Laplace Likelihood and LAD Estimation for Non-invertible MA(1)

F. Jay Breidt
Richard A. Davis
Nan-Jung Hsu
Murray Rosenblatt

Colorado State University
National Tsing-Hua University
U. of California, San Diego

(http://www.stat.colostate.edu/~rdavis/lectures)
Program

➤ Introduction
  • The MA(1) unit root problem
  • Why study non-invertible MA(1)’s?
    ▪ over-differencing
    ▪ random walk + noise

➤ Gaussian Likelihood Estimation
  • Identifiability
  • Limit results
  • Extensions
    ▪ non-zero mean
    ▪ heavy tails

➤ Laplace Likelihood/LAD estimation
  • Joint and exact likelihood
  • Limit results
  • Limit distribution/simulation comparisons
  • Pile-up probabilities
    ▪ joint likelihood
    ▪ exact likelihood
MA(1) unit root problem

MA(1):

\[ Y_t = Z_t - \theta Z_{t-1}, \quad \{Z_t\} \sim \text{IID } (0, \sigma^2) \]

Properties:

- \(|\theta| < 1 \Rightarrow Z_t = \sum_{j=0}^{\infty} \theta^j Y_{t-j} \) (invertible)
- \(|\theta| > 1 \Rightarrow Z_t = -\sum_{j=1}^{\infty} \theta^{-j} Y_{t+j} \) (non-invertible)
- \(|\theta| = 1 \Rightarrow Z_t \in \text{sp}\{Y_t, Y_{t-1}, \ldots,\} \quad \text{and} \quad Z_t \in \text{sp}\{Y_{t+1}, Y_{t+2}, \ldots,\} \Rightarrow P_{\text{sp}\{Y_s, s \neq 0\}} Y_0 = Y_0 \) (perfect interpolation)
- \(|\theta| < 1 \Rightarrow \hat{\theta}_{mle} \quad \text{is} \quad \text{AN}(\theta, (1 - \theta^2) / n) \)

MLE = maximum (Gaussian) likelihood, \( n = \text{sample size} \)

What if \( \theta = 1 \)?
Why study non-invertible MA(1)?

a) over-differencing

• linear trend model: \( X_t = a + bt + Z_t \).
  \[ Y_t = X_t - X_{t-1} = b + Z_t - Z_{t-1} \sim \text{MA}(1) \text{ with } \theta = 1. \]

• seasonal model: \( X_t = s_t + Z_t \), \( s_t \) seasonal component w/ period 12.
  \[ Y_t = X_t - X_{t-12} = Z_t - Z_{t-12} \sim \text{MA}(12) \text{ with } \theta = 1. \]

b) random walk + noise

\[ X_t = X_{t-1} + U_t \quad \text{(random walk signal)} \]
\[ Y_t = X_t + V_t \quad \text{(random walk signal + noise)} \]

Then
\[ Y_t - Y_{t-1} = U_t + V_t - V_{t-1} \sim \text{MA}(1) \]

with \( \theta = 1 \) if and only if \( \text{Var}(U_t) = 0. \)
Identifiability and Gaussian likelihood

Identifiability

• $|\theta| > 1 \Rightarrow Y_t = \epsilon_t - \theta^{-1} \epsilon_{t-1}$, where $\{\epsilon_t\} \sim \text{WN}(0, \theta^2 \sigma^2)$.

• $\{\epsilon_t\}$ is IID if and only if $\{Z_t\}$ is Gaussian (Breidt and Davis `91)

Gaussian Likelihood

$$L_G(\theta, \sigma^2) = L_G(1/\theta, \theta^2 \sigma^2) \Rightarrow \theta \text{ is only identifiably for } |\theta| \leq 1.$$ 

Notes:

i) this implies $L_G(\theta) = L_G(1/\theta)$ for the profile likelihood and $\theta = 1$ is a critical point, $L'_G(1) = 0$.

ii) a pile-up effect ensues, i.e.,

$$P(\hat{\theta} = 1) > 0$$

even if $\theta < 1$. 
Gaussian likelihood examples

100 observations from \( Y_t = Z_t - \theta_0 Z_{t-1} \), \( \{Z_t\} \sim \text{IID } (0, \sigma^2) \), Laplace pdf

\[ \theta_0 = 0.8 \quad \theta_0 = 1.0 \quad \theta_0 = 1.25 \]
MLE (Gaussian likelihood)

Idea: build parameter normalization into the likelihood function.

Model:  
\[ Y_t = Z_t - (1 - \frac{\beta}{n}) Z_{t-1}, \quad t = 1, \ldots, n. \]
\[ \beta = n(1 - \theta), \quad \theta = 1 - \frac{\beta}{n}, \quad \theta_0 = 1 - \frac{\gamma}{n} \]

Gaussian Likelihood:
\[ L_n(\beta) = l_n(1 - \frac{\beta}{n}) - l_n(1), \quad l_n(\ ) = \text{profile log-like}. \]

Theorem (Davis and Dunsmuir `96): Under \( \theta_0 = 1 - \frac{\gamma}{n} \),
\[ L_n(\beta) \to_d Z_{\gamma}(\beta) \quad \text{on} \quad C[0, \infty). \]

Results:

- \[ n(1 - \hat{\theta}_{mle}) \to \hat{\beta}_{mle} = \text{argmax } Z_{\gamma}(\beta) \]
- \[ n(1 - \hat{\theta}_{lm}) \to \hat{\beta}_{lm} = \text{arglocalmax } Z_{\gamma}(\beta) \]
- \[ P(\hat{\theta}_{lm} = 1) \to P(\hat{\beta}_{lm} = 0) = 0.6518 \quad \text{if } \gamma = 0. \]
Extensions of MLE (Gaussian likelihood)

i) non-zero mean (Chen and Davis `00): same type of limit, except pile-up is more excessive.

\[ P(\hat{\theta}_{mle} = 1) \to .955 \]

This makes hypothesis testing easy!

Reject \( H_0: \theta = 1 \) if \( \hat{\theta}_{mle} < 1 \) \hspace{1cm} (size of test is .045)

ii) heavy tails (Davis and Mikosch `98): \( \{Z_t\} \) symmetric alpha stable (S\(\alpha\)S). Then the max Gaussian likelihood estimator has the same normalizing rate, i.e.,

\[ n(1 - \hat{\theta}_{lm}) \to_d \hat{\beta}_{lm} \]

\[ P(\hat{\theta}_{lm} = 1) \to P(\hat{\beta}_{lm} = 0) \]

The pile-up decreases with increasing tail heaviness.
Comparison of limit cdf’s for different $\alpha$’s.

- $\alpha = 2.0$
- $\alpha = 1.5$
- $\alpha = 1.0$
- $\alpha = .75$
If noise distribution is non-Gaussian, the MA(1) parameter $\theta$ is identifiable for all real values.

Q1. For MLE (non-Gaussian) does one have $n$ or $n^{1/2}$ asymptotics?

Q2. Is there a pile-up effect?
Model. $Y_t = Z_t - \theta Z_{t-1}$, \{Z_t\} \sim \text{IID} (0, \sigma^2)$ with median 0 and $EZ^4 < \infty$.

Initial variable.

$$Z_{\text{init}} = \begin{cases} Z_0, & \text{if } |\theta| \leq 1, \\ Z_n - \sum_{t=1}^{n} Y_t, & \text{otherwise.} \end{cases}$$

Joint density: Let $Y_n=(Y_1, \ldots, Y_n)$, then

$$f(y_n, z_{\text{init}}) = f(z_0, z_1, \ldots, z_n) \left(1_{\{|\theta| \leq 1\}} + |\theta|^{-n} 1_{\{|\theta| > 1\}}\right),$$

where the $z_t$ are solved

- forward by: $z_t = Y_t + \theta z_{t-1}$, $t=1, \ldots, n$ for $|\theta| \leq 1$ with $z_0 = z_{\text{init}}$
- backward by: $z_{t-1} = \theta^{-1}(z_t - Y_t)$, $t=n, \ldots, 1$ for $|\theta| > 1$ with $z_n = z_{\text{init}} + Y_1 + \ldots + Y_n$

Note: integrate out $z_{\text{init}}$ to get exact likelihood.

$$f(y_n) = \int_{-\infty}^{\infty} f(y_n, z_{\text{init}}) dz_{\text{init}}$$
100 observations from $Y_t = Z_t - \theta_0 Z_{t-1}$, $\{Z_t\} \sim$ IID $(0, \sigma^2)$, Laplace pdf

$\theta_0 = 1.0$

$\theta_0 = .8$
100 observations from $Y_t = Z_t - \theta_0 Z_{t-1}$, \{\{Z_t\}\} $\sim$ IID \((0, \sigma^2)\), Laplace pdf

$\theta_0 = 0.8 \quad \theta_0 = 1.0 \quad \theta_0 = 1.25$

Exact likelihood

Joint likelihood at $z_{\text{max}}(\theta)$
(Joint) Laplace log-likelihood.

\[ L(\theta, z_{\text{init}}, \sigma) = -(n + 1) \log 2\sigma - \sigma^{-1} \sum_{t=0}^{n} |z_t| - n(\log |\theta|)1_{\{|\theta| > 1\}} \]

Maximizing wrt \( \sigma \), we obtain

\[ \hat{\sigma} = \frac{\sum_{t=0}^{n} |z_t|}{n + 1} \]

so that maximizing \( L \) is equivalent to minimizing

\[ l_n(\theta, z_{\text{init}}) = \begin{cases} \sum_{t=0}^{n} |z_t|, & \text{if } |\theta| \leq 1, \\ \sum_{t=0}^{n} |z_t| |\theta|, & \text{otherwise}. \end{cases} \]
Theorem 1. Under the parameterizations,

\[ \theta = 1 + \frac{\beta}{n} \quad \text{and} \quad z_{\text{init}} = Z_0 + \alpha \sigma / \sqrt{n}, \]

we have

\[ U_n(\beta, \alpha) = \sigma^{-1} (l_n(\theta, z_{\text{init}}) - l_n(1, Z_0)) \to_d U(\beta, \alpha) \]

on \( C(R^2) \), where

\[ U_n(\beta, \alpha) = \int_0^1 \left( \beta \int_0^{s-} e^{\beta(s-t)} dS(t) + \alpha e^{\beta s} \right) dW(s) \]
\[ + f(0) \int_0^1 \left( \beta \int_0^s e^{\beta(s-t)} dS(t) + \alpha e^{\beta s} \right)^2 ds \]

for \( \beta \leq 0 \), and

\[ U_n(\beta, \alpha) = \int_0^1 \left( -\beta \int_{s+}^{1} e^{\beta(s-t)} dS(t) + \alpha e^{-\beta(1-s)} \right) dW(s) \]
\[ + f(0) \int_0^1 \left( \beta \int_s^1 e^{\beta(s-t)} dS(t) + \alpha e^{-\beta(1-s)} \right)^2 dW(s) \]

for \( \beta > 0 \), in which \( S(t) \) and \( W(t) \) are the limits of the partial sum processes.
Joint Laplace likelihood — limit results

\[
S_n(t) = \frac{1}{\sigma\sqrt{n}} \sum_{i=0}^{[nt]} Z_i \rightarrow_d S(t), \quad W_n(t) = \frac{1}{\sigma\sqrt{n}} \sum_{i=0}^{[nt]} \text{sign}(Z_i) \rightarrow_d W(t).
\]

From the limit,

\[U_n(\beta, \alpha) \rightarrow_d U(\beta, \alpha)\]

it follows that

\[(n(\hat{\theta}_{lm} - 1), \sqrt{n}\sigma^{-1}(\hat{z}_{init}^L - Z_0)) \rightarrow_d (\hat{\beta}_{lm}, \hat{\alpha}_{lm})\]

where

\[(\hat{\beta}_{lm}, \hat{\alpha}_{lm}) = \text{arg(local)} \min U(\beta, \alpha).\]
Exact Laplace Likelihood:

\[ L_n(\theta, \sigma) = \int_{-\infty}^{\infty} f(y_n, z_{init}) dz_{init} \]

Theorem 2. For the MLE \( \tilde{\theta}_n, \tilde{\sigma}_n \), we have

\[ (n(\tilde{\theta}_{mle} - 1), \sqrt{n}(\tilde{\sigma}_{mle} - E \mid Z_0 \mid)) \rightarrow_d (\tilde{\beta}_{mle}, N), \]

where

\[ \tilde{\beta}_{mle} = \arg \min U^*(\beta), \quad N \sim N(0, \text{var}(\mid Z_0 \mid)), \]

and \( U^*(\beta) \) is a stochastic process defined in terms of \( S(t) \) and \( W(t) \).

In addition,

\[ n(\tilde{\theta}_{lm} - 1) \rightarrow_d \tilde{\beta}_{lm}, \quad \tilde{\beta}_{lm} = \arg \text{(local)} \min U^*(\beta). \]
Simulating from the limit process

Step 1. Simulate two indep sequences \((W_1, \ldots, W_m)\) and \((V_1, \ldots, V_m)\) of iid \(N(0,1)\) random variables with \(m=100000\).

Step 2. Form \(W(t)\) and \(V(t)\) by the partial sum processes,

\[
W(t) = \frac{[100000 \cdot t]}{\sqrt{100000}} \quad \text{and} \quad V(t) = \frac{[100000 \cdot t]}{\sqrt{100000}}.
\]

Step 3. Set \(S(t) = c_1 W(t) + c_2 V(t)\), where

\[
c_1 = E |Z_t| / \sigma \quad \text{and} \quad c_2 = \sqrt{\text{Var}(Z_t) / \sigma^2 - c_1^2}.
\]

Limit process depends only on \(c_1, c_2,\) and \(f(0)\).

Step 4. Compute \(U(\beta, \alpha)\) and \(U^*(\beta)\) from the definition.

Step 5. Determine the respective local and global minimizers of \(U(\beta, \alpha)\) and \(U^*(\beta)\) numerically.
Limit process

2 realizations of the limit processes, $U(\beta, \alpha(\beta))$, $U^*(\beta)$.

Laplace pdf

t(5) pdf
Limit distribution

red graph = Laplace pdf for $Z_t$
blue graph = Gaussian pdf for $Z_t$

Joint Lap Likelihood

$$n(\hat{\theta}_{lm} - 1) \rightarrow_d \hat{\beta}_{lm}$$

Exact Lap Likelihood

$$n(\tilde{\theta}_{lm} - 1) \rightarrow_d \tilde{\beta}_{lm}$$
red graph = Laplace pdf for $Z_t$

blue graph = Gaussian pdf for $Z_t$

Joint Lap Like

Exact Lap Like
### Laplace noise

\( \theta = 1, \ \sigma = 1 \)

1000 reps

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \hat{\theta}_{lm} )</th>
<th>( \hat{\theta}_{lm} )</th>
<th>( \hat{\sigma} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 20 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>-.0057</td>
<td>-.0033</td>
<td>-.0208</td>
</tr>
<tr>
<td>s.d.</td>
<td>.1438</td>
<td>.0656</td>
<td>.2430</td>
</tr>
<tr>
<td>rmse</td>
<td><strong>.1439</strong></td>
<td><strong>.0657</strong></td>
<td><strong>.2438</strong></td>
</tr>
<tr>
<td>asymp</td>
<td>.1207</td>
<td>.0526</td>
<td>.2236</td>
</tr>
<tr>
<td>( n = 50 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>.0000</td>
<td>.0004</td>
<td>.0293</td>
</tr>
<tr>
<td>s.d.</td>
<td>.0574</td>
<td>.0208</td>
<td>.1511</td>
</tr>
<tr>
<td>rmse</td>
<td><strong>.0574</strong></td>
<td><strong>.0208</strong></td>
<td><strong>.1539</strong></td>
</tr>
<tr>
<td>asymp</td>
<td>.0483</td>
<td>.0211</td>
<td>.1414</td>
</tr>
<tr>
<td>( n = 100 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>.0005</td>
<td>-.0003</td>
<td>-.0025</td>
</tr>
<tr>
<td>s.d.</td>
<td>.0303</td>
<td>.0107</td>
<td>.1000</td>
</tr>
<tr>
<td>rmse</td>
<td><strong>.0303</strong></td>
<td><strong>.0107</strong></td>
<td><strong>.1000</strong></td>
</tr>
<tr>
<td>asymp</td>
<td>.0241</td>
<td>.0105</td>
<td>.1000</td>
</tr>
<tr>
<td>( n = 200 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>.0005</td>
<td>.0000</td>
<td>-.0016</td>
</tr>
<tr>
<td>s.d.</td>
<td>.0140</td>
<td>.0058</td>
<td>.0718</td>
</tr>
<tr>
<td>rmse</td>
<td><strong>.0141</strong></td>
<td><strong>.0058</strong></td>
<td><strong>.0718</strong></td>
</tr>
<tr>
<td>asymp</td>
<td>.0121</td>
<td>.0053</td>
<td>.0707</td>
</tr>
</tbody>
</table>
Simulation results

Exact = MLE
Joint = maximize over $\theta$ and $z_{init}$
Cond = maximize over $\theta$ conditional on $z_{init} = 0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Exact $\tilde{\theta}_{ml}$</th>
<th>Joint $\hat{\theta}_{ml}$</th>
<th>Cond $\bar{\theta}_{ml}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>bias -.047</td>
<td>-.050</td>
<td>-.057</td>
</tr>
<tr>
<td></td>
<td>rmse .224</td>
<td>.213</td>
<td>.297</td>
</tr>
<tr>
<td>50</td>
<td>bias -.013</td>
<td>.002</td>
<td>-.026</td>
</tr>
<tr>
<td></td>
<td>rmse .096</td>
<td>.078</td>
<td>.171</td>
</tr>
<tr>
<td>100</td>
<td>bias .003</td>
<td>-.003</td>
<td>-.009</td>
</tr>
<tr>
<td></td>
<td>rmse .051</td>
<td>.034</td>
<td>.105</td>
</tr>
<tr>
<td>200</td>
<td>bias .000</td>
<td>.000</td>
<td>.007</td>
</tr>
<tr>
<td></td>
<td>rmse .028</td>
<td>.014</td>
<td>.070</td>
</tr>
</tbody>
</table>

Laplace noise
$\theta = 1, \sigma = 1$
1000 reps

Note:
- LM dominates ML
- joint dominates exact (rmse is half the size)
Theorem 3. (joint Laplace likelihood)

\[ P(\hat{\theta}_{lm} = 1) \rightarrow P(0 < Y < \int_{0}^{1} dS(s)dW(s)), \]

where

\[ Y = \int_{0}^{1} S(s)dW(s) - W(1)\int_{0}^{1} S(s)ds + \frac{W(1)}{2f(0)}(\int_{0}^{1} W(s)ds - W(1/2)) \]

Idea:

\[ P(\hat{\theta}_{lm} = 1) = P(\lim_{\beta \uparrow 0} \frac{\partial}{\partial \beta} U_n(\beta, \hat{\alpha}(\beta)) < 0 \text{ and } \lim_{\beta \downarrow 0} \frac{\partial}{\partial \beta} U_n(\beta, \hat{\alpha}(\beta)) > 0) \]

\[ \rightarrow P(\lim_{\beta \uparrow 0} \frac{\partial}{\partial \beta} U(\beta, \hat{\alpha}(\beta)) < 0 \text{ and } \lim_{\beta \downarrow 0} \frac{\partial}{\partial \beta} U(\beta, \hat{\alpha}(\beta)) > 0) \]

Now,

\[ \lim_{\beta \downarrow 0} \frac{\partial}{\partial \beta} U(\beta, \hat{\alpha}(\beta)) = Y \]

\[ \lim_{\beta \uparrow 0} \frac{\partial}{\partial \beta} U(\beta, \hat{\alpha}(\beta)) = Y - \int_{0}^{1} dS(s)dW(s) \quad \text{and the result follows.} \]
Theorem 4. (exact Laplace likelihood)

\[ P(\tilde{\theta}_{lm} = 1) \rightarrow P \left[ \frac{1}{2} < Y < \int_0^1 dS(s)dW(s) - \frac{1}{2} \right] \]

The pile-up probability is always smaller for the exact MLE than for the joint MLE (see Theorem 3).

Remark 1.

If \( Z_t \) has a Laplace density \( f(z) = \frac{1}{2\sigma} e^{-|z|/\sigma} \), then

\[ Y = \int_0^1 [W(1)s - W(s)] \, dV(s) + \frac{1}{2}. \]

where \( W(s) \) and \( V(s) \) are independent standard Brownian motions.
It follows that

\[ P(\hat{\theta}_{lm} = 1) \rightarrow P(0 < Y < \int_0^1 dS(s) dW(s)) \]

\[ = P(0 < \int_0^1 [W(1)s - W(s)] dV(s) + 0.5 < 1) \]

\[ = E \left[ P(-0.5 < \int_0^1 [W(1)s - W(s)] dV(s) < 0.5 \mid W(t), \ t \in [0,1]) \right] \]

\[ = E \left[ 2 \Phi \left( \frac{1}{2} \left\{ \int_0^1 [W(1)s - W(s)]^2 ds \right\}^{-1/2} \right) - 1 \right] \]

\[ \approx 0.820 \]

\[ P(\tilde{\theta}_{lm} = 1) \rightarrow P(1/2 < Y < \int_0^1 dS(s) dW(s)) - 1/2 \]

\[ = P(1/2 < Y < 1 - 1/2) \]

\[ = 0. \quad \Rightarrow \text{no pile-up} \]
Remark 2.

\[ P(\hat{\theta}_{lm} = 1) \rightarrow P(0 < Y < \int_{0}^{1} dS(s) dW(s)) \]

\[ = P(0 < Y < c_1), \quad \text{where} \quad c_1 = E |Z_t| / \sigma \]

\[ > 0. \]

On the other hand

\[ P(\tilde{\theta}_{lm} = 1) \rightarrow P(1/2 < Y < c_1 - 1/2) \]

\[ > 0 \quad \text{if and only if} \quad c_1 > 1. \]

That is, there is a pile-up if and only if \( c_1 > 1. \)

Remark 3. Pile-up probability tends to be larger if the density is more concentrated around 0.
Simulation results – pile-up probabilities

Pile-up probabilities for local maximum: $P(\hat{\theta}_{lm} = 1)$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Gau</th>
<th>Lap</th>
<th>Unif</th>
<th>t(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>.827</td>
<td>.796</td>
<td>.831</td>
<td>.796</td>
</tr>
<tr>
<td>50</td>
<td>.859</td>
<td>.806</td>
<td>.864</td>
<td>.823</td>
</tr>
<tr>
<td>100</td>
<td>.873</td>
<td>.819</td>
<td>.864</td>
<td>.817</td>
</tr>
<tr>
<td>200</td>
<td>.844</td>
<td>.819</td>
<td>.843</td>
<td>.831</td>
</tr>
<tr>
<td>500</td>
<td>.855</td>
<td>.809</td>
<td>.841</td>
<td>.846</td>
</tr>
<tr>
<td>$\infty$</td>
<td>.858</td>
<td>.820</td>
<td>.836</td>
<td>.827</td>
</tr>
</tbody>
</table>