Estimation for ARMA Processes with Stable Noise

Matt Calder & Richard A. Davis
Colorado State University
rdavis@stat.colostate.edu

- ARMA processes with stable noise
- Review of M-estimation
- Examples of M-estimation
  - LS
  - LAD
  - MLE
  - Bootstrapping
- Simulations

ARMA(p,q) model with heavy tailed noise.

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

a. $\{Z_t\} \sim \text{IID}(\alpha)$ with Pareto tails ($0 < \alpha < 2$)

b. $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$

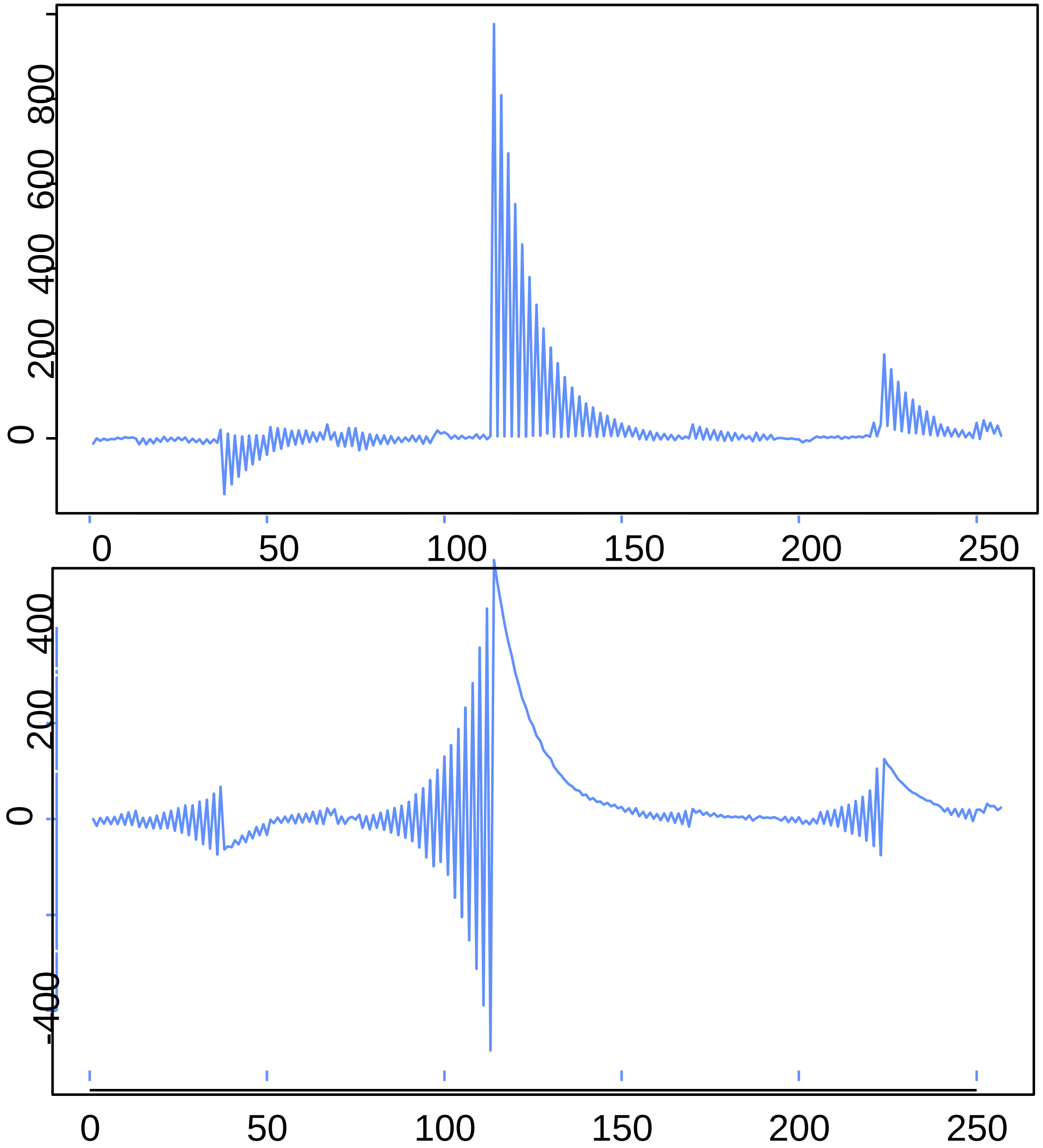
have no common zeroes and no zeroes on or inside the unit circle.

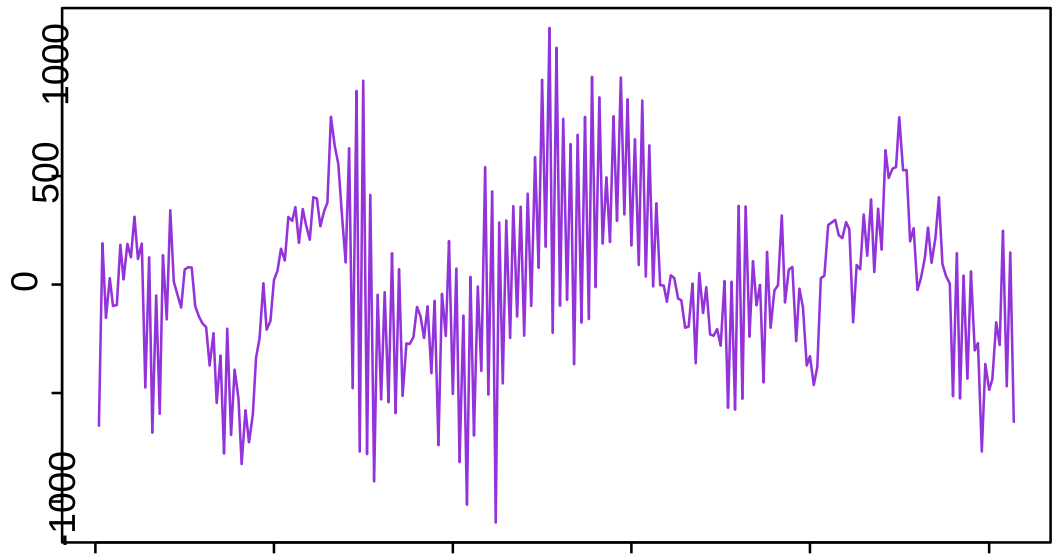
Shorthand notation.

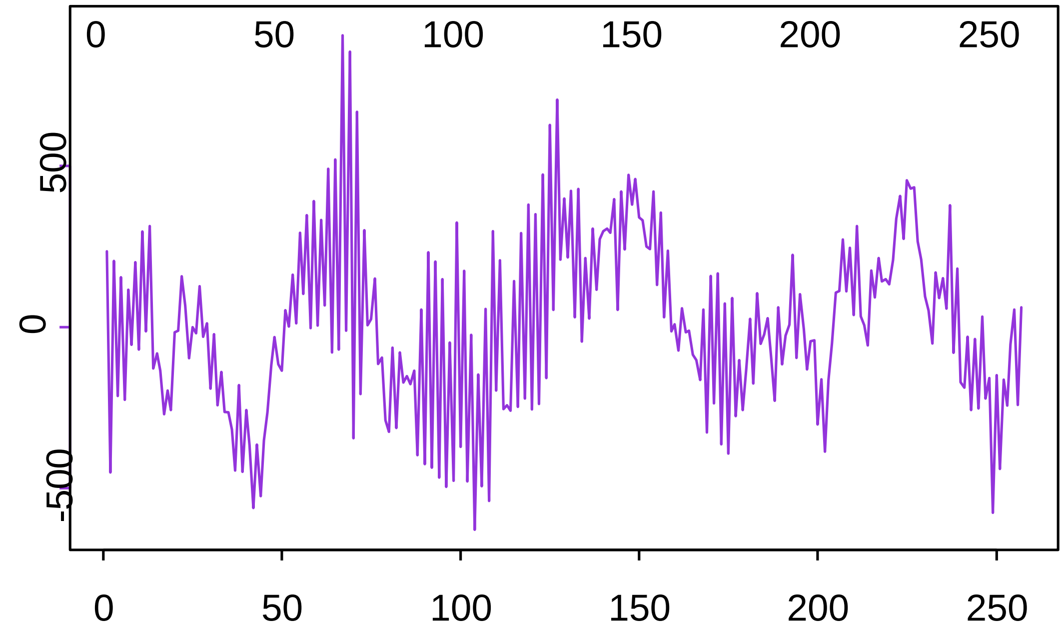
$\phi(B)X_t = \theta(B)Z_t$, $\{Z_t\} \sim \text{IID}(\alpha)$ (B=backward shift)

$\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)^T$ (parameter vector)

Examples of AR(2) models.




1000
500
0
-1000
0
50
100
150
200
250


500
0
-500
0
50
100
150
200
250

Review of M-Estimation for ARMA Models

Data. $X_1, \ldots, X_n$

Model. (ARMA(p,q)) $\phi(B)X_t = \theta(B)Z_t$, $\{Z_t\} \sim \mathrm{IID}(\alpha)$.

Loss function. $\rho(\,)$

Criterion. Minimize

$$T_n(\beta) = \sum_{t=1}^{n} \rho(Z_t(\beta)) \text{ with respect to } \beta,$$

where

$$Z_t(\beta) = 0 \text{ and } X_t = 0, \text{ for } t < 1,$$

$$Z_t(\beta) = \phi(B)X_t - \theta_1 Z_{t-1}(\beta) - \cdots - \theta_q Z_{t-q}(\beta), \quad t > 0.$$

Result (Davis, Knight & Liu `92, Davis `95).

If $\{Z_t\} \sim S\alpha S$ (symmetric α stable) and $\psi(x)= \rho'(x)$ satisfies

1. Lipschitz of order $\tau > \max(\alpha-1,0)$.
2. $E|\psi(Z_1)| < \infty$ if $\alpha < 1$,
3. $E\,\psi(Z_1) = 0$ and $Var(\psi(Z_1)) < \infty$, if $\alpha > 1$,

then

$$n^{1/\alpha}(\hat{\beta}_M - \beta) \xrightarrow{d} \eta$$

where $\hat{\beta}_M$ is the M-estimate of β. The limit random vector η is the minimizer of a stochastic process.

Example (MA(1)).

Data. $X_1, \ldots, X_n$

Model. $X_t = Z_t + \theta Z_{t-1}$,

$|\theta_0| < 1$, $\{Z_t\} \sim \text{IID}(\alpha)$

LAD estimation:

Minimize

$$T_n(\theta) = \sum_{t=1}^{n} \left(\rho(Z_t(\theta)) - \rho(Z_t(\theta_0)) \right)$$

$$= \sum_{t=1}^{n} \left(\rho(X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \cdots (-\theta)^{t-1} X_1) - \rho(Z_t(\theta_0)) \right)$$

Set $u = n^{1/\alpha}(\theta - \theta_0)$,

$$S_n(u) = T_n(\theta_0 + un^{-1/\alpha})$$

$$= \sum_{t=1}^{n} \left(\rho(Z_t(\theta_0 + un^{-1/\alpha})) - \rho(Z_t(\theta_0))\right)$$

(Not a convex function of u even if ρ is convex!)

Linearize $Z_t(\theta_0 + un^{-1/\alpha})$ to get

$$S_n(u) \sim \sum_{t=1}^{n} \left(\rho(Z_t(\theta_0) + un^{-1/\alpha} Z_t'(\theta_0)) - \rho(Z_t(\theta_0))\right)$$

where $-Z_t'(\theta_0)$ is the AR(1) process

$$Y_t = -\theta_0 Y_{t-1} + Z_t .$$

Result : $\hat{u}_n := \operatorname{argmin}(S_n(u))$

$$= n^{1/\alpha}(\hat{\theta}_M - \theta_0) \xrightarrow{d} \hat{\eta} := \operatorname{argmin}(S(u))$$

M-Estimation Examples

1. LS (least squares). $\rho(x) = x^2$ does not satisfy assumptions 1–3 of previous slide. However,

$$(n / \ln n)^{1/\alpha} (\hat{\beta}_{LS} - \beta) \xrightarrow{d} \eta_{LS}$$

Remark: Estimation procedures which are inherently second order based will have scaling factor $(n / \ln n)^{1/\alpha}$. Examples are

- moment estimation (Davis & Resnick `85, `86)
- Yule-Walker estimation for AR's (Davis & Resnick `85, `86)
- Whittle estimate (and max Gaussian likelihood?) (Mikosch, Gadrich, Kluppelberg and Adler `95)

Moreover,

$$\frac{\|\hat{\beta}_M - \beta\|}{\|\hat{\beta}_{LS} - \beta\|} \xrightarrow{p} 0$$

2. LAD (least absolute deviations). $\rho(x) = |x|$ does not satisfy assumptions 1–3 of previous slide either. However,

$$n^{1/\alpha}(\hat{\beta}_{LAD} - \beta) \xrightarrow{d} \eta_{LAD}$$

3. MLE (maximum likelihood). Suppose Z_t has pdf f and $\rho(x) = -\ln f(x)$. Then $\hat{\beta}_{MLE}$, which minimizes

$$T_n(\beta) = \sum_{t=1}^{n} -\ln f\big(Z_t(\beta)\big),$$

is an *approximate* MLE estimator. If one chooses f to be the symmetric λ–stable density f_λ, then $\rho(x) = -\ln f_\lambda(x)$ satisfies the assumptions of the result mentioned previously so that

$$n^{1/\alpha}(\hat{\beta}_{MLE,\lambda} - \beta) \xrightarrow{d} \eta_\lambda$$

Call $\beta_{MLE,\lambda}$ the maximum (λ-stable) likelihood estimate.

Remark. Can minimize

$$T_n(\beta) = \sum_{t=1}^{n} -\ln f_{\lambda}(Z_t(\beta))$$

with respect to both λ and β to obtain pseudo-MLE's of both parameters.

Bootstrapping the M-Estimate (Davis and Wu `95).

Data. $X_1, \ldots, X_n$

Model. $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$, $\{Z_t\} \sim \text{IID}(\alpha)$

M-estimate. $\hat{\phi}$

Estimated residuals. $\hat{Z}_t = X_t - \hat{\phi}_1 X_{t-1} + \cdots - \hat{\phi}_p X_{t-p}$

Bootstrap sample. $X_t^* = \hat{\phi}_1 X_{t-1}^* + \cdots + \hat{\phi}_p X_{t-p}^* + Z_t^*$

for $t = 1, \ldots, m_n$, where $\{Z_t^*\} \sim \text{IID}(F_n)$, F_n = empirical df of $\hat{Z}_{p+1}, \ldots, \hat{Z}_n$.

BS M-estimate. $\hat{\phi}^*$

Result. If $m_n / n \longrightarrow 0$, then

$$P(m_n^{1/\alpha}(\hat{\phi}^* - \hat{\phi}) \in \bullet \mid \mathbf{X}_n) \xrightarrow{p} P(\hat{\eta} \in \bullet) .$$

Removing the dependence on normalizing constants.

Let

$$M_n = \max\{|X_1|, \ldots, |X_n|\}$$

$$M_m^* = \max\{|X_1^*|, \ldots, |X_m^*|\}.$$

Then

$$M_n(\hat{\phi} - \phi) \xrightarrow{d} \hat{w}$$

and

$$P(M_m^*(\hat{\phi}^* - \hat{\phi}) \in \bullet \mid \mathbf{X}_n) \xrightarrow{p} P(\hat{w} \in \bullet) .$$

Simulation Comparison.

Principal objectives:

- compare performance of $\hat{\beta}_{MLE,\alpha}$, $\hat{\beta}_{LAD}$, and $\hat{\beta}_{LS}$.
- compare performance of $\hat{\beta}_{MLE,\alpha}$, $\hat{\beta}_{MLE,1}$, $\hat{\beta}_{MLE,\hat{\alpha}}$, and $\hat{\beta}_{LAD}$.
- investigate performance of the MLE estimator of α.

3 Models ($\{Z_t\} \sim S\alpha S$)

AR(1) : $X_t = .4\, X_{t-1} + Z_t$

MA(1): $X_t = Z_t + .8\, Z_{t-1}$

ARMA(1,1): $X_t = .4\, X_{t-1} + Z_t + .8\, Z_{t-1}$

Sample size = 200, replications = 10,000

$\alpha = 1.75$

Model	True Values	$\hat{\beta}_{LAD}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{MLE,\hat{\alpha}}$
M.1	$\phi = .4$	.397 (.0465)	.397(.0474)	.398(.0394)
M.2	$\theta = .8$	.802 (.0317)	.803(.0330)	.803(.0270)
M.3	$\phi = .4$	.389 (.0456)	.397(.0525)	.398(.0440)
	$\theta = .8$	.807 (.0353)	.804(.0363)	.803(.0296)

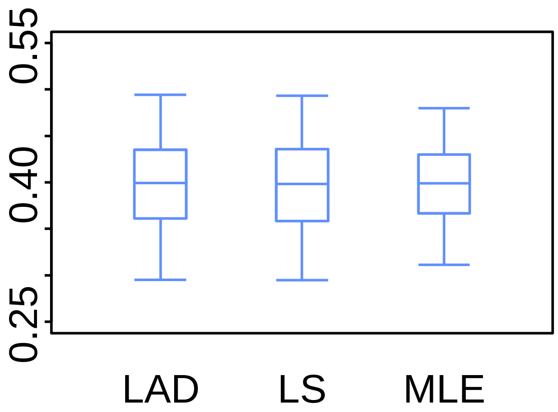
$\alpha = 1.0$

Model	True Values	$\hat{\beta}_{LAD}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{MLE,\hat{\alpha}}$
M.1	$\phi = .4$	.3995 (.0073)	.3971(.0263)	.3999(.0061)
M.2	$\theta = .8$	.8000 (.0043)	.8005(.0207)	.7997(.0039)
M.3	$\phi = .4$	.3997 (.0083)	.3979(.0320)	.3999(.0071)
	$\theta = .8$	.8005 (.0048)	.8012(.0232)	.7996(.0048)

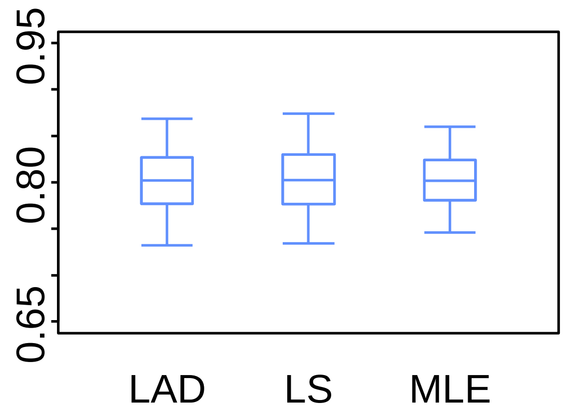
$\alpha = .50$

Model	True Values	$\hat{\beta}_{LAD}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{MLE,\hat{\alpha}}$
M.1	$\phi = .4$	.3997 (.00058)	.3988(.01022)	.4000(.00007)
M.2	$\theta = .8$	.7978 (.00226)	.8004(.01052)	.7576(.04242)
M.3	$\phi = .4$	.3988 (.00142)	.3986(.01396)	.3998(.00071)
	$\theta = .8$	.8000 (.00012)	.8014(.01179)	.7529(.04715)

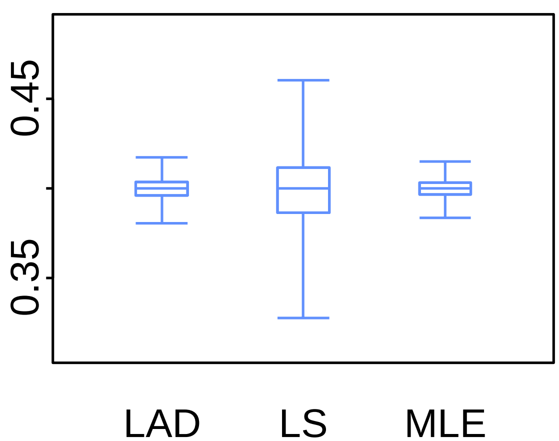
M.1, ϕ α=1.75



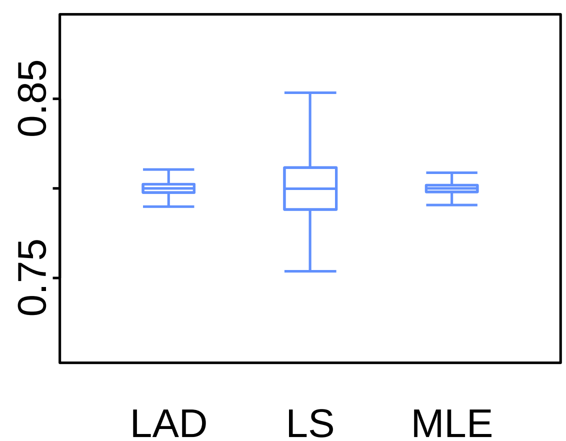
M.2, θ α=1.75



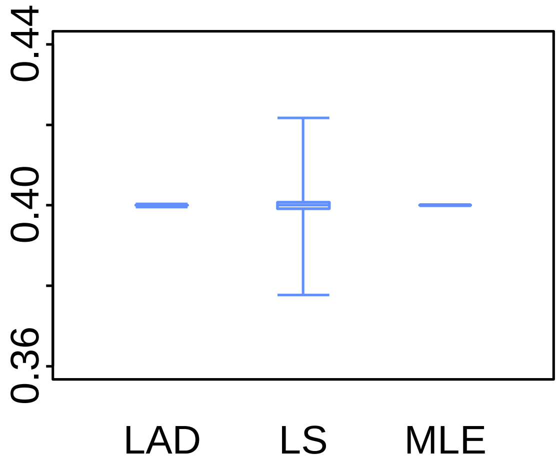
M.1, ϕ α=1.0



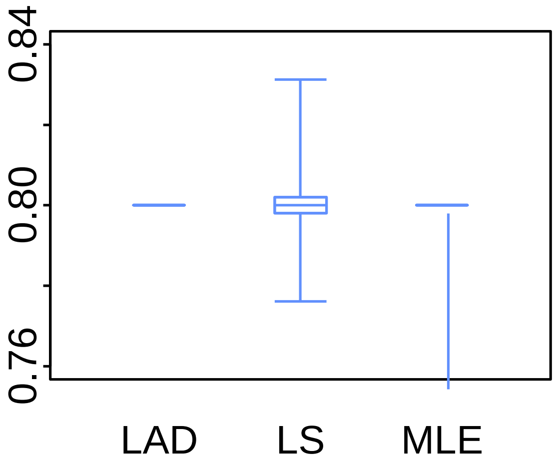
M.2, θ α=1.0



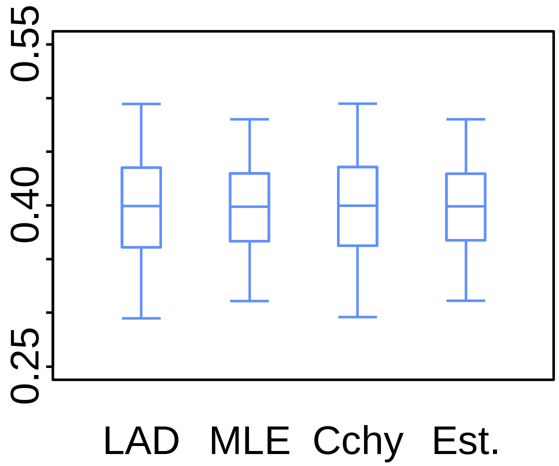
M.1, ϕ, α=0.5



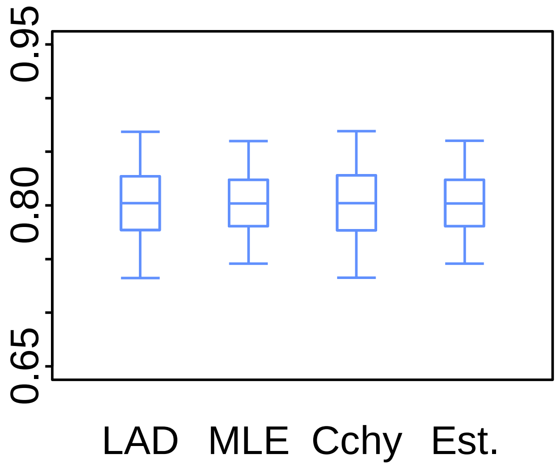
M.2, θ α=0.5



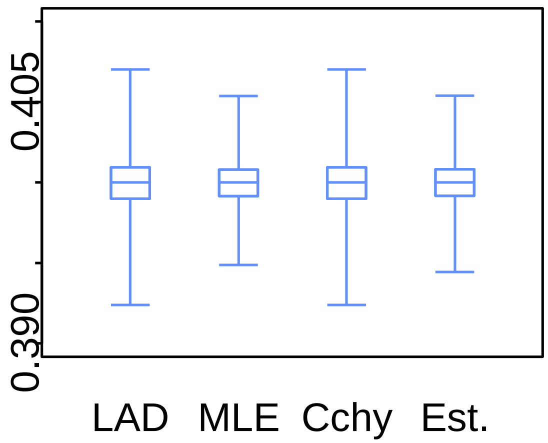
M.1, ϕ α=1.75



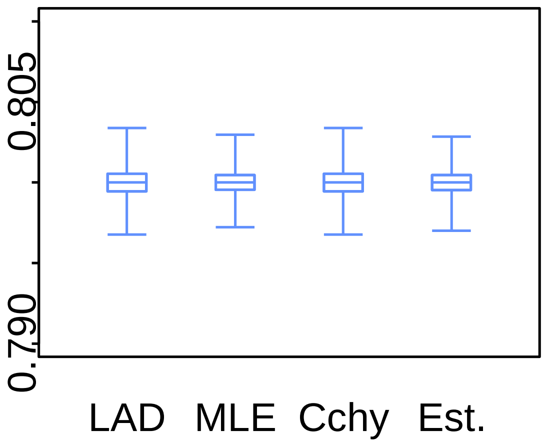
M.2, θ α =1.75



M.1, ϕ α=0.8

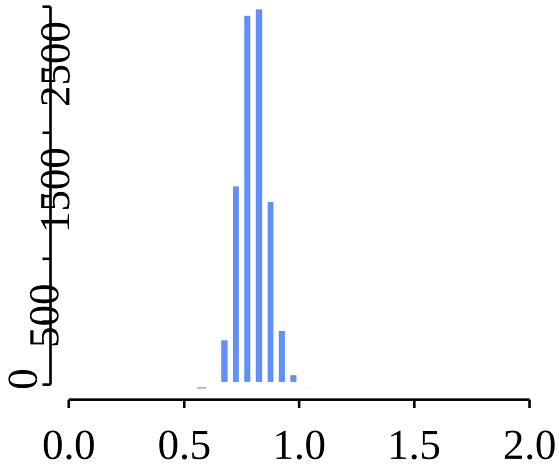


M.2, θ α=0.8

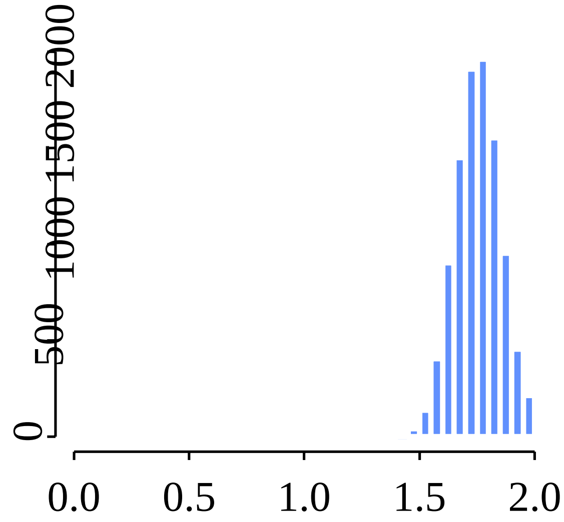


Estimation of α

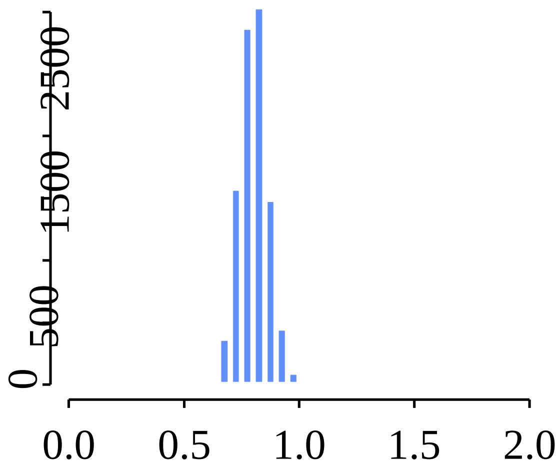
M.1, ϕ   α = 0.80



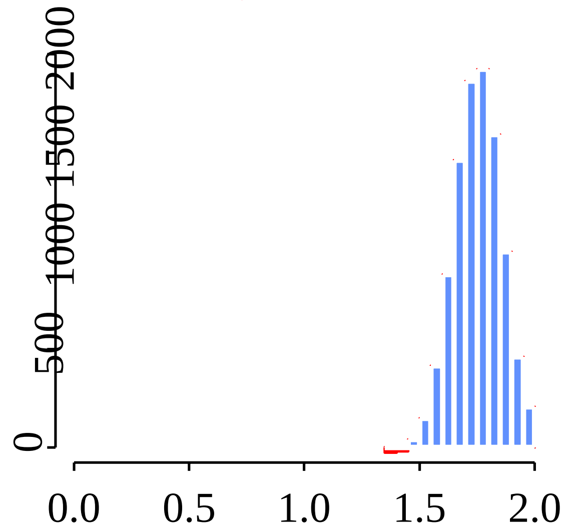
M1, ϕ   α = 1.75



M.2, θ α = 0.80

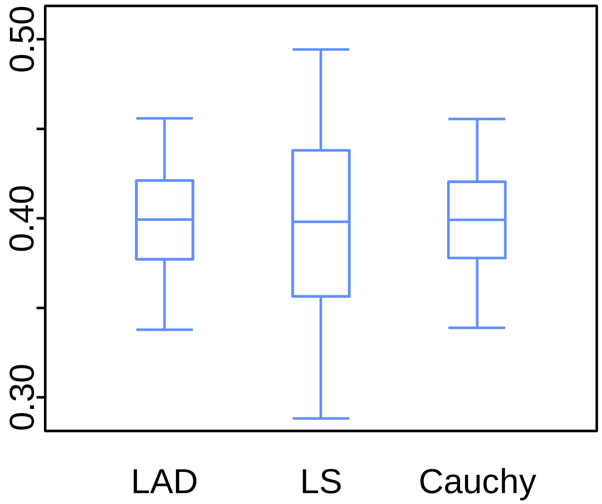


M.2, θ α = 1.75

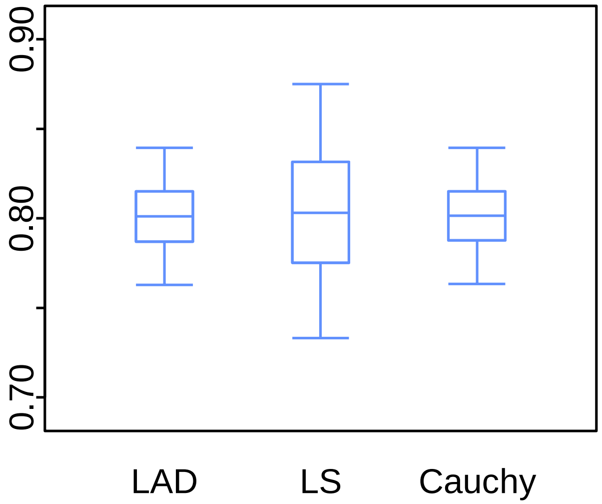


Robustness of the estimates. Z_1 has a Pareto tail with exponent $\alpha = 4.0$.

M.1, ϕ $\alpha = 4.0$



M.2, θ $\alpha = 4.0$



Conclusions

- Parameters of an ARMA model with heavy tailed noise can be estimated quite well.
- LAD estimates almost as good as MLE when noise is stable.
- Drawbacks of MLE:
  - computationally more difficult
  - requires an estimate of α
- MLE estimation of α performs well but is not very robust.