Another Look at Estimation for MA(1) Processes With a Unit Root

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

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

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

Properties:

• \( |\theta| < 1 \Rightarrow Z_t = \sum_{j=0}^{\infty} \theta^j Y_{t-j} \) (invertible)

• \( |\theta| > 1 \Rightarrow Z_t = -\sum_{j=1}^{\infty} \theta^j Y_{t+j} \) (non-invertible)

• \( |\theta| = 1 \Rightarrow Z_t \in \text{sp}\{Y_t, Y_{t-1},\ldots\} \quad \text{and} \quad Z_t \in \text{sp}\{Y_{t+1}, Y_{t+2},\ldots,\} \)

\[ \Rightarrow \ P_{\text{sp}\{Y_t, s\neq 0\}} Y_0 = Y_0 \quad \text{(perfect interpolation)} \]

• \( |\theta| < 1 \Rightarrow \hat{\theta}_{MLE} \text{ is } \text{AN}(0, (1 - \theta^2) / n) \)

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

• 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

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

\[ Y_t = X_t + V_t \text{ (random walk signal + noise)} \]

Then

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

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

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

- \(|\theta| > 1 \implies Y_t = \varepsilon_t - \theta^{-1} \varepsilon_{t-1}, \) where \( \{\varepsilon_t\} \sim \text{WN}(0, \theta^2\sigma^2). \)

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

- \( \{\varepsilon_t\} \) 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

$\theta_0 = 0.8 \quad \theta_0 = 1.0 \quad \theta_0 = 1.25$
Gaussian MLE for near-unit roots

Idea: build parameter normalization into the likelihood function.

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

\[ \beta = n(1-\theta), \quad \theta = 1 - \beta/n, \quad \theta_0 = 1 - \gamma/n \]

Gaussian Likelihood:

\[ L_n(\beta) = l_n(1 - \beta/n) - l_n(1), \quad l_n(\cdot) = \text{profile log-like}. \]

Theorem (Davis and Dunsmuir `96): Under \( \theta_0 = 1 - \gamma/n \),

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

Results:

- \( n(1 - \hat{\theta}_{\text{mle}}) \to \hat{\beta}_{\text{mle}} = \arg\max Z_\gamma(\beta) \)
- \( n(1 - \hat{\theta}_L) \to \hat{\beta}_L = \arg\max \text{local } Z_\gamma(\beta) \)
- \( P(\hat{\theta}_L = 1) \to P(\hat{\beta}_L = 0) = .6518 \quad \text{if } \gamma = 0. \)
Extensions of MLE (Gaussian likelihood)

i) non-zero mean (Chen and Davis `00): same type of limit, except pile-up is more excessive.
   \[ P(\hat{\theta}_{\text{mle}} = 1) \to .955 \]

This makes hypothesis testing easy!

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

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)

• Some results are preliminary 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.

\[
Z^{\text{init}} = \begin{cases} 
Z_0, & \text{if } |\theta| \leq 1, \\
Z_n - \sum^n_{t=1} Y_t, & \text{otherwise.}
\end{cases}
\]

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: \( z_t = Y_t + \theta z_{t-1}, \ t = 1, \ldots, n \) 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 \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}

$\theta_0 = 1.0$

$\theta_0 = 0.8$
Laplace likelihood, Laplace noise

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

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

Exact likelihood

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

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

Maximizing wrt $\sigma$, we obtain

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

so that maximizing $L$ is equivalent to minimizing

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

Result 1. Under the parameterizations,
\[ \theta = 1 + \frac{\beta}{n} \quad \text{and} \quad z_{\text{init}} = Z_0 + \frac{\alpha \sigma}{n^{1/2}}, \]
we have
\[ U_n(\beta, \alpha) = \sigma^{-1}(l_n(\theta, z_{\text{init}}) - l_n(1, Z_0)) \to_d U(\beta, \alpha) \]
where
\[ U(\beta, \alpha) = \int_0^1 \left( \beta \int_0^s e^{\beta(s-t)} dS(t) + \alpha e^{\beta s} \right) dW(s) \]
\[ + f(0) \int_0^1 \left( \beta \int_0^s e^{\beta(s-t)} dS(t) + \alpha e^{\beta s} \right)^2 ds \]
for \( \beta \leq 0 \), and
\[ U(\beta, \alpha) = \int_0^1 \left( -\beta \int_{t}^{1} e^{\beta(s-t)} dS(t) + \alpha e^{-\beta(1-s)} \right) dW(s) \]
\[ + f(0) \int_0^1 \left( -\beta \int_{s}^{1} e^{\beta(s-t)} dS(t) + \alpha e^{-\beta(1-s)} \right)^2 ds \]
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

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

So for the optimizer of the Joint likelihood

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

where

$$(\hat{\beta}_j, \hat{\alpha}_j) = \arg(\text{local}) \min U(\beta, \alpha).$$
Consistent estimation of noise?

Note that the previous results imply that

\[ \hat{\theta}^{\text{init}} = Z_0 + \frac{\sigma}{\sqrt{n}} \hat{\alpha} = Z_0 + O_p \left( n^{-1/2} \right) \]

so that an unobserved random noise can be consistently estimated.

Does this make any sense?

Recall that in the unit root case,

\[ Z_0 \in \overline{\text{sp}\{Y_1, Y_2, \ldots, Y_n, \ldots\}} \]

so that in fact, consistent estimation is possible.
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 local optimizer of the Exact likelihood,

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

where

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

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

\[
W(t) = \sum_{j=1}^{[100000 \cdot t]} W_j / \sqrt{100000} \quad \text{and} \quad V(t) = \sum_{j=1}^{[100000 \cdot t]} V_j / \sqrt{100000}.
\]

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 \text{Local} and \text{Global} minimizers of \(U(\beta, \alpha)\) and \(U^*(\beta)\) 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
red graph = Laplace pdf for $Z_t$

blue graph = Gaussian pdf for $Z_t$

Joint Lap Like

Exact Lap Like
### Simulation results: Global Exact and Global Joint

**Exact** = MLE

**Joint** = maximize over $\theta$ and $z_{init}$

Laplace noise

$\theta = 1, \sigma = 1$

1000 reps

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

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\hat{\theta}_E$</th>
<th>$\hat{\theta}_J$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 20$</td>
<td>bias</td>
<td>-.047</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>.224</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>bias</td>
<td>-.013</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>.096</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>bias</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>.051</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>bias</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>.028</td>
</tr>
</tbody>
</table>
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. (Joint Laplace likelihood)

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

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

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

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

Now,

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

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

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

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

The pile-up probability is always zero for the Exact, and always positive for the 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 the Joint estimator has pile-up probability

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

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

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

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

\[ \approx 0.820 \]

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 Joint: $P(\hat{\theta}_j = 1)$

<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>

(No pile-up probabilities for Exact.)
Summary and Future Work

• Reviewed MA(1) unit root and near-unit root with Gaussian likelihood
  • $1/n$ asymptotics, pile-up even if $\theta < 1$

• New results for MA(1) unit root with Least Absolute Deviations
  • $1/n$ asymptotics for \textit{Joint} or \textit{Exact}
  • \textit{Joint} beats \textit{Exact};
  • \textit{Joint} has pile-up and \textit{Exact} does not

• Further work:
  • Nail down preliminary results, conduct further simulations
  • Other non-Gaussian criterion functions (MLE)?
  • Non-zero mean?
  • Near-unit root? $(1-\gamma/n)$
  • Performance of \textit{Joint} with Gaussian likelihood?