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)
MA(1) unit root problem

MA(1): (world’s simplest time series model!)

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

Properties:

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

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

What if \(\theta = 1\)?
Why study MA(1) with a unit root?

a) differencing (to remove non-stationarity)

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

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

b) random walk + noise

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

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

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

Identifiability

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

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

- $\{\epsilon_t\}$ is a special case of an All-Pass Model (Breidt, Davis, Trindade `01, Andrews et al. `05a, `05b)

Gaussian Likelihood

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

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, non-Gaussian data

100 observations from $Y_t = Z_t - \theta_0 Z_{t-1}$, $\{Z_t\} \sim$ IID, Laplace pdf

$\theta_0 = .8 \quad \theta_0 = 1.0 \quad \theta_0 = 1.25$
Gaussian MLE for near-unit roots

Idea: build parameter normalization into the likelihood function.

Model: \[ Y_t = Z_t - \left(1 - \frac{\beta}{n}\right) Z_{t-1}, \ t = 1, \ldots, n. \]

\[ \beta = n(1-\theta), \quad \theta = 1 - \frac{\beta}{n}, \quad \theta_0 = 1 - \frac{\gamma}{n} \]

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 \( \theta_0 = 1 - \frac{\gamma}{n} \),

\[ L_n(\beta) \xrightarrow{d} Z_\gamma(\beta) \quad \text{on} \quad C[0, \infty). \]

Results:

- \( n(1 - \hat{\theta}_{mle}) \rightarrow \hat{\beta}_{mle} = \text{argmax} \ Z_\gamma(\beta) \)
- \( n(1 - \hat{\theta}_L) \rightarrow \hat{\beta}_L = \text{arglocalmax} \ Z_\gamma(\beta) \)
- \( P(\hat{\theta}_L = 1) \rightarrow P(\hat{\beta}_L = 0) = 0.6518 \quad \text{if} \ \gamma = 0. \)
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 \( \hat{\theta}_{mle} < 1 \) (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.,

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

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

The pile-up decreases with increasing tail heaviness.
Laplace likelihood/LAD estimation

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 $1/n$ or $1/n^{1/2}$ asymptotics?

Q2. Is there a pile-up effect?

Look at this problem with non-Gaussian likelihood

- Specifically, consider Laplace likelihood / Least Absolute Deviations for unit root only (not near-unit root)
- Preliminary results only!
Non-Gaussian likelihood – Joint and Exact

Model. \( Y_t = Z_t - \theta Z_{t-1}, \{Z_t\} \sim \text{IID} \) 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

\[
f(y_n, z^{\text{init}}) = f(z_0, z_1, \ldots, z_n)(1_{\{|\theta| \leq 1\}} + |\theta|^{-n} 1_{\{|\theta| > 1\}}),
\]

where the \( z_t \) are solved

forward by: \( z_t = Y_t + \theta z_{t-1}, \quad t = 1, \ldots, n \) for \(|\theta| \leq 1\) with \( z_0 = z^{\text{init}} \)

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

Note: integrate out \( z^{\text{init}} \) to get \textit{Exact} likelihood.

\[
f(y_n) = \int_{-\infty}^{\infty} f(y_n, z^{\text{init}}) dz^{\text{init}}
\]
100 observations from $Y_t = Z_t - \theta_0 Z_{t-1}$, \(\{Z_t\} \sim \text{IID Laplace pdf}\)
Laplace likelihood, Laplace noise

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

$\theta_0 = .8 \quad \theta_0 = 1.0 \quad \theta_0 = 1.25$

Exact likelihood

Joint likelihood at $z_{\text{max}}(\theta)$
(Joint) Laplace log-likelihood. ($\sigma = E|Z_0|$ is a scale parameter)

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

Maximizing wrt $\sigma$, we obtain

$$\hat{\sigma} = \frac{\sum_{t=0}^{n} |z_t|}{(n + 1)}$$

so that maximizing $L$ is equivalent to minimizing

$$l_n(\theta, z^{init}) = \begin{cases} \sum_{t=0}^{n} |z_t|, & \text{if } |\theta| \leq 1, \\ \sum_{t=0}^{n} |z_t| |\theta|, & \text{otherwise.} \end{cases}$$
Joint Laplace likelihood — limit results

Result 1. Under the parameterizations,

\[ \theta = 1 + \frac{\beta}{n} \quad \text{and} \quad z^{\text{init}} = Z_0 + \alpha \sigma / n^{1/2}, \]

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 \( C(\mathbb{R}^2) \), where

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

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

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

for \( \beta > 0 \).
Joint Laplace likelihood — limit results

The limits contain correlated Brownian Motions $S(t)$ and $W(t)$, obtained as the limits of the partial sum processes

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

From the limit,

$$U_n(\beta, \alpha) \to_d U(\beta, \alpha),$$

it suggests (from the continuous mapping theorem?) that

$$\text{limit(optimum(criterion))} = \text{optimum(limit(criterion))}.$$

So for the Local optimizer of the Joint likelihood

$$\left( n(\hat{\theta}_{LJ} - 1), \sqrt{n}\sigma^{-1}(\hat{z}_{LJ}^{\text{init}} - Z_0) \right) \to_d \left( \hat{\beta}_{LJ}, \hat{\alpha}_{LJ} \right)$$

where

$$\left( \hat{\beta}_{LJ}, \hat{\alpha}_{LJ} \right) = \text{arg(local) min } U(\beta, \alpha).$$
Joint Laplace likelihood — limit results

Might expect a similar result to hold for the Global optimizer of the Joint likelihood

\[
\left( n\left( \hat{\theta}_{GJ} - 1 \right), \sqrt{n} \sigma^{-1} \left( \hat{z}_{GJ}^{\text{init}} - Z_0 \right) \right) \rightarrow_d \left( \hat{\beta}_{GJ}, \hat{\alpha}_{GJ} \right)
\]

where

\[
(\hat{\beta}_{GJ}, \hat{\alpha}_{GJ}) = \text{argmin} \ U(\beta, \alpha).
\]
Exact Laplace likelihood — limit results

Exact Laplace Likelihood:

\[ L_n(\theta, \sigma) = \int_{-\infty}^{\infty} f(y_n, z^{\text{init}}) dz^{\text{init}} \]

Result 2. For the Global optimizer of the Exact likelihood,

\[ n(\hat{\theta}_{GE} - 1) \rightarrow_d \hat{\beta}_{GE}, \]

where

\[ \hat{\beta}_{GE} = \arg \min U^*(\beta), \]

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

In addition, for the Local optimizer of the Exact likelihood

\[ n(\hat{\theta}_{LE} - 1) \rightarrow_d \hat{\beta}_{LE}, \hat{\beta}_{LE} = \arg (\text{local}) \min U^*(\beta). \]
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,

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

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

\[
c_1 = \sqrt{\text{Var}(Z_t)} / E^2 |Z_0|^{-1}.
\]

Limit process depends only on \(c_1\) 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 Joint limit \(U(\beta,\alpha)\) and Exact limit \(U^*(\beta)\) numerically.
Simulated realizations of the limit processes

Simulate Joint and Exact limit processes, $U(\beta, \alpha(\beta))$, $U^*(\beta)$.

- Simulate realization of each limit process, joint and exact
- Compute local and global optima
- Repeat…
- Build up limit distribution functions

\[ t(5) \text{ pdf} \]
Limit cdf

red graph = Laplace pdf for $Z_t$

blue graph = Gaussian pdf for $Z_t$

Joint Lap Like

Exact Lap Like
### Local results

<table>
<thead>
<tr>
<th>$n$</th>
<th>bias</th>
<th>s.d.</th>
<th>rmse</th>
<th>asymp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exact LE</td>
<td>Joint LJ</td>
<td>$\hat{\sigma}$</td>
<td></td>
</tr>
<tr>
<td>$n = 20$</td>
<td>-.0057</td>
<td>-.0033</td>
<td>-.0208</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.1438</td>
<td>.0656</td>
<td>.2430</td>
<td></td>
</tr>
<tr>
<td>$n = 50$</td>
<td>.0000</td>
<td>.0004</td>
<td>.0293</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0574</td>
<td>.0208</td>
<td>.1511</td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>.0005</td>
<td>-.0003</td>
<td>-.0025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0303</td>
<td>.0107</td>
<td>.1000</td>
<td></td>
</tr>
<tr>
<td>$n = 200$</td>
<td>.0005</td>
<td>.0000</td>
<td>-.0016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0140</td>
<td>.0058</td>
<td>.0718</td>
<td></td>
</tr>
</tbody>
</table>

### Laplace noise

$\theta = 1$, $\sigma = 1$

1000 reps
Simulation results: Global Exact and Global Joint

Global Exact = MLE
Global Joint = maximize over $\theta$ and $z_{init}$

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

Note:
- Local dominates Global
- Joint dominates Exact (rmse is half the size)

<table>
<thead>
<tr>
<th>$n$</th>
<th>Exact $\hat{\theta}_{GE}$</th>
<th>Joint $\hat{\theta}_{GJ}$</th>
<th>Local $\hat{\theta}_{LJ}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n=20$</td>
<td>bias .047</td>
<td>bias .050</td>
<td>bias -.003</td>
</tr>
<tr>
<td></td>
<td>rmse .224</td>
<td>rmse .213</td>
<td>rmse .144</td>
</tr>
<tr>
<td>$n=50$</td>
<td>bias -.013</td>
<td>bias .002</td>
<td>bias .000</td>
</tr>
<tr>
<td></td>
<td>rmse .096</td>
<td>rmse .078</td>
<td>rmse .057</td>
</tr>
<tr>
<td>$n=100$</td>
<td>bias .003</td>
<td>bias -.003</td>
<td>bias .000</td>
</tr>
<tr>
<td></td>
<td>rmse .051</td>
<td>rmse .034</td>
<td>rmse .011</td>
</tr>
<tr>
<td>$n=200$</td>
<td>bias .000</td>
<td>bias .000</td>
<td>bias .000</td>
</tr>
<tr>
<td></td>
<td>rmse .028</td>
<td>rmse .014</td>
<td>rmse .006</td>
</tr>
</tbody>
</table>
Analysis of pile-up probabilities

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

- When is there a local optimum at $\theta = 1$?
- Check derivatives
- Negative derivative from the left
- Positive derivative from the right
- Local optimum at $\theta = 1$
Pile-up probabilities (Joint)

Result 3. (Local Joint Laplace likelihood)

\[ P(\hat{\theta}_{LJ} = 1) \rightarrow P(0 < Y < 1), \]

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: look at derivatives

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

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

Now,

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

\[ \lim_{\beta \uparrow 0} \frac{\partial}{\partial \beta} U(\beta, \hat{\alpha}(\beta)) = Y - 1 \]

and the result follows.
Result 4. (Local Exact Laplace likelihood)

\[ P(\hat{\theta}_{LE} = 1) \rightarrow P\left[ \frac{1}{2} < Y < 1 - \frac{1}{2} \right] = 0 \]

The pile-up probability is always zero for the Local Exact, and always positive for the Local Joint (see Result 3).

Remark. (Laplace pile-up)

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 Local Joint has pile-up probability

\[ P(\hat{\theta}_{LJ} = 1) \rightarrow P(0 < Y < 1) \]

\[ = P(0 < \int_0^1 [W(1)s - W(s)] \, dV(s) + .5 < 1) \]

\[ = E \left[ P(-.5 < \int_0^1 [W(1)s - W(s)] \, dV(s) < .5 \mid W(t), \, t \in [0,1]) \right] \]

\[ = E \left[ 2\Phi \left( .5 \left\{ \int_0^1 [W(1)s - W(s)]^2 \, ds \right\}^{-1/2} \right) - 1 \right] \]

\[ \approx 0.820 \]

But no pile-up probability for Local Exact:

Remark: if Local does not pile up, Global does not pile up

if Local does pile up, Global probably does as well
Simulation results – pile-up probabilities

Pile-up probabilities for \textbf{Local Joint}: \( P(\hat{\theta}_{LJ} = 1) \)

\[
\begin{array}{|c|c|c|c|c|}
\hline
n & \text{Gau} & \text{Lap} & \text{Unif} & t(5) \\
\hline
20 & .827 & .796 & .831 & .796 \\
50 & .859 & .806 & .864 & .823 \\
100 & .873 & .819 & .864 & .817 \\
200 & .844 & .819 & .843 & .831 \\
500 & .855 & .809 & .841 & .846 \\
\infty & .858 & .820 & .836 & .827 \\
\hline
\end{array}
\]

(No pile-up probabilities for \textbf{Local Exact.})