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.**

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

\[ \beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)^T \quad (\text{parameter vector}) \]
Examples of AR(2) models.
Review of M-Estimation for ARMA Models

Data. $X_1, \ldots, X_n$

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

Loss function. $\rho(\cdot)$

Criterion. Minimize

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

where

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

$Z_t(\beta) = \phi(B)X_t - \theta_1 Z_{t-1}(\beta) - \cdots - \theta_q Z_{t-q}(\beta)$, $t > 0$. 
**Result** (Davis, Knight & Liu `92, Davis `95).

If \( \{Z_t\} \sim S\alpha S \) (symmetric \( \alpha \) 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 \( \text{Var}(\psi(Z_1)) < \infty \), if \( \alpha > 1 \),

then

\[ n^{1/\alpha} (\hat{\beta}_M - \beta) \overset{d}{\rightarrow} \eta \]

where \( \hat{\beta}_M \) is the M-estimate of \( \beta \). The limit random vector \( \eta \) is the minimizer of a stochastic process.
Example (MA(1)).

Data. \( X_1, \ldots, X_n \)

Model. \( X_t = Z_t + \theta Z_{t-1}, \quad |\theta_0| < 1, \{Z_t\} \sim \text{IID}(\alpha) \)

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 \( u = n^{1/\alpha}(\theta - \theta_0), \)
\[ S_n(u) = T_n \left( \theta_0 + un^{-1/\alpha} \right) \]
\[ = \sum_{t=1}^{n} \left( \rho(Z_t (\theta_0 + un^{-1/\alpha})) - \rho(Z_t (\theta_0)) \right) \]

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

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

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

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

\[ Y_t = -\theta_0 Y_{t-1} + Z_t \]

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

\[ = n^{1/\alpha} (\hat{\theta}_M - \theta_0) \quad \text{d} \rightarrow \quad \hat{\eta} := \text{argmin}(S(u)) \]
M-Estimation Examples

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

\[
\left( \frac{n}{\ln n} \right)^{1/\alpha} (\hat{\beta}_{LS} - \beta) \xrightarrow{d} \eta_{LS}
\]

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,

\[
\frac{\|\hat{\beta}_M - \beta\|}{\|\hat{\beta}_{LS} - \beta\|} \overset{p}{\to} 0
\]

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

\[
n^{1/\alpha} (\hat{\beta}_{LAD} - \beta) \overset{d}{\to} \eta_{LAD}
\]
3. **MLE (maximum likelihood).** Suppose $Z_t$ has pdf $f$ and 
\[ \rho(x) = -\ln f(x). \] 
Then $\hat{\beta}_{\text{MLE}}$, which minimizes 
\[ T_n(\beta) = \sum_{t=1}^{n} -\ln f(Z_t(\beta)), \] 
is an approximate MLE estimator. If one chooses $f$ to be 
the symmetric $\lambda$–stable density $f_\lambda$, then $\rho(x) = -\ln f_\lambda(x)$ 
satisfies the assumptions of the result mentioned previously so that 
\[ n^{1/\alpha} (\hat{\beta}_{\text{MLE},\lambda} - \beta) \xrightarrow{d} \eta_\lambda \] 
Call $\beta_{\text{MLE},\lambda}$ the maximum ($\lambda$-stable) likelihood estimate.
Remark. Can minimize

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

with respect to both \( \lambda \) and \( \beta \) 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 \( \{Z_t^*\} \sim \text{IID}(F_n) \), \( F_n = \text{empirical df of} \)

\( \hat{Z}_{p+1}, \ldots, \hat{Z}_n \).

BS M-estimate. \( \hat{\phi}^* \)
\textbf{Result.} If \( m_n / n \rightarrow 0 \), then

\[
P(m_n^{1/\alpha} (\hat{\phi}^* - \bar{\phi}) \in \bullet \mid X_n) \xrightarrow{p} P(\hat{\eta} \in \bullet).\]

Removing the dependence on normalizing constants.

Let

\[
M_n = \max\{|X_1|, \ldots, |X_n|\}
\]

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

Then

\[M_n (\hat{\phi} - \bar{\phi}) \xrightarrow{d} \hat{w}\]

and

\[
P(M_m^* (\hat{\phi}^* - \bar{\phi}) \in \bullet \mid X_n) \xrightarrow{p} P(\hat{w} \in \bullet).
\]
Simulation Comparison.

Principal objectives:

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

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

AR(1): $X_t = 0.4 X_{t-1} + Z_t$

MA(1): $X_t = Z_t + 0.8 Z_{t-1}$

ARMA(1,1): $X_t = 0.4 X_{t-1} + Z_t + 0.8 Z_{t-1}$

Sample size = 200, replications = 10,000
### α = 1.75

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

### α = 1.0

<table>
<thead>
<tr>
<th>Model</th>
<th>True Values</th>
<th>$\hat{\beta}_{LAD}$</th>
<th>$\hat{\beta}_{LS}$</th>
<th>$\hat{\beta}_{MLE, \alpha}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.1</td>
<td>$\phi = .4$</td>
<td>.3995 (.0073)</td>
<td>.3971 (.0263)</td>
<td>.3999 (.0061)</td>
</tr>
<tr>
<td>M.2</td>
<td>$\theta = .8$</td>
<td>.8000 (.0043)</td>
<td>.8005 (.0207)</td>
<td>.7997 (.0039)</td>
</tr>
<tr>
<td>M.3</td>
<td>$\phi = .4$</td>
<td>.3997 (.0083)</td>
<td>.3979 (.0320)</td>
<td>.3999 (.0071)</td>
</tr>
<tr>
<td></td>
<td>$\theta = .8$</td>
<td>.8005 (.0048)</td>
<td>.8012 (.0232)</td>
<td>.7996 (.0048)</td>
</tr>
</tbody>
</table>
\( \alpha = .50 \)

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

M.1, $\phi \alpha = 0.80$

M.1, $\phi \alpha = 1.75$

M.2, $\theta \alpha = 0.80$

M.2, $\theta \alpha = 1.75$
Robustness of the estimates. $Z_1$ has a Pareto tail with exponent $\alpha = 4.0$.

M.1, $\phi$ $\alpha = 4.0$

M.2, $\theta$ $\alpha = 4.0$
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 $\alpha$
• MLE estimation of $\alpha$ performs well but is not very robust.