{N}(0,1)$, then $K$ multivariate normal. For $\alpha = 2$, coincidence of results with 1-dim. location. For $\alpha = 1$, convergence of results to those for 1-dim. location as $p + q \to \infty$.
- ARCH(1): $y_t = \sqrt{1 + \theta y_{t-1}^2}\, u_t$, $u_t$ i.i.d. $\sim \mathcal{N}(0,1)$
  For $\alpha = 2$, coincidence of results with 1-dim. scale.





## 8.2 Robust Neigborhoods About $P$

- (1-dim. location) symmetric contamination nbd of size $s \in [0,1)$ :

$$F = (1-s)\,\mathcal{N}(0,1) + s\,H, \quad H \text{ symmetric}$$

- $r/\sqrt{n}$ - nbds at sample size $n$:

$$Q_n = (1 - \tfrac{r}{\sqrt{n}})P + \tfrac{r}{\sqrt{n}}\,H$$

  (location, unconditional regression; scale: H symmetric)

- conditional regression $r/\sqrt{n}$ - nbds, with radius curve $\varepsilon(x)$:

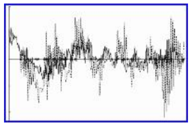
$$Q_n(du\,|\,x) = (1 - \tfrac{r}{\sqrt{n}}\,\varepsilon(x))\Phi(du) + \tfrac{r}{\sqrt{n}}\,\varepsilon(x)\,H(du\,|\,x),$$

- in time series: contaminated transition probabilities
  $Q_n(dy_t|\bar{y}_{t-1})$ where $\bar{y}_{t-1} := y_{t-1},\ldots,y_1$

$$= (1 - \tfrac{r}{\sqrt{n}}\,\varepsilon(\bar{y}_{t-1}))P(dy_t|\bar{y}_{t-1}) + \tfrac{r}{\sqrt{n}}\,\varepsilon(\bar{y}_{t-1})\,H_n(dy_t\,|\,\bar{y}_{t-1})$$

- $\|\varepsilon\|_\alpha \le 1$: $\mathrm{E}\,\varepsilon \le 1$ $(\alpha = 1)$, $\mathrm{E}\,\varepsilon^2 \le 1$ $(\alpha = 2)$, $\varepsilon \le 1$ $(\alpha = \infty)$
  $\mathrm{E}$ is taken under the ideal measure $P$, resp. ideal regressor distr.





## 8.3 Relative Maximum Risk Over Nbds

We use the estimate which is optimally robust for the neighborhood model of an assumed radius while this radius may not be true.

- relative Var (in Huber[64] model):

  (minmax) M-estimates of location, $\sum_{i=1}^{n} \psi(y_i - S_n) \approx 0$

$$\mathrm{relVar}\,(\psi_{s_0}, s) = \frac{\mathrm{maxVar}\,(\psi_{s_0}, s)}{\mathrm{maxVar}\,(\psi_s, s)}, \qquad 0 \le s < 1$$

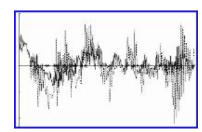
- relative MSE ($r/\sqrt{n}$ - neighborhoods, Ri[94]):

  (minmax) asy. linear estimates with influence curves

$$\sqrt{n}\,(S_n - \theta) - n^{-1/2} \sum_{i=1}^{n} \eta(y_i) \longrightarrow 0 \quad \text{in } P\text{-prob.}$$

$$\mathrm{relMSE}\,(\eta_{r_0}, r) = \frac{\mathrm{maxMSE}\,(\eta_{r_0}, r)}{\mathrm{maxMSE}\,(\eta_r, r)}, \qquad 0 \le r < \infty$$





# 8.4 Location (1-dimensional)

## 8.4.a Minimax asymptotic variance

- Minimax M-estimate for $s \in [0,1)$:

$$\psi_s(u) = (-m_s) \vee u \wedge m_s\,, \qquad \frac{s}{1-s}\, m_s = \mathrm{E}(|u| - m_s)_+$$

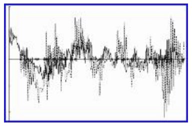
- Maximal asymptotic variance of $\psi_{s_0}$ under $s$:

$$\max\mathrm{Var}\,(\psi_{s_0}, s) = \frac{(1-s)\,\mathrm{E}\,\psi_{s_0}^2 + s\,m_0^2}{\bigl[(1-s)\,\mathrm{E}\,\psi'_{m_0}\bigr]^2}$$

- Median ($s=1$): $\psi_1(u) = \mathrm{sign}\,(u) = \lim\limits_{s\to 1} \frac{1}{m_s}\psi_s(u)$,

$$\max\mathrm{Var}\,(\psi_1, s) = \frac{\pi}{2(1-s)^2}\,, \qquad \mathrm{relVar}\,(\psi_1, s) \longrightarrow 1 \quad (s \to 1)$$





# 8.4 Location (1-dimensional)

## 8.4.b Minimax asymptotic MSE

- Minimax IC for $r \in [0, \infty)$: $\eta_r(u) = A_r u \min\left\{1, \dfrac{c_r}{|u|}\right\}$,

$$1 = A_r \operatorname{E} u^2 \min\left\{1, \tfrac{c_r}{|u|}\right\}, \qquad r^2\, c_r = \operatorname{E}\left(|u| - c_r\right)_+$$

- Median ($r = \infty$): $\eta_\infty\,(u) = b_{\min} \operatorname{sign}(u)$,
- Minimal bias (of asy. lin. est.): $b_{\min} = \sqrt{\frac{\pi}{2}}$
- Maximal MSE of $\eta_{r_0}$ under $r$:

$$\operatorname{maxMSE}(\eta_{r_0}, r) = A_{r_0}^2 \;\operatorname{E}\min\left\{u^2,\, c_{r_0}^2\right\} \;+\; r^2\; A_{r_0}^2\, c_{r_0}^2$$

- **Coincidence:** for $\boldsymbol{s = r^2/(1+r^2)}$, $\boldsymbol{s_0 = r_0^2/(1+r_0^2)}$

$$\begin{aligned} \mathbf{(1-s)\,maxMSE}\,(\boldsymbol{\eta_{r_0}, r}) &= \mathbf{maxVar}\,(\boldsymbol{\psi_{s_0}, s}) \\ \Longrightarrow \mathbf{relVar}\,(\boldsymbol{\psi_{s_0}, s}) &= \mathbf{relMSE}\,(\boldsymbol{\eta_{r_0}, r}) \end{aligned}$$

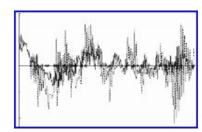
UNIVERSITÄT BAYREUTH

Mathematics VII

*Helmut Rieder*

Robust Estimation for Time Series Models

**06/12/2003**



# 8.5 Scale (1-dimensional) — MSE

## 8.5.a Scale—contamination balls

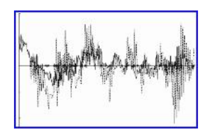
- Minimax IC for $r \in [0, \infty)$: $\eta_r(u) = A_r(u^2 - \alpha_r^2)\min\left\{1, \dfrac{c_r}{|u^2 - \alpha_r^2|}\right\}$,

$$\begin{aligned} 0 &= \mathrm{E}(u^2 - \alpha_r^2)\min\left\{1, \frac{c_r}{|u^2 - \alpha_r^2|}\right\}, \\ A_r^{-1} &= \mathrm{E}(u^2 - \alpha_r^2)^2\min\left\{1, \frac{c_r}{|u^2 - \alpha_r^2|}\right\}, \\ r^2\, c_r &= \mathrm{E}(|u^2 - \alpha_r^2| - c_r)_+ \end{aligned}$$

- MAD ($r = \infty$): $\eta_\infty(u) = b_{\min}\,\mathrm{sign}\,(|u| - \alpha_\infty)$, $\hat\theta = \alpha_\infty^{-1}\mathrm{med}(|u_i|)$
- Minimal bias (of asy. lin. est.): $b_{\min} = (4\alpha_\infty\varphi(\alpha_\infty))^{-1} = 1.166$
- $0 < \alpha_r$ decreasing from $\alpha_0 = 1$ to $\alpha_\infty := \Phi^{-1}(3/4) = 0.674$
- clipping of $|u|$ from below and above iff $r \geq 0.92$;
  clipping of $|u|$ only from above for $r \leq 0.92$.
- For $r_0, r \in [0, \infty)$, the maximal MSE is

$$\mathrm{maxMSE}\,(\eta_{r_0}, r) = A_{r_0}^2\,\mathrm{E}\min\{|u^2 - \alpha_{r_0}^2|^2,\ c_{r_0}^2\} \ +\ r^2 A_{r_0}^2 c_{r_0}^2$$





## 8.5.b Scale—total variation balls

- Minimax IC for $r \in [0, \infty)$:

$$\eta_r(u) = A_r\{[\,g_r \vee u^2 \wedge (\,g_r + c_r)] - 1\}$$

$$0 = \mathrm{E}(\,g_r - u^2)_+ - \mathrm{E}(u^2 - g_r - c_r)_+$$

$$1 = A_r\,\mathrm{E}\,u^2\{[\,g_r \vee u^2 \wedge (\,g_r + c_r)] - 1\}$$

$$r^2\,c_r = \mathrm{E}(\,g_r - u^2)_+$$

- MADv ($r = \infty$):

$$\eta_\infty\,(u) = \omega_v^{\min}\{P(|u| < 1)\,\mathrm{I}(|u| > 1) - P(|u| > 1)\,\mathrm{I}(|u| < 1)\}$$

- Minimal bias (of asy. lin. est.):

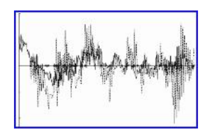
$$\omega_v^{\min} = (\mathrm{E}\,\Lambda_+)^{-1} = \sqrt{\tfrac{\pi}{2}\,e} \approx 2.066$$

- clipping of $|u|$ always from above and below

- For $r_0, r \in [0, \infty)$, the maximal MSE is

$$\mathrm{maxMSE}\,(\eta_{r_0}, r) = A_{r_0}^2\,\mathrm{E}\,\{[\,g_{r_0} \vee u^2 \wedge (\,g_{r_0} + c_{r_0})] - 1\}^2 + r^2 A_{r_0}^2 c_{r_0}^2$$





# 8.6 Location (dimension $k \geq 1$) — MSE

**Minimax asymptotic MSE**

- Minimax IC for $r \in [0, \infty)$: $\eta_r(u) = \alpha_r u \min\left\{1, \dfrac{c_r}{|u|}\right\}$,

$$k = \alpha_r \;\mathrm{E}\,|u|^2 \min\left\{1, \tfrac{c_r}{|u|}\right\}, \qquad r^2\, c_r = \mathrm{E}(|u| - c_r)_+$$

- min$L_1$ $(r = \infty)$: $\sum_{i=1}^n |u_i - \hat\theta|_2 = \min_\theta !$ $\quad \eta_\infty\,(u) = b_{\min} \frac{u}{|u|}$

- **Minimal bias** (of asy. lin. est.):

$$b_{\min} = \frac{k}{\mathrm{E}\,|\Lambda|} = \frac{k\,\Gamma(\frac{k}{2})}{\sqrt{2}\,\Gamma(\frac{k+1}{2})}, \qquad \frac{b_{\min}}{\sqrt{k}} \to 1, \;\; \frac{\mathrm{E}\,|\eta_\infty|^2}{k} \to 1$$

- Maximal MSE of $\eta_{r_0}$ under $r$:

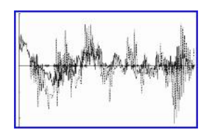
$$\mathrm{maxMSE}\,(\eta_{r_0}, r) = \alpha_{r_0}^2 \,\mathrm{E}\min\{|u|^2,\, c_{r_0}^2\} \;+\; r^2\; \alpha_{r_0}^2 c_{r_0}^2$$

- Relative MSE: $\eta_\infty$ becomes radius–minimax

$$\frac{\mathbf{maxMSE}\,(\boldsymbol{\eta}_{\boldsymbol{r}_0}, \boldsymbol{r})}{\mathbf{maxMSE}\,(\boldsymbol{\eta}_\infty, \boldsymbol{r})} \longrightarrow \mathbf{1} \qquad \textbf{[}\boldsymbol{\infty}\textbf{loc]}$$

uniformly for $0 \leq r_0, r \leq$ any $r_1 < \infty$, as $k \to \infty$.





# 8.7 Regression (k-dim.) — MSE

### 8.7.a Minimax asymptotic MSE $(* = c, \alpha = 1)$

- Minimax IC for $r \in [0,\infty)$: $\eta_r(x,u) = \alpha_r x\, u \min\left\{1, \dfrac{c_r}{|xu|}\right\}$,

$$k = \alpha_r \,\mathrm{E}\,|x|^2 u^2 \min\left\{1, \tfrac{c_r}{|xu|}\right\}, \qquad r^2\, c_r = \mathrm{E}\,(|xu| - c_r)_+$$

- weighted min$L_1$ $(r = \infty)$: $\eta_\infty\,(x,u) = b_{\min} \frac{x}{|x|}\,\mathrm{sign}\,(u)$

- Minimal bias (of asy. lin. est.): $b_{\min} = \frac{k}{\mathrm{E}\,|\Lambda|} = \sqrt{\frac{\pi}{2}}\,\frac{k}{\mathrm{E}\,|x|}$

- Maximal MSE of $\eta_{r_0}$ under $r$:

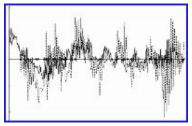
$$\mathrm{maxMSE}\,(\eta_{r_0}, r) = \alpha_{r_0}^2\,\mathrm{E}\min\{|x|^2u^2,\, c_{r_0}^2\} \;+\; r^2\;\alpha_{r_0}^2 c_{r_0}^2$$

- RelMSE same for all $\theta$, but depends on $K(dx)$.
  Convergence to 1-dim. location:

$$\mathbf{relMSE}\,(\boldsymbol{\eta_{r_0}}, \boldsymbol{r}) \longrightarrow \mathbf{relMSE}\,(\boldsymbol{\eta_{r_0}^{\mathrm{1loc}}}, \boldsymbol{r}) \qquad [\mathbf{1loc}]$$

uniformly for $0 \le r_0, r \le$ any $r_1 < \infty$, as $k \to \infty$.





## 8.7.b Minimax asymptotic MSE ($* = c, \alpha = 2$)

- Minimax IC for $r \in [0, \infty)$: $\eta_r(x,u) = \alpha_r x\, u \min\left\{1, \dfrac{c_r}{|u|}\right\}$,

$$k = \alpha_r \,\mathrm{E}\,|x|^2 \cdot \mathrm{E}\,u^2 \min\left\{1, \tfrac{c_r}{|u|}\right\},\; r^2\, c_r = \mathrm{E}\,(|u| - c_r)_+$$

- min$L_1$: $\eta_\infty\,(x,u) = \mathcal{K}^{-1} \frac{x}{\mathrm{E}\,|u|}\,\mathrm{sign}\,(u), \quad \mathcal{K}^{-1} = \mathrm{E}\,x x' = \gamma \cdot \mathbb{I}_k$

- Minimal bias (of asy. lin. est.): $b_{\min} = \dfrac{\sqrt{\mathrm{tr}\,\mathcal{K}^{-1}}}{\mathrm{E}\,|u|} = \sqrt{\dfrac{\pi k}{2\gamma}}$

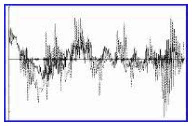
- Maximal MSE of $\eta_{r_0}$ under $r$:

$$\begin{aligned} \mathrm{maxMSE}\,(\eta_{r_0}, r) &= \mathrm{E}\,|x|^2 \alpha_{r_0}^2 \left\{\mathrm{E}\,\min\{u^2,\, c_{r_0}^2\} \;+\; r^2\, c_{r_0}^2\right\} \\ &= \frac{k^2}{\mathrm{E}\,|x|^2}\,\mathrm{maxMSE}\,(\eta_{r_0}^{1\mathrm{loc}}, r) \end{aligned}$$

- Relative MSE identical with one-dimensional location,

$$\mathbf{relMSE}\,(\boldsymbol{\eta}_{r_0}, \boldsymbol{r}) = \mathbf{relMSE}\,(\boldsymbol{\eta}_{r_0}^{\mathbf{1loc}}, \boldsymbol{r})$$





### 8.7.d Minimax asymptotic MSE $(* = h, \alpha = 1)$

- Minimax IC for $r \in [0, \infty)$: $\eta_r(x,u) = \alpha_r x\, u \min\left\{1, \dfrac{c_r}{|x|}\right\}$,

$$k = \alpha_r \,\mathrm{E}\, u^2 \cdot \mathrm{E}\, |x|^2 \min\left\{1, \tfrac{c_r}{|x|}\right\}, \qquad 8\, r^2\, c_r = \mathrm{E}\,(|x| - c_r)_+$$

- weighted min$L_2$ $(r = \infty)$: $\eta_\infty\,(x,u) = \frac{1}{\sqrt{8 \mathrm{E}\, u^2}}\, b_{\min}\, \frac{x}{|x|}\, u$
- Minimal bias (of asy. lin. est.): $b_{\min} = \sqrt{8}\, \frac{k}{\mathrm{E}\,|x|}$
- Maximal MSE of $\eta_{r_0}$ under $r$:

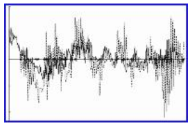
$$\mathrm{maxMSE}\,(\eta_{r_0}, r) = \alpha_{r_0}^2 \mathrm{E}\, u^2 \cdot \mathrm{E}\, \min\{|x|^2,\, c_{r_0}^2\} \;+\; 8\, r^2\, \alpha_{r_0}^2 c_{r_0}^2$$

- Relative MSE related to $k$-dimensional location

$$\mathbf{relMSE}\,(\boldsymbol{\eta}_{\boldsymbol{r_0}}, \boldsymbol{r}) = \mathbf{relMSE}\,(\boldsymbol{\eta}^{\mathbf{kloc}}_{\sqrt{8}\, \boldsymbol{r_0}}, \sqrt{8}\, \boldsymbol{r}), \quad \boldsymbol{K(dx)} = \boldsymbol{\mathcal{N}(0, \sigma^2 \mathbb{I}_k)}$$

For $K(dx) = \mathrm{Ufo}\,(B_k(0,m))$ the **convergence** $[\boldsymbol{\infty}\mathbf{loc}]$ holds: $\eta_\infty$ becomes radius–minimax as $k \to \infty$ (as for $k$-dim. location).





# 8.8 Relative MSE – numerical evaluation

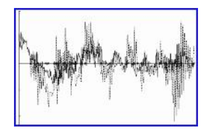
**Location ($P_\theta = \mathcal{N}(\theta, \mathbb{I}_k)$, $r/\sqrt{n}$-contamination balls)**

| k | $\eta_\infty$ | $\rho = 0$ | $\rho = \frac{1}{3}$ | $\rho = \frac{1}{2}$ | $r_0$ | $r_1(\frac{1}{3})$ | $r_1(\frac{1}{2})$ |
|---|---|---|---|---|---|---|---|
| 1 | 1.571 | 1.181 | 1.088 | 1.044 | 0.621 | 0.548 | 0.574 |
| 2 | 1.273 | 1.121 | 1.063 | 1.032 | 0.627 | 0.527 | 0.558 |
| 3 | 1.178 | 1.091 | 1.049 | 1.026 | 0.611 | 0.496 | 0.529 |
| 5 | 1.104 | 1.062 | 1.035 | 1.018 | 0.577 | 0.450 | 0.481 |
| 10 | 1.051 | 1.035 | 1.020 | 1.011 | 0.520 | 0.385 | 0.413 |
| 15 | 1.0339 | 1.0247 | 1.0143 | 1.0078 | 0.485 | 0.351 | 0.375 |

**Scale ($P_\sigma = \mathcal{N}(0, \sigma^2)$, $r/\sqrt{n}$-contamination balls)**

| $\eta_\infty$ | $\rho = 0$ | $\rho = \frac{1}{3}$ | $\rho = \frac{1}{2}$ | $r_0$ | $r_1(\frac{1}{3})$ | $r_1(\frac{1}{2})$ |
|---|---|---|---|---|---|---|
| 2.721 | 1.505 | 1.207 | 1.099 | 0.499 | 0.481 | 0.551 |





**Scale ($P_\sigma = \mathcal{N}(0,\sigma^2)$, $r/\sqrt{n}$-total variation balls)**

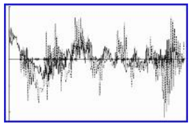
| $\eta_\infty$ | $\rho = 0$ | $\rho = \frac{1}{3}$ | $\rho = \frac{1}{2}$ | $r_0$ | $r_1(\frac{1}{3})$ | $r_1(\frac{1}{2})$ |
|---|---|---|---|---|---|---|
| 1.850 | 1.254 | 1.115 | 1.056 | 0.265 | 0.237 | 0.249 |

**Regression ($y = x\theta + u,\ u \sim \mathcal{N}(0,\sigma_u^2), x \sim K(dx)$ )**

conditional $r/\sqrt{n}$-contamination balls, $\alpha = 1$, $K(dx) = \mathrm{Ufo}\,(B_k(0,m))$

| $k$ | $\eta_\infty$ | $\rho = 0$ | $\rho = \frac{1}{3}$ | $\rho = \frac{1}{2}$ | $r_0$ | $r_3$ | $r_2$ |
|---|---|---|---|---|---|---|---|
| 1 | 2.094 | 1.271 | 1.122 | 1.060 | 0.566 | 0.517 | 0.540 |
| 2 | 1.767 | 1.227 | 1.107 | 1.053 | 0.595 | 0.532 | 0.558 |
| 3 | 1.677 | 1.209 | 1.100 | 1.049 | 0.604 | 0.536 | 0.562 |
| 5 | 1.616 | 1.194 | 1.094 | 1.047 | 0.611 | 0.540 | 0.565 |
| 10 | 1.584 | 1.185 | 1.090 | 1.045 | 0.617 | 0.545 | 0.570 |
| 15 | 1.577 | 1.183 | 1.089 | 1.044 | 0.619 | 0.546 | 0.572 |





conditional $r/\sqrt{n}$-contamination balls, $\alpha = 1$, $K(dx) = \mathcal{N}(0, \sigma^2 \mathbb{I}_k)$

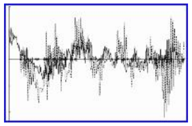
| $k$ | $\eta_\infty$ | $\rho = 0$ | $\rho = \frac{1}{3}$ | $\rho = \frac{1}{2}$ | $r_0$ | $r_3$ | $r_2$ |
|---|---|---|---|---|---|---|---|
| 1 | 2.467 | 1.347 | 1.146 | 1.070 | 0.515 | 0.474 | 0.496 |
| 2 | 2.000 | 1.287 | 1.127 | 1.062 | 0.555 | 0.499 | 0.525 |
| 3 | 1.851 | 1.258 | 1.117 | 1.057 | 0.569 | 0.506 | 0.534 |
| 5 | 1.735 | 1.231 | 1.107 | 1.053 | 0.584 | 0.514 | 0.542 |
| 10 | 1.651 | 1.207 | 1.098 | 1.049 | 0.598 | 0.526 | 0.553 |
| 15 | 1.624 | 1.199 | 1.095 | 1.047 | 0.605 | 0.532 | 0.558 |

conditional $r/\sqrt{n}$-contamination balls, $* = c$, $\alpha = 2$:
same numbers as for 1-dim. location, $\forall\, K(dx)$!

**ARMA($p$,$q$)** with shift ($* = c$, $\alpha = 2$):
with $K = \mathcal{L}_{\rm id}(H)$, same relMSE as 1-dim. location, $\forall\, \theta$.
**ARMA(1,1)** with $* = c$, $\alpha = 1$, $\mu = 0$; for $\phi = -.7$, $\xi = .35$:
$2.163\,(\eta_\infty)$, $1.3138\,(\rho = 0)$, $0.5414\,(r_0)$ ($\approx$ 2-dim. regr.)

**ARCH(1):** conditional $r/\sqrt{n}$-balls ($* = c, v$)
$\alpha = 2$: same relMSE as 1-dim. scale, and independent of ARCH(1)-parameter $\theta$;
$0 < \theta < \exp(-\operatorname{E} \log u_1^2) \approx 1.887364$
$\alpha = 1$: similar relMSE as 1-dim. scale; but dependent on $\theta$.





conditional $r/\sqrt{n}$-Hellinger balls, $\alpha = 1$, $K(dx) = \mathcal{N}(0, \sigma^2 \mathbb{I}_k)$:
same numbers as for k-dim. location $(r_i = r_i^{\rm kloc}/\sqrt{8}\,;\; i = 0, 2, 3)$

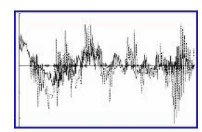
conditional $r/\sqrt{n}$-Hellinger balls, $\alpha = 1$, $K(dx) = \mathrm{Ufo}\,(B_k(0, m))$

| k | $\eta_\infty$ | $\rho = 0$ | $\rho = \frac{1}{3}$ | $\rho = \frac{1}{2}$ | $r_0$ | $r_3$ | $r_2$ |
|---|---|---|---|---|---|---|---|
| 1 | 1.333 | 1.101 | 1.055 | 1.029 | 0.255 | 0.231 | 0.238 |
| 2 | 1.125 | 1.055 | 1.032 | 1.017 | 0.247 | 0.211 | 0.220 |
| 3 | 1.067 | 1.035 | 1.021 | 1.011 | 0.232 | 0.191 | 0.199 |
| 5 | 1.029 | 1.018 | 1.011 | 1.006 | 0.207 | 0.162 | 0.169 |
| 10 | 1.008 | 1.006 | 1.004 | 1.002 | 0.170 | 0.124 | 0.129 |
| 15 | 1.004 | 1.003 | 1.002 | 1.000 | 0.149 | 0.104 | arb. |

$r/\sqrt{n}$-Hellinger balls, $\alpha = 2$: $\quad \mathrm{relMSE} \equiv 1!$

In our models—except scale, ARCH(1)—the results for $r/\sqrt{n}$-total variation $(* = v)$ agree with results for $2r/\sqrt{n}$-contamination $(* = c)$; hence, in these cases, same relMSE and half the least favorable radii.





# Comparison With Semiparametrics

## 8.9 Efficiency Breakdown

**Symmetric location** about $\theta \in \mathbb{R}$:

$$y_i = \theta + u_i\,, \quad u_i \text{ i.i.d.} \sim dF = f\,d\lambda$$

where $\mathcal{I}_F^{\rm loc} = \int (\Lambda_F^{\rm loc})^2\,dF$ finite, $\Lambda_F^{\rm loc} = -\dot f/f$, $F$ symmetric about 0.

**Proposition 8.9.1** Given $F_0$ sym., $\mathcal{I}_{F_0}^{\rm loc}$ finite, and $\psi_0 \colon \mathbb{R}$ odd, such that

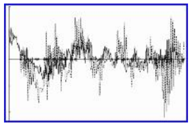
$$\mathrm{Var}_{\rm loc}(\psi_0, F_0) = \frac{\int \psi_0^2\,dF_0}{\left(\int \psi_0 \Lambda_{F_0}^{\rm loc}\,dF_0\right)^2} \in (0,\infty)$$

(e.g., $\psi_0 = \Lambda_{F_0}^{\rm loc}$). Assume $\psi_0$ abs. continuous in some nondeg. interval with bounded derivative. Then,

$$\sup_{F \in \mathcal{F}(F_0,\varepsilon)} \mathrm{Var}_{\rm loc}(\psi_0, F) \cdot \mathcal{I}_F^{\rm loc} = \infty \quad \forall \varepsilon \in (0,1)$$

where $\mathcal{F}(F_0,\varepsilon) = \left\{\, (1-\varepsilon)F_0 + \varepsilon\,H \;\middle|\; H \text{ sym.}, \mathcal{I}_H^{\rm loc} < \infty \,\right\}$.





**Proof** employs $F \approx (1-\varepsilon)F_0 + \frac{\varepsilon}{2}(\mathrm{I}_{-a} + \mathrm{I}_a)$, and uses representation,

$$\mathcal{I}_F^{\mathrm{loc}} = \sup_{\varphi \in \mathcal{C}_c^1} \frac{\left(\int \dot{\varphi}\, dF\right)^2}{\int \varphi^2\, dF}$$

Fisher information of location is convex, and vaguely lower semicontinuous, but not upper s.c. (in total variation).

**Scale model** with scale $\sigma \in (0,\infty)$:

$$y_i = \sigma u_i\,, \quad u_i \text{ i.i.d.} \sim dF = f\, d\lambda$$

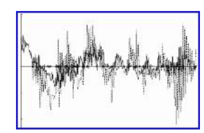
except possibly a point mass at $0$, where $u \mapsto uf(u)$ is absolutely continuous on $(-\infty,0)$ and on $(0,\infty)$, and $\mathcal{I}_F^{\mathrm{scal}} = \int_{\neq 0} (\Lambda_F^{\mathrm{scal}})^2\, dF$ is finite, where $\Lambda_F^{\mathrm{scal}} = u\Lambda_F^{\mathrm{loc}} - 1$. For general $F$ define

$$\mathcal{I}_F^{\mathrm{scal}} = \sup_{\varphi \in \mathcal{C}_c^1} \frac{\left(\int u\dot{\varphi}\, dF\right)^2}{\int \varphi^2\, dF}$$

**Proposition 8.9.2** $\mathcal{I}_F^{\mathrm{scal}} < \infty$ iff $dF = f\, d\lambda$ on $\mathbb{R} \setminus \{0\}$, $uf$ is abs. cts. on $(-\infty,0)$ and on $(0,\infty)$, and $\int_{\neq 0} (\Lambda_F^{\mathrm{scal}})^2\, dF < \infty$;
in which case $\mathcal{I}_F^{\mathrm{scal}}$ = this integral.
Hence also $\mathcal{I}_F^{\mathrm{scal}}$ is convex, vaguely l.s.c., but not u.s.c. (in total variation).

**Proposition 8.9.3** Given $F_1$ of finite $\mathcal{I}_{F_1}^{\mathrm{scal}}$, and $\psi_1 : \mathbb{R} \to \mathbb{R}$, $\int \psi_1\, dF_1 = 0$,





such that

$$\mathrm{Var}_{\mathrm{scal}}(\psi_1, F_1) = \frac{\int \psi_1^2 \, dF_1}{\left(\int \psi_1 \Lambda_{F_1}^{\mathrm{scal}} \, dF_1\right)^2} \in (0, \infty)$$

(e.g., $\psi_1 = \Lambda_{F_1}^{\mathrm{scal}}$). Assume $\psi_1$ is abs. continuous on two nondeg. intervals with bounded derivative, and positive on one, negative on the other. Then, for every $\varepsilon \in (0,1)$,

$$\sup_{F \in \mathcal{F}(F_1,\varepsilon), S_{\psi_1}(F)=1} \mathrm{Var}_{\mathrm{scal}}(\psi_1, F) \cdot \mathcal{I}_F^{\mathrm{scal}} = \infty$$

where $\mathcal{F}(F_1, \varepsilon) = \left\{ (1-\varepsilon)F_1 + \varepsilon\, H \;\middle|\; \mathcal{I}_H^{\mathrm{scal}} < \infty \right\}$.

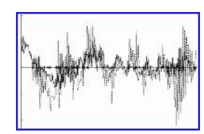
(symmetry version: in addition, $\psi_1$ even, $F_1$, $H$ sym. about 0.)

## 8.10 Nonuniform Convergence of Adaptive Fully Efficient Estimators

Structured (location/scale type) model $\{ P_\theta \}$, parameter $\theta \in \mathbb{R}^k$, $L_2$-differentiable transition probabilitites; observations $y_i$ driven by innovations $u_i$ i.i.d.$\sim F$; Fisher information of form

$$\mathcal{I}_{\theta,F} = \mathcal{K}_\theta \, \mathcal{I}_F^{\mathrm{loc/scal}} \tag{0.1}$$





Adaptive estimators $(S_n)$:

$$\mathcal{V}^n_{\theta,F} := \mathcal{L}_{\theta,F}\Big\{\sqrt{n}\,\mathcal{I}^{1/2}_{\theta,F}(S_n-\theta)\Big\} \longrightarrow \mathcal{N}(0,\mathbb{I}_k) \qquad (0.2)$$

for every $\theta$ and $F$ of finite $\mathcal{I}_F$. Fix $\theta$ and such an $F_0$.

**Theorem 8.10.1** Assume that: $\forall\, n\, \forall\, \delta_n > 0\, \exists\, \tau_n > 0$ such that,

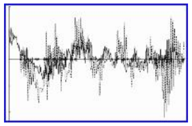
$$d_v(F,F_0) < \tau_n \Longrightarrow d_v\big(\mathcal{L}_{\theta,F}(Y_{1:n}), \mathcal{L}_{\theta,F_0}(Y_{1:n})\big) < \delta_n \qquad (0.3)$$

in total variation distance $d_v$, where $Y_{1:n} = (Y_1,\dots,Y_n)$.
Then, for every sequence $\varepsilon_n \in (0,1)$, $\varepsilon_n \to 0$ (arbitrarily fast!),

$$\liminf_{n\to\infty} \sup_{F\in\mathcal{F}(F,\varepsilon_n)} d_\kappa(\mathcal{V}^n_{\theta,F}, \mathcal{N}(0,\mathbb{I}_k)) \ge 1 - \frac{1}{2^k} \qquad (0.4)$$

in Kolmogorv $d_\kappa$, and $\mathcal{F}(F_0,\varepsilon_n) = \big\{\, (1-\varepsilon_n)F_0 + \varepsilon_n\, H \,\big|\, \mathcal{I}_H < \infty \,\big\}$.
(At all instances, it suffices to argue with $F = (1-\varepsilon_n)F_0 + \varepsilon_n\, H$, $H$ symmetric, $H \approx \mathrm{I}_0$.)





## Examples of Models

**(a)** 1-dim. location; cf. Klaassen[80] using equivariance; Pfanzagl and Wefelmeyer [82]; Beran[74]; Stone[75]

**(b)** scale

**(c)** linear regression: $y = x'\theta + u$; here $\mathcal{K}_\theta = \mathrm{E}xx'$; Bickel[82]

**(d)** MA(q): $Y_t = \mu + \theta(B)U_t$; here $\mathcal{K}_\theta = \mathrm{E}_\theta\, \widetilde{H}_\theta \widetilde{H}_\theta'$ with $\widetilde{H}_\theta' = \big(H_\theta', 1/\theta(1)\big)$, $H_\theta' = \theta^{-1}(B)(U_{-1}, \ldots, U_{-q})$

**(e)** AR(p) and ARMA(p,q) — however without (0.3); weaker version:

**Proposition 8.10.2** Let $(S_n)$ satisfy (0.2) and suppose that

$$\text{each } S_n \text{ is a continuous function.} \qquad (0.5)$$

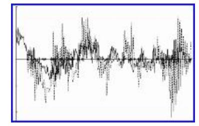
Then, for every sequence $\varepsilon_n \in (0,1)$, $\varepsilon_n \to 0$ (arbitrarily fast!),

$$\liminf_{n\to\infty} \sup_{F \in \mathcal{F}(F,\varepsilon_n)} d_\kappa(\mathcal{V}^n_{\theta,F}, \mathcal{N}(0, \mathbb{I}_k)) > 0 \qquad (0.6)$$

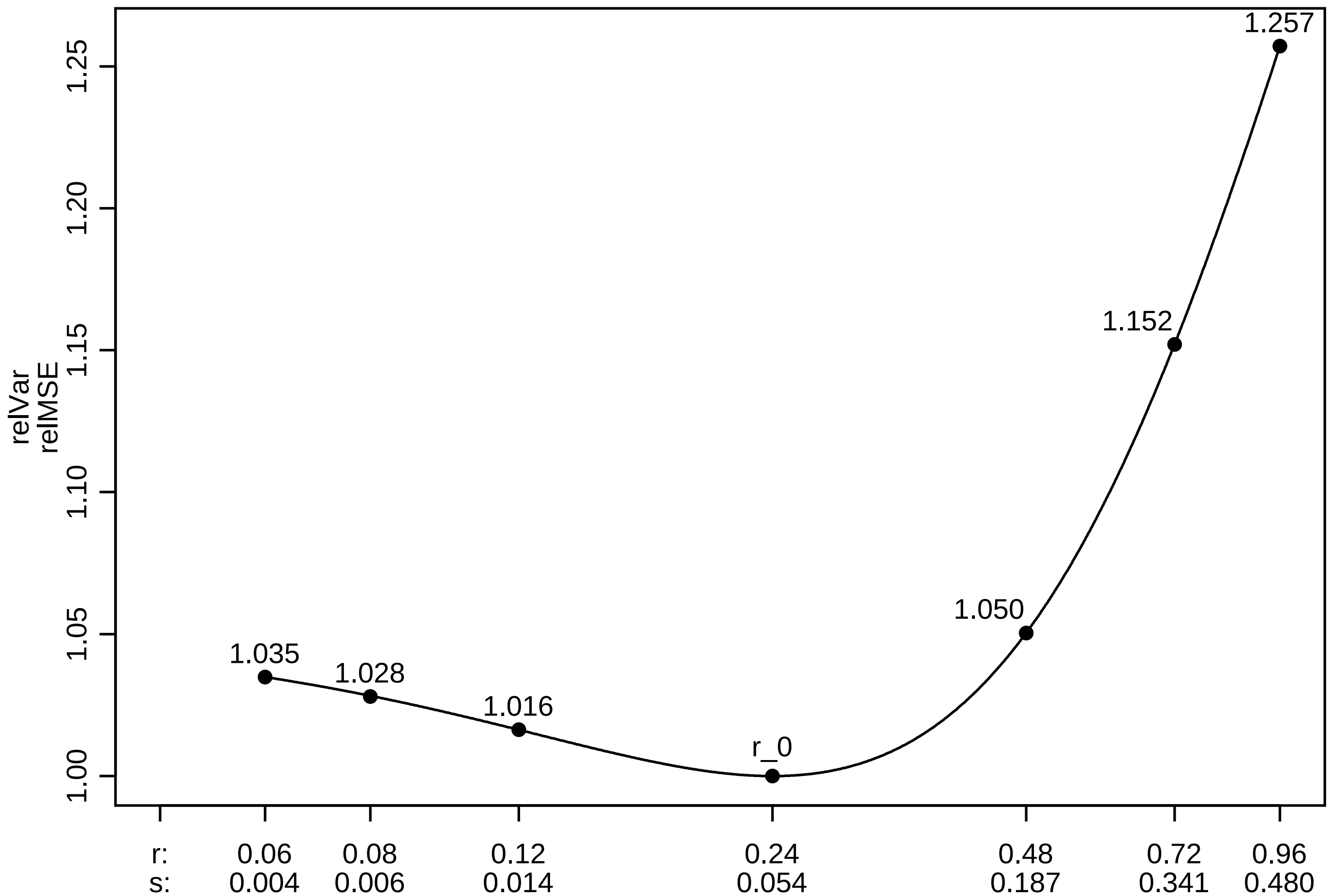
Thus estimates by Beran[76], Kreiss[87] are nonuniform.

**Need to adapt robustly—replacing inverse Fisher information by minmax asy. MSE over $r/\sqrt{n}$-total variation neighborhoods.**





relVar, relMSE: 1-Dimensional Location (Var=1.050 at r, s = 0)



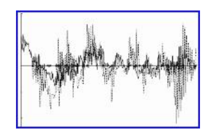

UNIVERSITÄT BAYREUTH

Mathematics VII

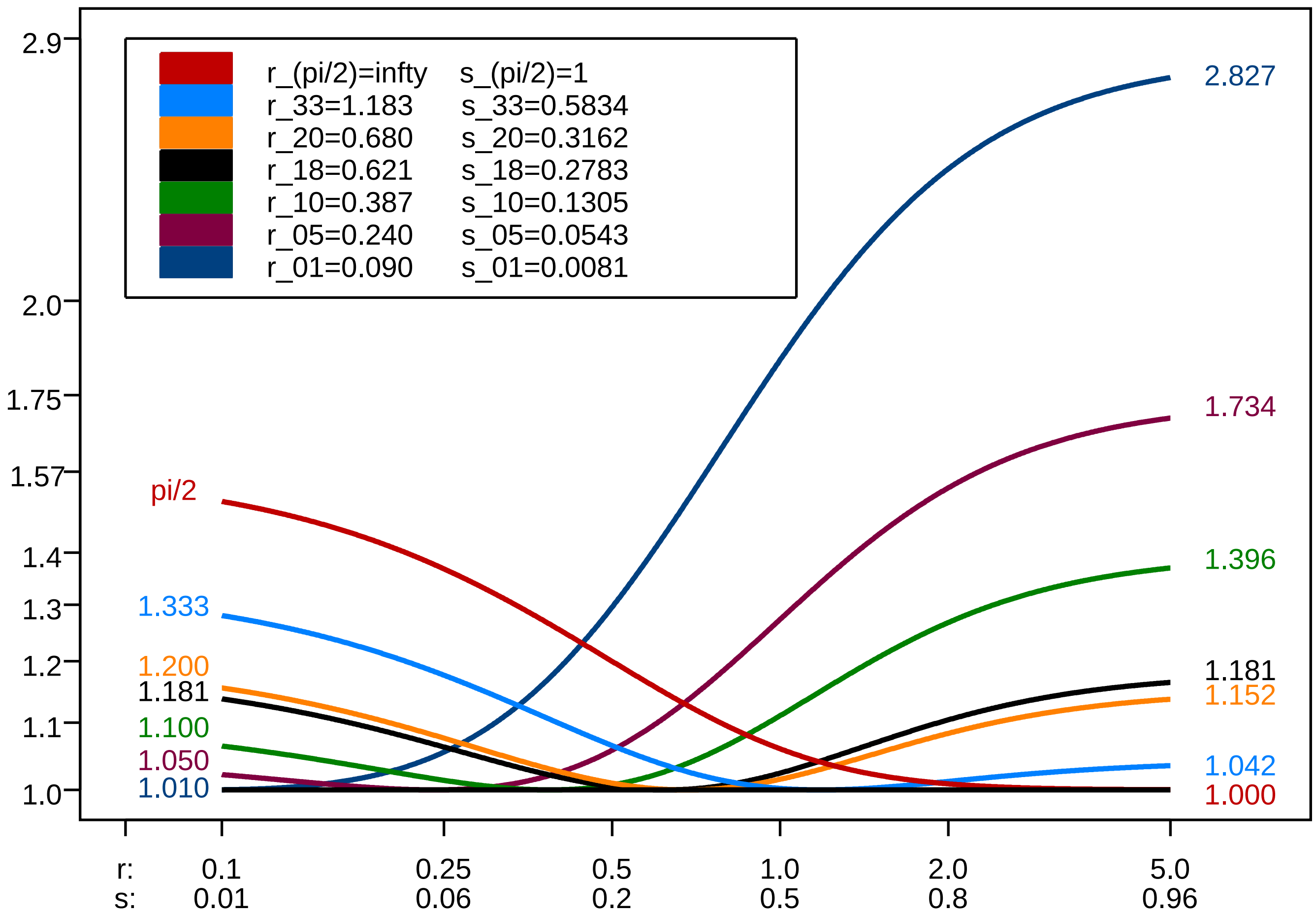
*Helmut Rieder*

**Robust Estimation for Time Series Models**

**06/12/2003**



1-dim. Location: relMSE, relVar vs. r, rsp. s=r^2/(1+r^2)



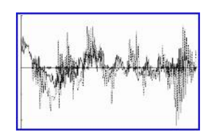

UNIVERSITÄT BAYREUTH

Mathematics VII

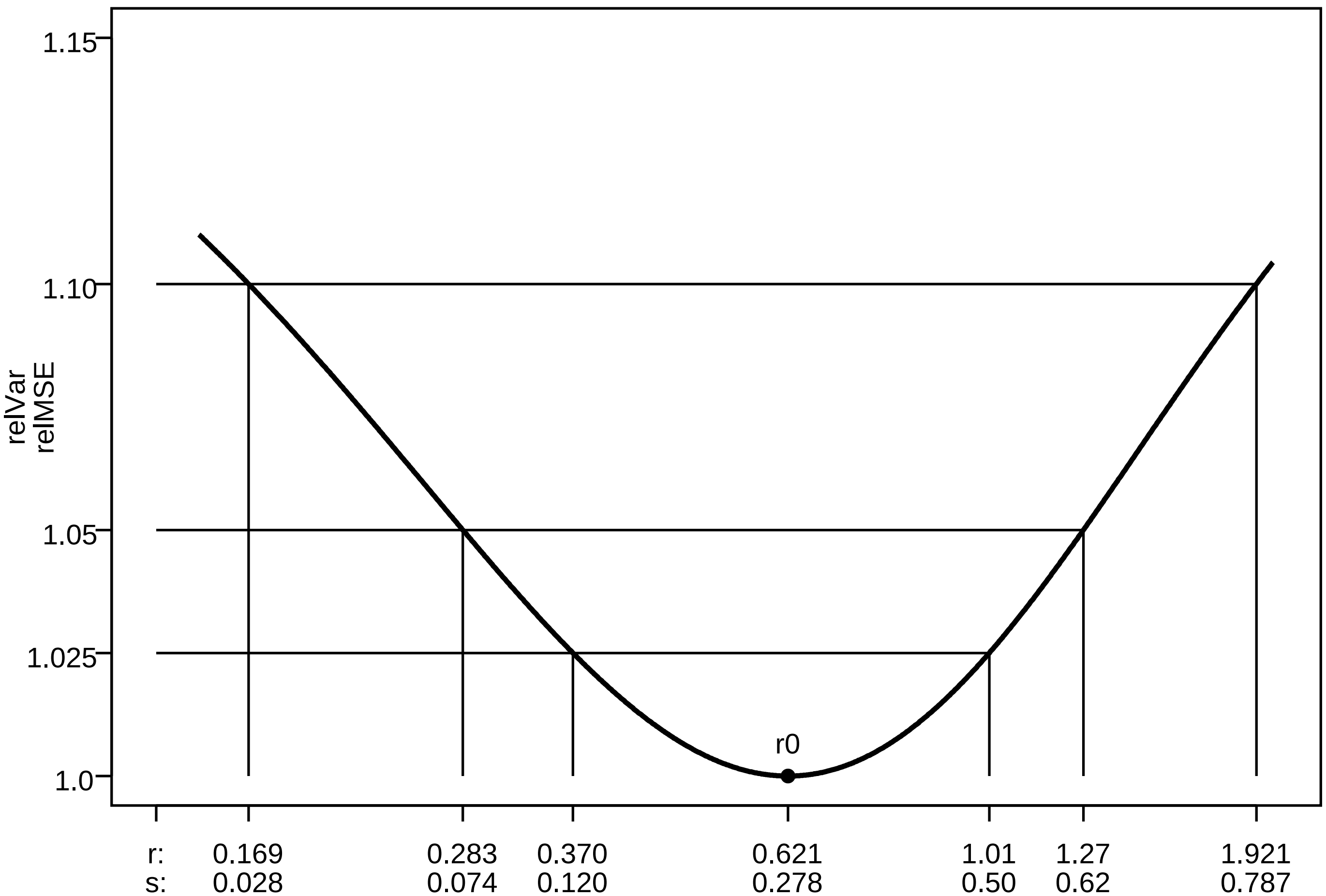
*Helmut Rieder*

**Robust Estimation for Time Series Models**

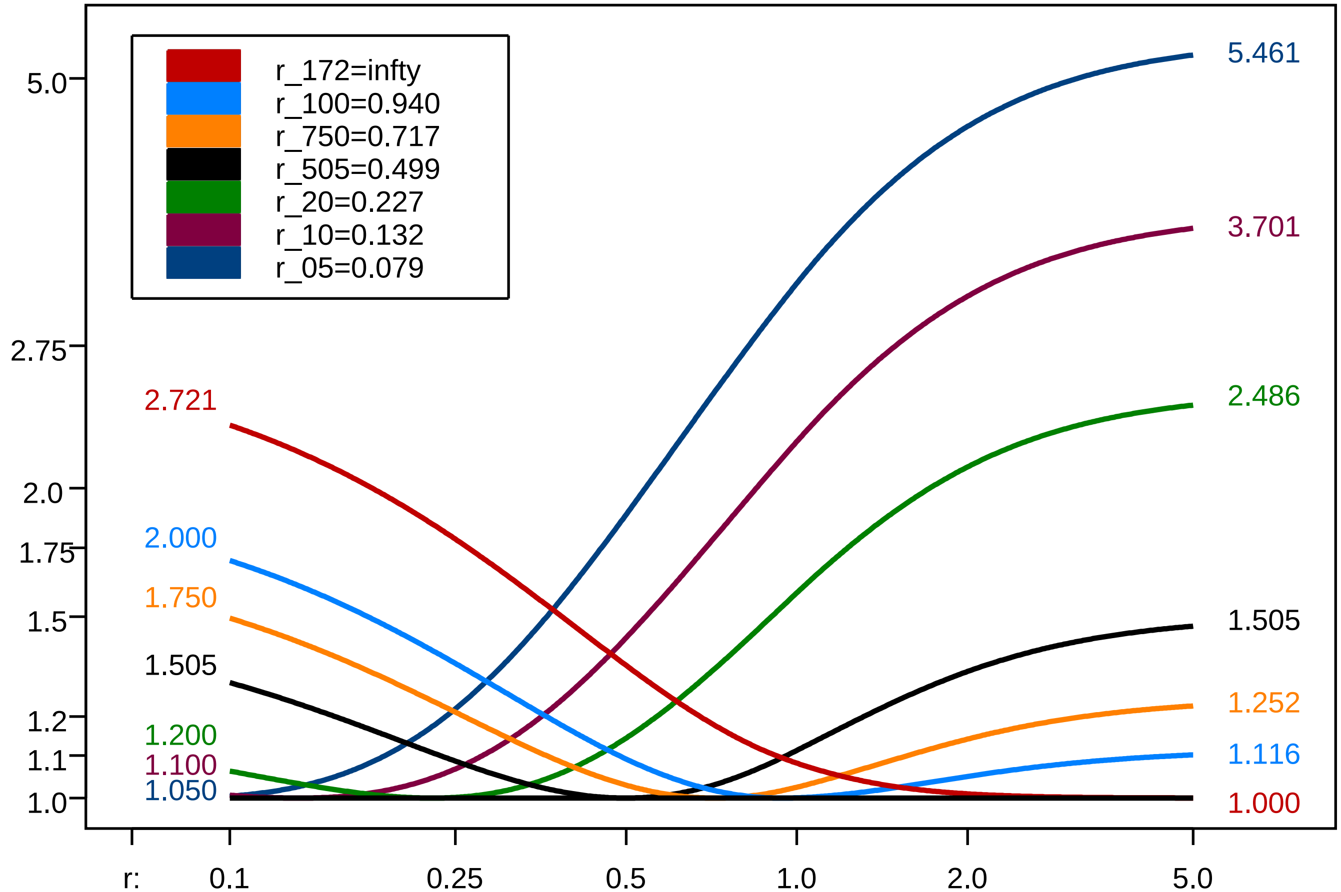
**06/12/2003**



## relVar, relMSE: 1-Dimensional Location (Var = 1.181 at r, s = 0)



relMSE: 1-Dimensional Scale (contamination)



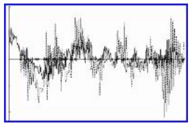

UNIVERSITÄT BAYREUTH

Mathematics VII

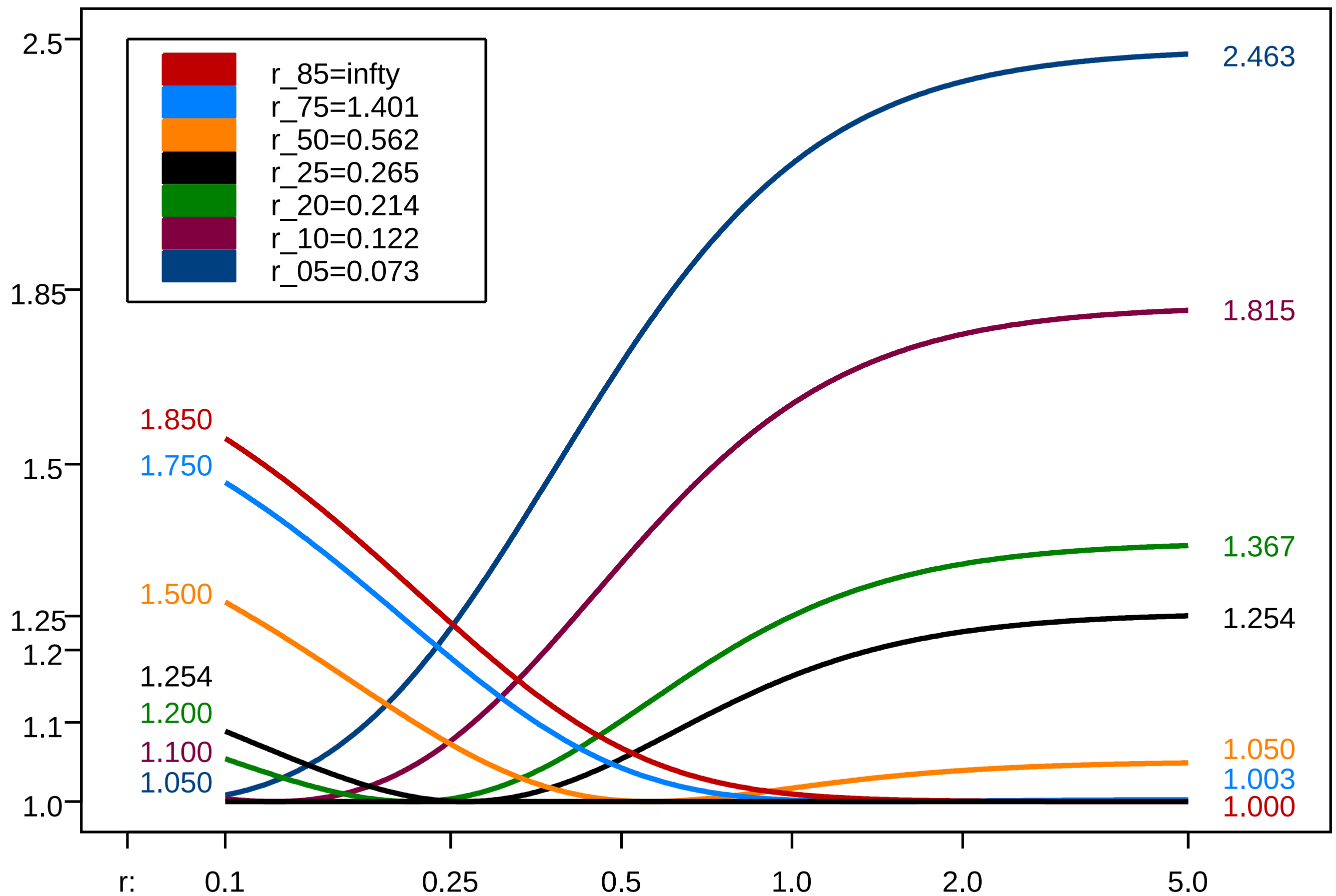
*Helmut Rieder*

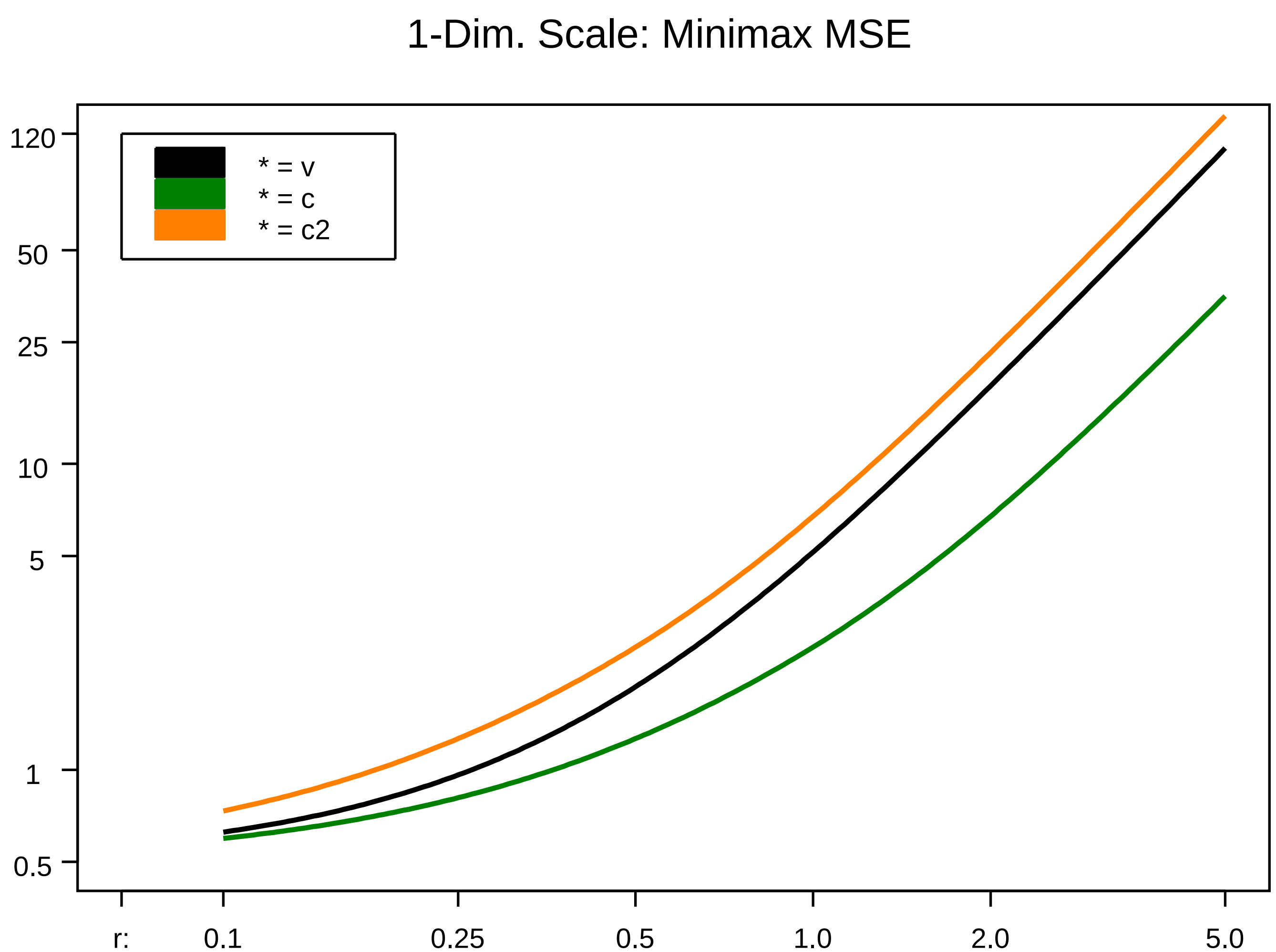
**Robust Estimation for Time Series Models**

**06/12/2003**



## relMSE: 1-Dimensional Scale (total variation)





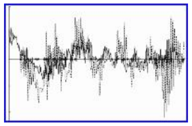

UNIVERSITÄT BAYREUTH

Mathematics VII

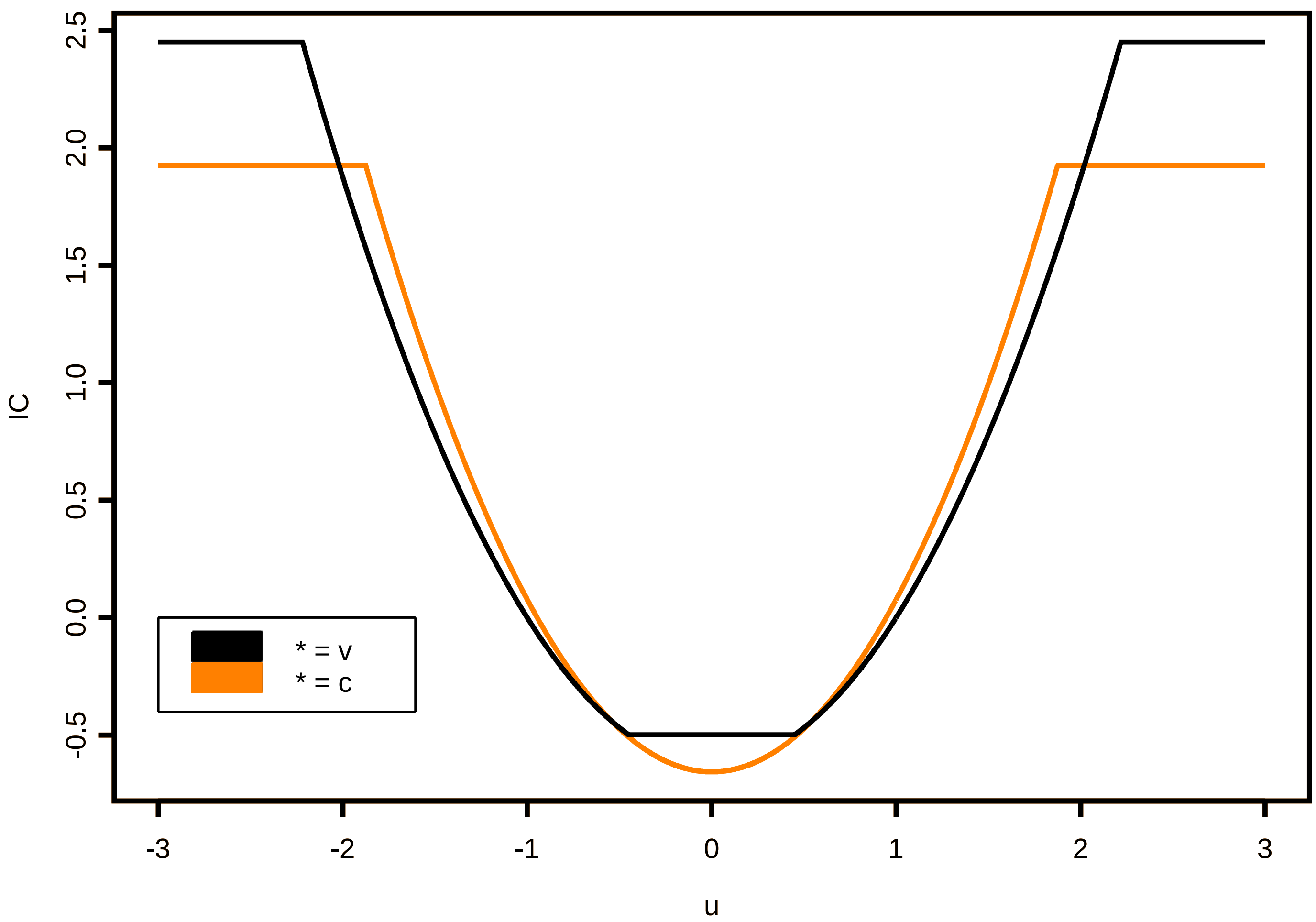
*Helmut Rieder*

**Robust Estimation for Time Series Models**

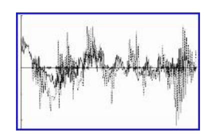
**06/12/2003**



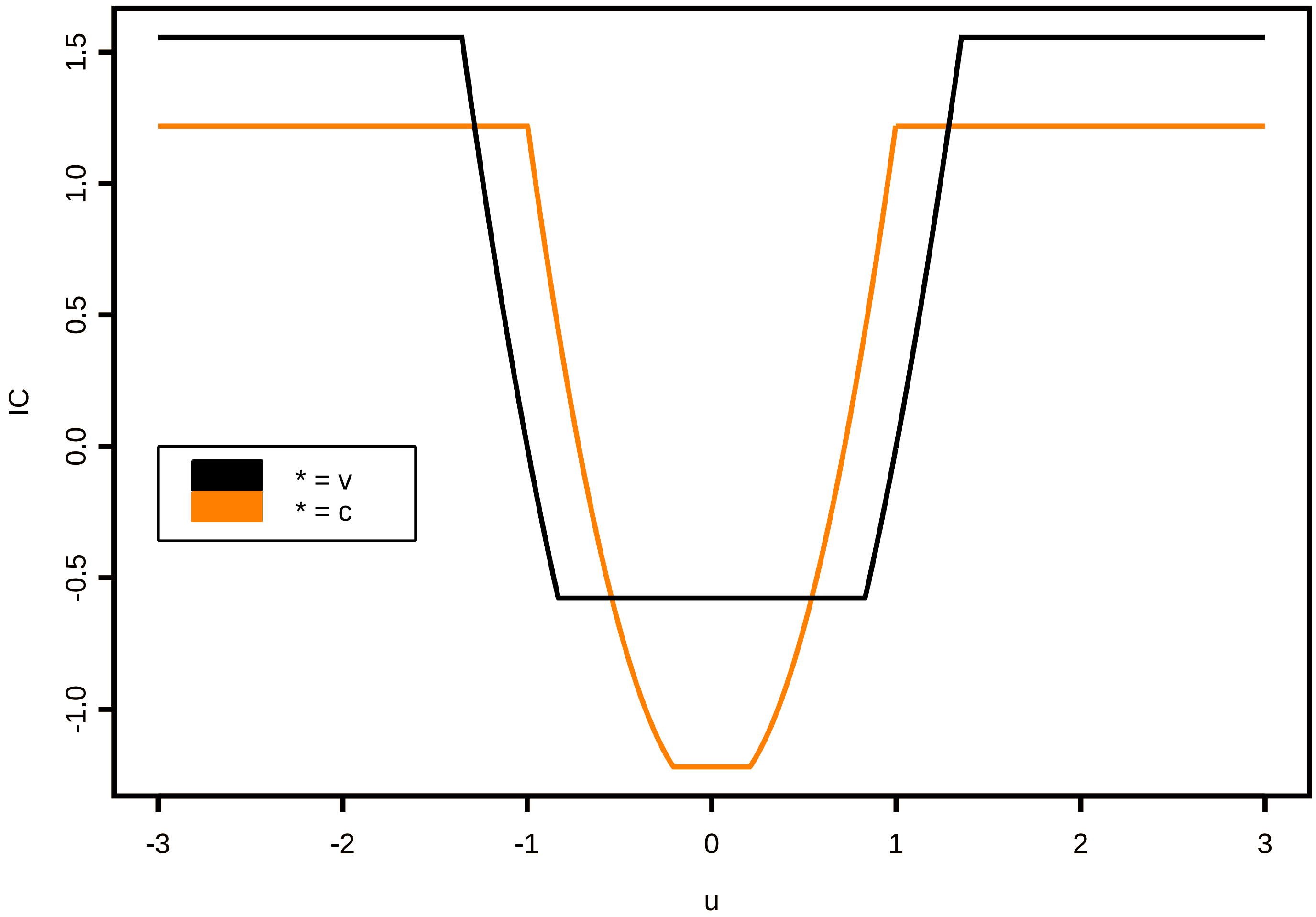
1-Dimensional Scale: IC-comparison for r=0.2 (*=c) with r=0.1 (*=v)



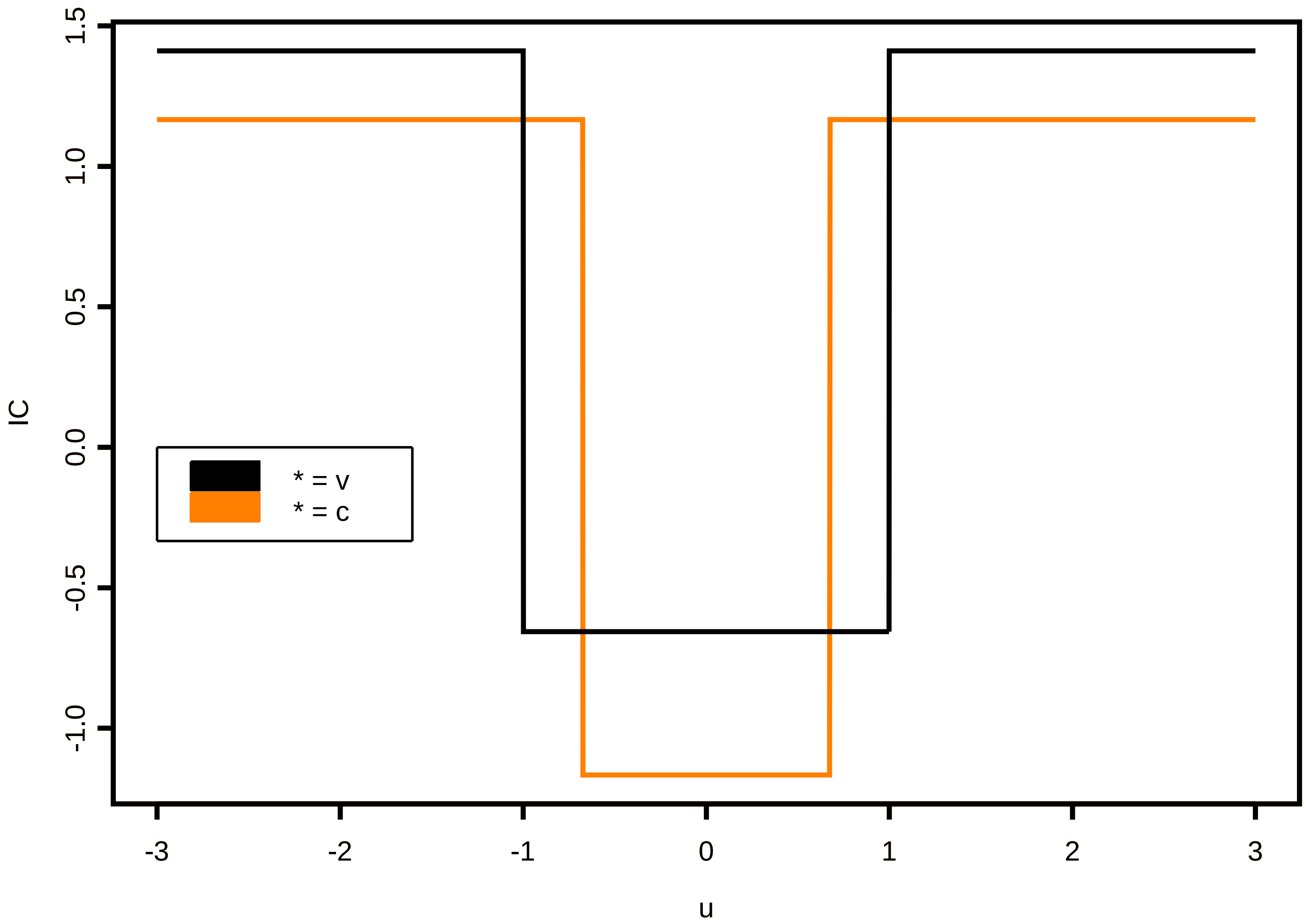


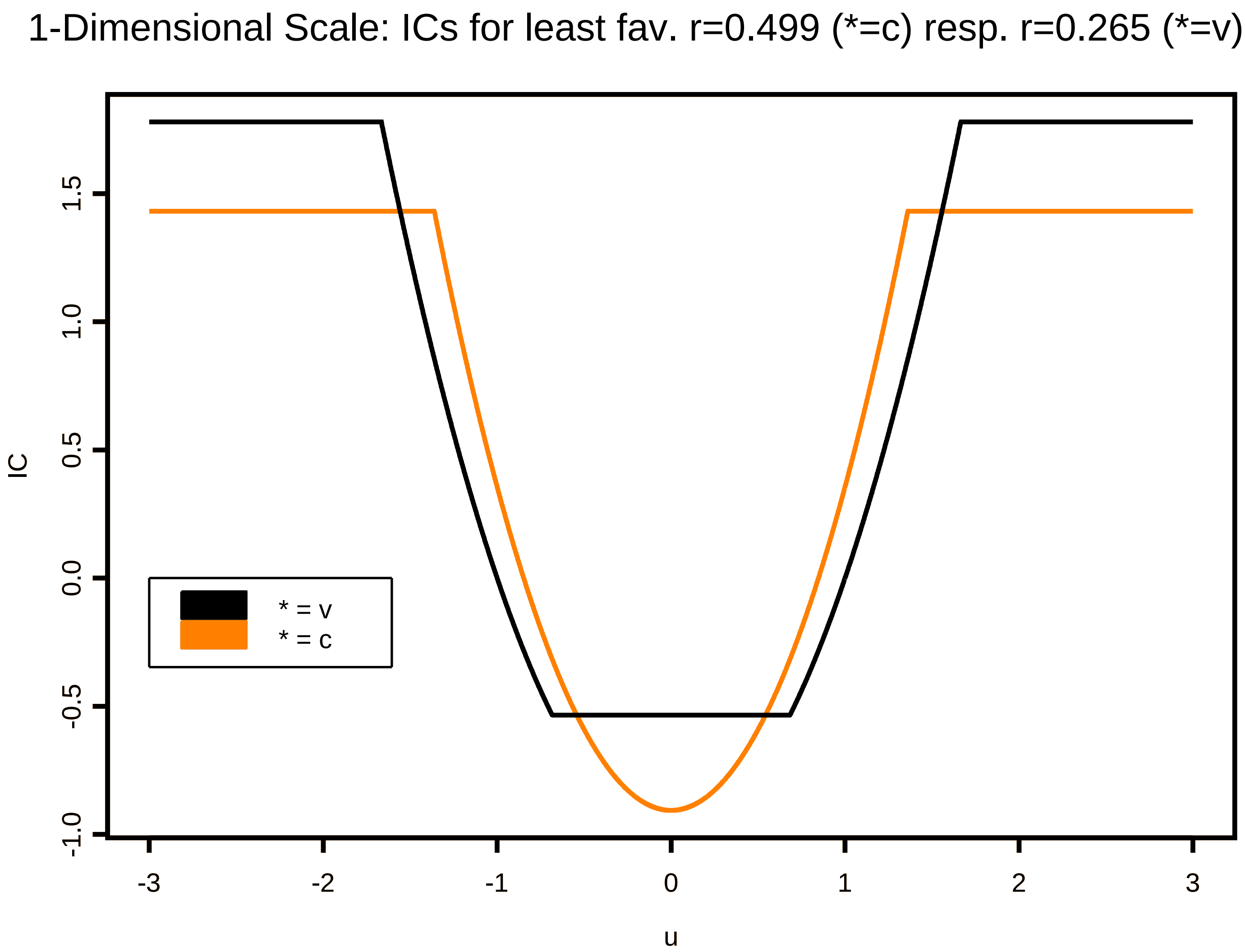



1-Dimensional Scale: IC-comparison for r=1.0 (*=c) with r=0.5 (*=v)



1-Dimensional Scale: IC-comparison for lower cases



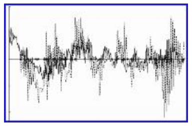




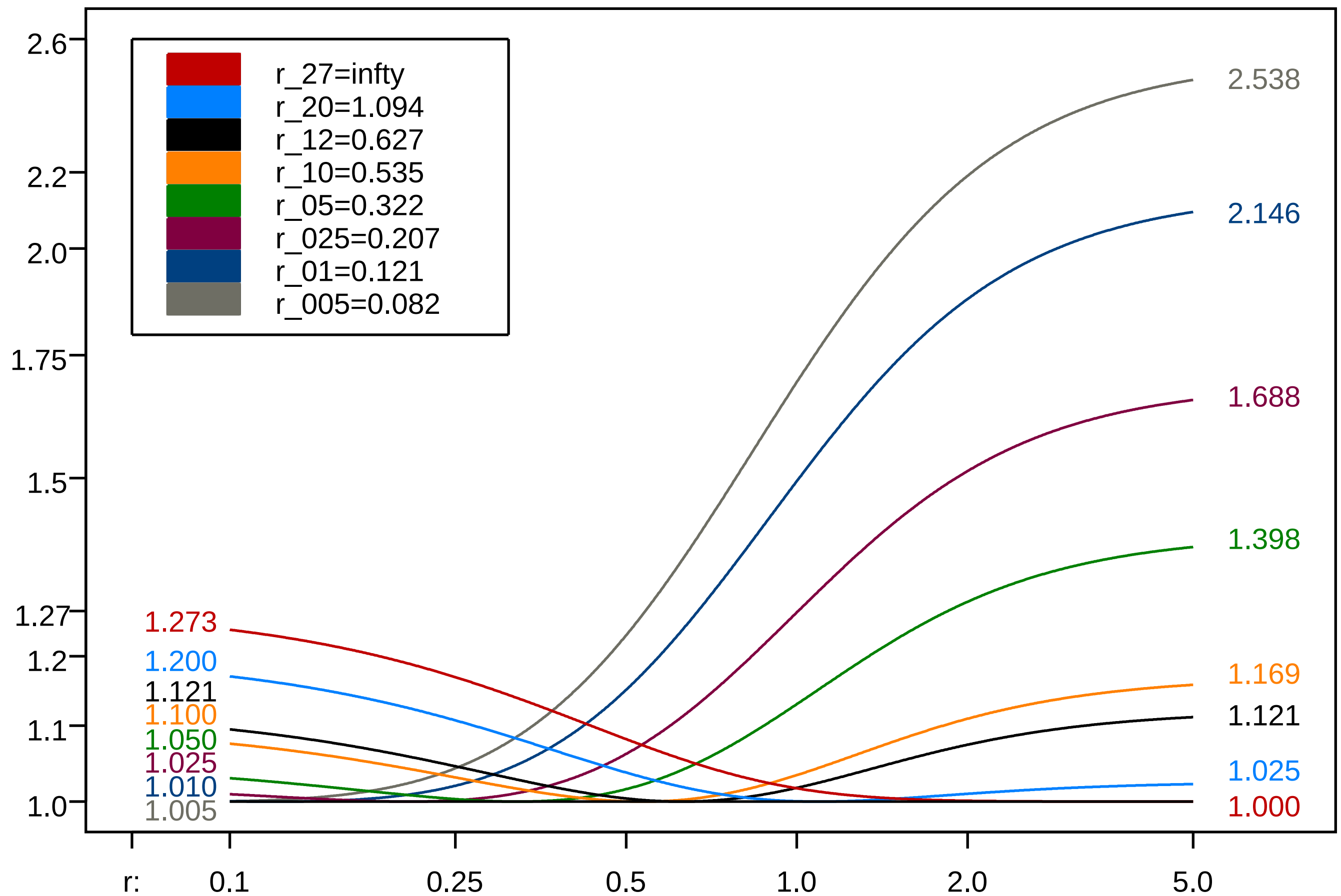

*Helmut Rieder*

**Robust Estimation for Time Series Models**

**06/12/2003**



relMSE: 2-Dimensional Location



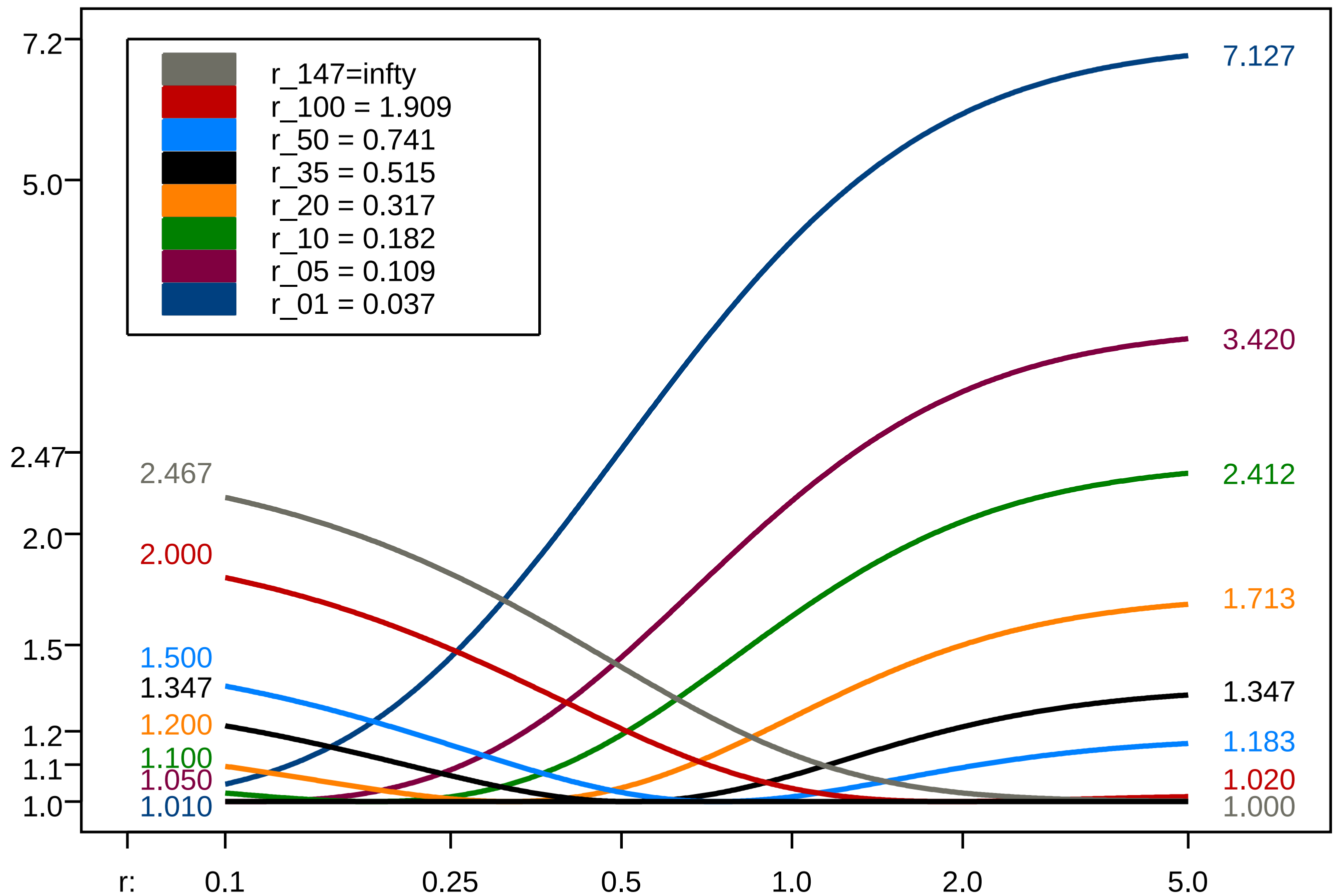

Mathematics VII

*Helmut Rieder*

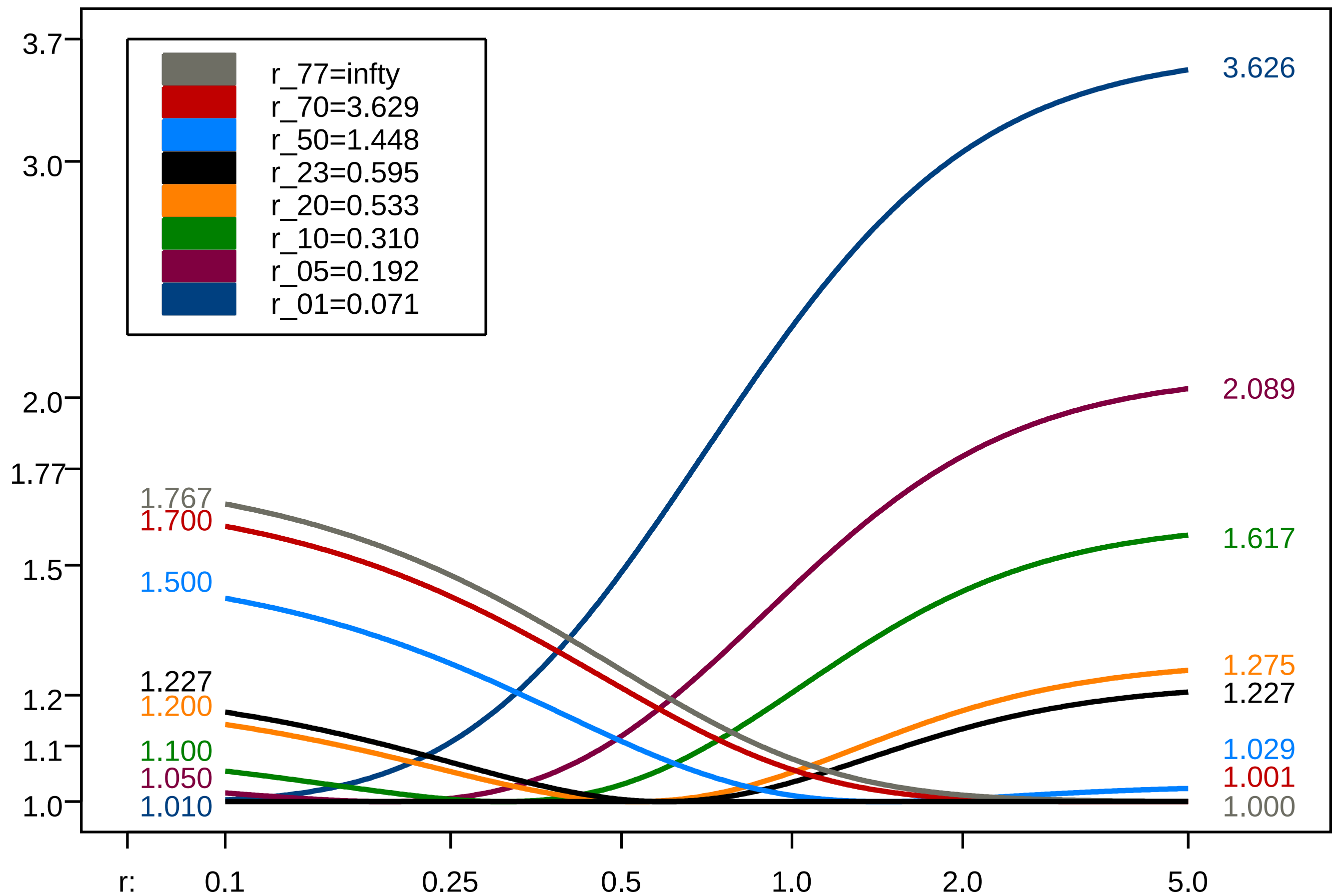
**Robust Estimation for Time Series Models**

**06/12/2003**

relMSE: Regression (*=c, alpha=1, K normal, dim=1)



relMSE: Regression (*=c, alpha=1, K uniform, dim=2)



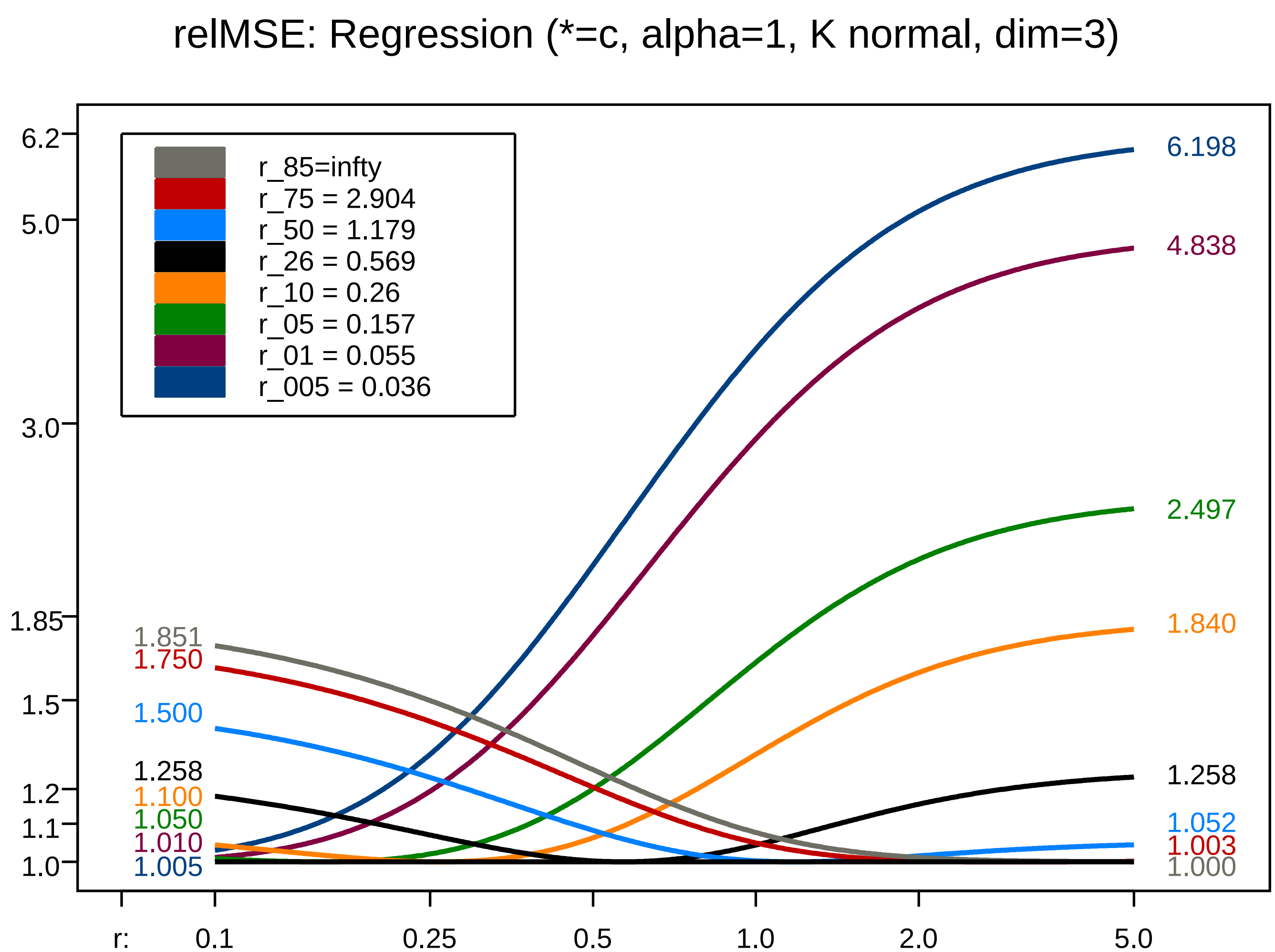

UNIVERSITÄT BAYREUTH

Mathematics VII

*Helmut Rieder*

**Robust Estimation for Time Series Models**

**06/12/2003**


relMSE: Regression (*=c, alpha=1, K normal, dim=3)
r_85=infty
r_75 = 2.904
r_50 = 1.179
r_26 = 0.569
r_10 = 0.26
r_05 = 0.157
r_01 = 0.055
r_005 = 0.036
6.2
5.0
3.0
1.85
1.5
1.2
1.1
1.0
1.851
1.750
1.500
1.258
1.100
1.050
1.010
1.005
6.198
4.838
2.497
1.840
1.258
1.052
1.003
1.000
r:
0.1
0.25
0.5
1.0
2.0
5.0