R-Estimation for Allpass Time Series Models

Richard A. Davis

Department of Statistics
Colorado State University

http://www.stat.colostate.edu/~rdavis/lectures/

Joint work with
Beth Andrews, Colorado State University
Jay Breidt, Colorado State University
Introduction
• properties of financial time series
• motivating example
• all-pass models and their properties

Estimation
• likelihood approximation
• MLE, R-estimation, and LAD
• asymptotic results
• order selection

Empirical results
• simulation

Noninvertible MA processes
• preliminaries
• a two-step estimation procedure
• Microsoft trading volume

Summary
Financial Time Series

Log returns, \( X_t = 100*(\ln (P_t) - \ln (P_{t-1})) \), of financial assets often exhibit:

- heavy-tailed marginal distributions
  \[ P(|X_1| > x) \sim C x^{-\alpha}, \quad 0 < \alpha < 4. \]

- lack of serial correlation
  \( \hat{\rho}_X(h) \) near 0 for all lags \( h > 0 \) (MGD sequence)

- \( |X_t| \) and \( X_t^2 \) have slowly decaying autocorrelations
  \( \hat{\rho}_{|X|}(h) \) and \( \hat{\rho}_{X^2}(h) \) converge to 0 slowly as \( h \to \infty \)

- process exhibits ‘stochastic volatility’

Nonlinear models \( X_t = \sigma_t Z_t, \{Z_t\} \sim \text{IID}(0,1) \)

- ARCH and its variants (Engle `82; Bollerslev, Chou, and Kroner 1992)

- Stochastic volatility (Clark 1973; Taylor 1986)
All-pass model of order 2 (t3 noise)

ACF : (allpass)2

model
sample

ACF : (allpass)
All-pass Models

Causal AR polynomial: \( \phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \), \( \phi(z) \neq 0 \) for \( |z| \leq 1 \).

Define MA polynomial:

\[
\theta(z) = -z^p \phi(z^{-1})/\phi_p = -(z^p - \phi_1 z^{p-1} - \cdots - \phi_p)/\phi_p
\]

\( \neq 0 \) for \( |z| \geq 1 \) (MA polynomial is non-invertible).

Model for data \( \{X_t\} : \phi(B)X_t = \theta(B)Z_t \), \( \{Z_t\} \sim \text{IID (non-Gaussian)} \)

\( B^kX_t = X_{t-k} \)

Examples:

All-pass(1): \( X_t - \phi X_{t-1} = Z_t - \phi^{-1} Z_{t-1} \), \( |\phi| < 1 \).

All-pass(2): \( X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t + \phi_1/\phi_2 Z_{t-1} - 1/\phi_2 Z_{t-2} \)
Properties:

- causal, non-invertible ARMA with MA representation

\[ X_t = \frac{B^p \phi(B^{-1})}{\phi_p \phi(B)} Z_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \]

- uncorrelated (flat spectrum)

\[ f_X(\omega) = \frac{\left| e^{-ip\omega}\right|^2 \left| \phi(e^{i\omega}) \right|^2}{\phi_p^2 \left| \phi(e^{-i\omega}) \right|^2} \frac{\sigma^2}{2\pi} \]

\[ = \frac{\sigma^2}{\phi_p^2 2\pi} \]

- zero mean
- data are dependent if noise is non-Gaussian (e.g. Breidt & Davis 1991).
- squares and absolute values are correlated.
- \( X_t \) is heavy-tailed if noise is heavy-tailed.
Estimation for All-Pass Models

- Second-order moment techniques do not work
  - least squares
  - Gaussian likelihood

- Higher-order cumulant methods
  - Giannakis and Swami (1990)
  - Chi and Kung (1995)

- Non-Gaussian likelihood methods
  - likelihood approximation assuming known density
  - quasi-likelihood

- Other
  - LAD- least absolute deviation
  - R-estimation (minimum dispersion)
Approximating the likelihood

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

Model: \(X_t = \phi_{01}X_{t-1} + \cdots + \phi_{0p}X_{t-p} - (Z_{t-p} - \phi_{01}Z_{t-p+1} - \cdots - \phi_{0p}Z_{t}) / \phi_{0r}\)

where \(\phi_{0r}\) is the last non-zero coefficient among the \(\phi_{0j}\)’s.

Noise: \(z_{t-p} = \phi_{01}z_{t-p+1} + \cdots + \phi_{0p}z_t - (X_t - \phi_{01}X_{t-1} - \cdots - \phi_{0p}X_{t-p})\),

where \(z_t = Z_t / \phi_{0r}\).

More generally define,

\[ z_{t-p}(\phi) = \begin{cases} 
0, & \text{if } t = n + p, \ldots, n + 1, \\
\phi_1 z_{t-p+1}(\phi) + \cdots + \phi_p z_t(\phi) - \phi(B) X_t, & \text{if } t = n, \ldots, p + 1. 
\end{cases} \]

Note: \(z_t(\phi_0)\) is a close approximation to \(z_t\) (initialization error)
Log-likelihood:

$$L(\phi, \sigma) = -(n - p) \ln(\sigma / | \phi_q |) + \sum_{t=1}^{n-p} \ln f(\sigma^{-1} \phi_q z_t(\phi))$$

where $f_\sigma(z) = \sigma^{-1} f(z/\sigma)$.

Least absolute deviations: choose Laplace density

$$f(z) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2} \frac{|z|}{\kappa}\right)$$

and log-likelihood becomes

constant $- (n - p) \ln \kappa - \sum_{t=1}^{n-p} \sqrt{2} | z_t(\phi) | / \kappa$, $\kappa = \sigma / \phi_q$

Concentrated Laplacian likelihood

$$l(\phi) = \text{constant} - (n - p) \ln \sum_{t=1}^{n-p} | z_t(\phi) |$$

Maximizing $l(\phi)$ is equivalent to minimizing the absolute deviations

$$m_n(\phi) = \sum_{t=1}^{n-p} | z_t(\phi) |.$$
Assumptions for MLE

Assume \( \{Z_t\} \) iid \( f_\sigma(z) = \sigma^{-1} f(\sigma^{-1} z) \) with

- \( \sigma \) a scale parameter
- mean 0, variance \( \sigma^2 \)
- further smoothness assumptions (integrability, symmetry, etc.) on \( f \)
- Fisher information:

\[
\tilde{I} = \sigma^{-2} \int (f'(z))^2 / f(z) \, dz
\]
Assumptions for MLE

Assume \( \{Z_t\} \) iid \( f_\sigma(z) = \sigma^{-1} f(\sigma^{-1} z) \) with

- \( \sigma \) a scale parameter
- mean 0, variance \( \sigma^2 \)
- further smoothness assumptions (integrability, symmetry, etc.) on \( f \)

Fisher information:

\[
\tilde{I} = \sigma^{-2} \int (f''(z))^2 / f(z) dz
\]

Results

Let \( \gamma(h) = \text{ACVF of AR model with AR poly } \phi_0(.) \) and

\[
\Gamma_p = [\gamma(j-k)]_{j,k=1}^p
\]

\[
\sqrt{n}(\hat{\phi}_{\text{MLE}} - \phi_0) \xrightarrow{D} N(0, \frac{1}{2(\sigma^2 \tilde{I} - 1)} \sigma^2 \Gamma_p^{-1})
\]
Further comments on MLE

Let \( \alpha=(\phi_1, \ldots, \phi_p, \sigma/|\phi_p|, \beta_1, \ldots, \beta_q) \), where \( \beta_1, \ldots, \beta_q \) are the parameters of pdf \( f \).

Set

\[
\hat{I} = \sigma_0^{-2} \int (f'(z;\beta_0))^2 / f(z;\beta_0) \, dz
\]

\[
\hat{K} = \alpha_{0,p+1}^{-2} \left\{ \int z^2 (f'(z;\beta_0))^2 / f(z;\beta_0) \, dz - 1 \right\}
\]

\[
L = -\alpha_{0,p+1}^{-1} \int z \frac{f'(z;\beta_0)}{f(z;\beta_0)} \frac{\partial f(z;\beta_0)}{\partial \beta_0} \, dz
\]

\[
I_f(\beta_0) = \int \frac{1}{f(z;\beta_0)} \frac{\partial f(z;\beta_0)}{\partial \beta_0} \frac{\partial f^T(z;\beta_0)}{\partial \beta_0} \, dz \quad \text{(Fisher Information)}
\]
Under smoothness conditions on $f$ wrt $\beta_1, \ldots, \beta_q$ we have

$$\sqrt{n}(\hat{\alpha}_{\text{MLE}} - \alpha_0) \xrightarrow{D} N(0, \Sigma^{-1}),$$

where

$$\Sigma^{-1} = \begin{bmatrix}
\frac{1}{2(\sigma_0^2 I - 1)} & 0 & 0 \\
0 & (\hat{K} - L' I_f^{-1} L)^{-1} & -\hat{K}^{-1} L' (I_f - L\hat{K}^{-1} L')^{-1} \\
0 & -(I_f - L\hat{K}^{-1} L')^{-1} L\hat{K}^{-1} & (I_f - L\hat{K}^{-1} L')^{-1}
\end{bmatrix}$$

Note: $\hat{\phi}_{\text{MLE}}$ is asymptotically independent of $\hat{\alpha}_{p+1, \text{MLE}}$ and $\hat{\beta}_{\text{MLE}}$. 
Asymptotic Covariance Matrix

- For LS estimators of AR(p):
  \[ \sqrt{n} (\hat{\phi}_{LS} - \phi_0) \overset{d}{\to} N(0, \sigma^2 \Gamma_p^{-1}) \]

- For LAD estimators of AR(p):
  \[ \sqrt{n} (\hat{\phi}_{LAD} - \phi_0) \overset{d}{\to} N(0, \frac{1}{4\sigma^2 f^2(0)} \sigma^2 \Gamma_p^{-1}) \]

- For LAD estimators of AP(p):
  \[ \sqrt{n} (\hat{\phi}_{LAD} - \phi_0) \overset{d}{\to} N(0, \frac{\text{Var}(|Z_1|)}{2(2\sigma^2 f_\sigma(0) - E|Z_1|^2)} \sigma^2 \Gamma_p^{-1}) \]

- For MLE estimators of AP(p):
  \[ \sqrt{n} (\hat{\phi}_{MLE} - \phi_0) \overset{d}{\to} N(0, \frac{1}{2(\sigma^2 \hat{I} - 1)} \sigma^2 \Gamma_p^{-1}) \]
Laplace: \((\text{LAD} = \text{MLE})\)

\[
\frac{\text{Var}(|Z_1|)}{2(2\sigma^2 f_\sigma(0) - E|Z_1|)^2} = \frac{1}{2} = \frac{1}{2(\sigma^2 \hat{\theta} - 1)}
\]

Students \(t_\nu, \nu > 2:\)

\[
\text{LAD: } \frac{(\nu - 2)}{8 \Gamma^2 (\nu + 1)/2} \left( \pi (\nu - 1)^2 \Gamma^2 (\nu/2) - 4(\nu - 2)\Gamma^2 ((\nu + 1)/2) \right)
\]

\[
\text{MLE: } \frac{1}{2(\sigma^2 \hat{\theta} - 1)} = \frac{(\nu - 2)(\nu + 3)}{12}
\]

Student’s \(t_3:\)

\[
\text{LAD: } 0.7337
\]

\[
\text{MLE: } 0.5
\]

\[
\text{ARE: } 0.7337/0.5 = 1.4674
\]
R-Estimation:

Minimize the objective function

\[
S(\phi) = \sum_{t=1}^{n-p} \phi\left(\frac{t}{n - p + 1}\right)z_{(t)}(\phi)
\]

where \(\{z_{(t)}(\phi)\}\) are the ordered \(\{z_t(\phi)\}\), and the weight function \(\phi\) satisfies:

- \(\phi\) is differentiable and nondecreasing on \((0,1)\)
- \(\phi'\) is uniformly continuous
- \(\phi(x) = -\phi(1-x)\)
R-Estimation:

Minimize the objective function

$$S(\phi) = \sum_{t=1}^{n-p} \phi\left(\frac{t}{n-p+1}\right)z_{(t)}(\phi)$$

where \(\{z_{(t)}(\phi)\}\) are the ordered \(\{z_t(\phi)\}\), and the weight function \(\phi\) satisfies:

- \(\phi\) is differentiable and nondecreasing on \((0,1)\)
- \(\phi'\) is uniformly continuous
- \(\phi(x) = -\phi(1-x)\)

Remarks:

- \(S(\phi) = \sum_{t=1}^{n-p} \phi\left(\frac{R_t(\phi)}{n-p+1}\right)z_t(\phi)\)
- For LAD, take \(\phi(x) = \begin{cases} -1, & 0 < x < 1/2, \\ 1, & 1/2 < x < 1. \end{cases}\)
Assumptions for R-estimation

Assume \( \{Z_t\} \) iid with density function \( f(\text{distr } F) \)
- mean 0, variance \( \sigma^2 \)

Assume weight function \( \varphi \) is nondecreasing and continuously differentiable with \( \varphi(x) = -\varphi(1-x) \)
Assumptions for R-estimation

☞ Assume \( \{Z_t\} \) iid with density function \( f(\text{distr } F) \)
  
  • mean 0, variance \( \sigma^2 \)

☞ Assume weight function \( \phi \) is nondecreasing and continuously differentiable with \( \phi(x) = -\phi(1-x) \)

Results

☞ Set

\[
\tilde{J} = \int_0^1 \phi^2(s)ds, \quad \tilde{K} = \int_0^1 F^{-1}(s)\phi(s)ds, \quad \tilde{L} = \int_0^1 f(F^{-1}(s))\phi'(s)ds
\]

☞ If \( \sigma^2\tilde{L} > \tilde{K} \), then

\[
\sqrt{n}(\hat{\phi}_R - \phi_0) \xrightarrow{D} N(0, \frac{\sigma^2\tilde{J} - \tilde{K}^2}{2(\sigma^2\tilde{L} - \tilde{K})^2} \sigma^2\Gamma_p^{-1})
\]
Further comments on R-estimation

\( \varphi(x) = x - 1/2 \) is called the Wilcoxon weight function

By formally choosing \( \varphi(x) = \begin{cases} -1, & 0 < x < 1/2, \\ 1, & 1/2 < x < 1. \end{cases} \) we obtain

\[
\frac{\sigma^2 \tilde{J} - \tilde{K}^2}{2(\sigma^2 \tilde{L} - \tilde{K})^2} \sigma^2 \Gamma_p^{-1} = \frac{\text{Var}(\mid Z_1 \mid)}{2(2\sigma^2 f_\sigma(0) - E \mid Z_1 \mid)^2} \sigma^2 \Gamma_p^{-1}.
\]

That is \( R = \text{LAD}, \) asymptotically.

The \( R \)-estimation objective function is smoother than the \( \text{LAD} \)-objective function and hence easier to minimize.
Objective Functions

R-estimation
LAD

phi

22.0 22.5 23.0 23.5 24.0 24.5
Summary of asymptotics

☞ Maximum likelihood:

\[ \sqrt{n} (\hat{\phi}_{\text{MLE}} - \phi_0) \xrightarrow{D} N(0, \frac{1}{2(\sigma^2 \widetilde{I} - 1)} \sigma^2 \Gamma_p^{-1}) \]

☞ R-estimation

\[ \sqrt{n} (\hat{\phi}_R - \phi_0) \xrightarrow{D} N(0, \frac{\sigma^2 \widetilde{J} - \widetilde{K}^2}{2(\sigma^2 \widetilde{L} - \widetilde{K})^2} \sigma^2 \Gamma_p^{-1}) \]

☞ Least absolute deviations:

\[ \sqrt{n} (\hat{\phi}_{\text{LAD}} - \phi_0) \xrightarrow{D} N(0, \frac{\text{Var}(|Z_1|)}{2(2\sigma^2 f_\sigma(0) - E |Z_1|)^2} \sigma^2 \Gamma_p^{-1}) \]
Laplace:  \( \text{LAD}=\text{MLE} \)

\[
R: \quad \frac{\sigma^2 \tilde{J} - \tilde{K}^2}{2(\sigma^2 \tilde{L} - \tilde{K})^2} = \frac{5}{6} \quad \text{(using } \varphi(x) = x-1/2, \text{ Wilcoxon)}
\]

\[\text{LAD}=\text{MLE}: \quad 1/2\]

**Students \( t_\nu \):**

<table>
<thead>
<tr>
<th>( \nu )</th>
<th>LAD</th>
<th>R</th>
<th>MLE</th>
<th>LAD/R</th>
<th>MLE/R</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.733</td>
<td>0.520</td>
<td>0.500</td>
<td>1.411</td>
<td>0.962</td>
</tr>
<tr>
<td>6</td>
<td>6.22</td>
<td>3.01</td>
<td>3.00</td>
<td>2.068</td>
<td>0.997</td>
</tr>
<tr>
<td>9</td>
<td>16.8</td>
<td>7.15</td>
<td>7.00</td>
<td>2.354</td>
<td>0.980</td>
</tr>
<tr>
<td>12</td>
<td>32.6</td>
<td>13.0</td>
<td>12.5</td>
<td>2.510</td>
<td>0.964</td>
</tr>
<tr>
<td>15</td>
<td>53.4</td>
<td>20.5</td>
<td>19.5</td>
<td>2.607</td>
<td>0.952</td>
</tr>
<tr>
<td>20</td>
<td>99.6</td>
<td>36.8</td>
<td>34.5</td>
<td>2.707</td>
<td>0.937</td>
</tr>
<tr>
<td>30</td>
<td>234</td>
<td>83.6</td>
<td>77.0</td>
<td>2.810</td>
<td>0.921</td>
</tr>
</tbody>
</table>
Central Limit Theorem (R-estimation)

• Think of \( u = n^{1/2}(\phi - \phi_0) \) as an element of \( \mathbb{R}^p \)

• Define

\[
S_n(u) = \sum_{t=1}^{n-p} \left( \varphi\left( \frac{R_t(\phi_0 + n^{-1/2}u)}{n - p + 1} \right) z_t(\phi_0 + n^{-1/2}u) \right) - \sum_{t=1}^{n-p} \left( \varphi\left( \frac{R_t(\phi_0 + n^{-1/2}u)}{n - p + 1} \right) z_t(\phi_0) \right),
\]

where \( R_t(\phi) \) is the rank of \( z_t(\phi) \) among \( z_1(\phi), \ldots, z_{n-p}(\phi) \).

• Then \( S_n(u) \to S(u) \) in distribution on \( C(\mathbb{R}^p) \), where

\[
S(u) = |\phi_{0r}|^{-1} (\sigma^2 \tilde{L} - \tilde{K}) u' \sigma^{-2} \Gamma_p u + u'N, \quad N \sim N(0, 2(\sigma^2 \tilde{J} - \tilde{K}^2) |\phi_{0r}|^{-2} \sigma^{-2} \Gamma_p),
\]

• Hence,

\[
\text{arg min } S_n(u) = n^{1/2} \left( \hat{\phi}_R - \phi_0 \right)
\]

\[
\to \text{arg min } S(u)
\]

\[
= - \frac{|\phi_{0r}|}{2(\sigma^2 \tilde{L} - \tilde{K})} \sigma^2 \Gamma_p^{-1} N \sim N(0, \frac{\sigma^2 \tilde{J} - \tilde{K}^2}{2(\sigma^2 \tilde{L} - \tilde{K})^2} |\phi_{0r}|^2 \sigma^2 \Gamma_p^{-1})
\]
Order Selection:

**Partial ACF** From the previous result, if true model is of order \( r \) and fitted model is of order \( p > r \), then

\[
n^{1/2} \hat{\phi}_{p,R} \rightarrow N(0, \frac{\sigma^2 \bar{J} - \bar{K}^2}{2(\sigma^2 \bar{L} - \bar{K})^2})
\]

where \( \hat{\phi}_{p,R} \) is the pth element of \( \hat{\phi}_R \).

**Procedure:**

1. Fit high order (P-th order), obtain residuals and estimate the scalar,

\[
\theta^2 = \frac{\sigma^2 \bar{J} - \bar{K}^2}{2(\sigma^2 \bar{L} - \bar{K})^2}
\]

by empirical moments of residuals and density estimates.
2. Fit AP models of order $p=1,2,\ldots,P$ via LAD and obtain $p$-th coefficient $\hat{\phi}_{p,p}$ for each.

3. Choose model order $r$ as the smallest order beyond which the estimated coefficients are statistically insignificant.
Sample realization of all-pass of order 2

(a) Data From Allpass Model

(b) ACF of Allpass Data

(c) ACF of Squares

(d) ACF of Absolute Values
Simulation results:

• 1000 replicates of all-pass models

• model order parameter value
  1 $\phi_1 = .5$
  2 $\phi_1 = .3, \phi_2 = .4$

• noise distribution is t with 3 d.f.

• sample sizes $n=500, 5000$

• estimation method is LAD
To guard against being trapped in local minima, we adopted the following strategy.

• 250 random starting values were chosen at random. For model of order $p$, $k$-th starting value was computed recursively as follows:

1. Draw $\phi_{11}^{(k)}, \phi_{22}^{(k)}, \ldots, \phi_{pp}^{(k)}$ iid uniform $(-1,1)$.
2. For $j=2, \ldots, p$, compute

$$
\begin{bmatrix}
\phi_{j1}^{(k)} \\
\vdots \\
\phi_{jj}^{(k)}
\end{bmatrix} =
\begin{bmatrix}
\phi_{j-1,1}^{(k)} \\
\vdots \\
\phi_{j-1,j-1}^{(k)}
\end{bmatrix} - \phi_{jj}^{(k)}
$$

• Select top 10 based on minimum function evaluation.

• Run Hooke and Jeeves with each of the 10 starting values and choose best optimized value.
<table>
<thead>
<tr>
<th>N</th>
<th>Asymptotic mean</th>
<th>std dev</th>
<th>Empirical mean</th>
<th>std dev</th>
<th>%coverage</th>
<th>rel eff*</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>$\phi_1=0.5$</td>
<td>0.0332</td>
<td>0.4979</td>
<td>0.0397</td>
<td>94.2</td>
<td>11.8</td>
</tr>
<tr>
<td>5000</td>
<td>$\phi_1=0.5$</td>
<td>0.0105</td>
<td>0.4998</td>
<td>0.0109</td>
<td>95.4</td>
<td>9.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>Asymptotic mean</th>
<th>std dev</th>
<th>Empirical mean</th>
<th>std dev</th>
<th>%coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>$\phi_1=0.3$</td>
<td>0.0351</td>
<td>0.2990</td>
<td>0.0456</td>
<td>92.5</td>
</tr>
<tr>
<td></td>
<td>$\phi_2=0.4$</td>
<td>0.0351</td>
<td>0.3965</td>
<td>0.0447</td>
<td>92.1</td>
</tr>
<tr>
<td>5000</td>
<td>$\phi_1=0.3$</td>
<td>0.0111</td>
<td>0.3003</td>
<td>0.0118</td>
<td>95.5</td>
</tr>
<tr>
<td></td>
<td>$\phi_2=0.4$</td>
<td>0.0111</td>
<td>0.3990</td>
<td>0.0117</td>
<td>94.7</td>
</tr>
</tbody>
</table>

*Efficiency relative to maximum absolute residual kurtosis:

$$
\left| \frac{1}{n-p} \sum_{t=1}^{n-p} \left( \frac{z_t(\hat{\phi})}{v_2^{1/2}} \right)^4 - 3 \right|, \quad v_2 = \frac{1}{n-p} \sum_{t=1}^{n-p} (z_t(\hat{\phi}) - \bar{z}(\hat{\phi}))^2
$$

34
### MLE Simulations Results using t-distr(3.0)

<table>
<thead>
<tr>
<th>N</th>
<th>Asymptotic</th>
<th></th>
<th>Empirical</th>
<th></th>
<th>%coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>std dev</td>
<td>mean</td>
<td>std dev</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>φ₁=.5</td>
<td>.0274</td>
<td>.4971</td>
<td>.0315</td>
<td>93.0</td>
</tr>
<tr>
<td></td>
<td>φ₂=.4</td>
<td>.4480</td>
<td>3.112</td>
<td>.5008</td>
<td>95.8</td>
</tr>
<tr>
<td>5000</td>
<td>φ₁=.5</td>
<td>.0087</td>
<td>.4997</td>
<td>.0091</td>
<td>93.4</td>
</tr>
<tr>
<td></td>
<td>φ₂=.4</td>
<td>.1417</td>
<td>3.008</td>
<td>.1533</td>
<td>94.0</td>
</tr>
<tr>
<td>500</td>
<td>φ₁=.3</td>
<td>.0290</td>
<td>.2993</td>
<td>.0345</td>
<td>90.6</td>
</tr>
<tr>
<td></td>
<td>φ₂=.4</td>
<td>.0290</td>
<td>.3964</td>
<td>.0350</td>
<td>90.1</td>
</tr>
<tr>
<td></td>
<td>v=3.0</td>
<td>.4480</td>
<td>3.079</td>
<td>.4722</td>
<td>94.8</td>
</tr>
<tr>
<td>5000</td>
<td>φ₁=.3</td>
<td>.0092</td>
<td>.2999</td>
<td>.0095</td>
<td>94.0</td>
</tr>
<tr>
<td></td>
<td>φ₂=.4</td>
<td>.0092</td>
<td>.3999</td>
<td>.0094</td>
<td>94.6</td>
</tr>
<tr>
<td></td>
<td>v=3.0</td>
<td>.1417</td>
<td>3.008</td>
<td>.1458</td>
<td>95.2</td>
</tr>
</tbody>
</table>
**R-Estimation:** Minimize the objective fcn

\[ S(\phi) = \sum_{t=1}^{n-p} \left( \frac{t}{n-p+1} - \frac{1}{2} \right) z_{(t)}(\phi) \]

where \( \{z_{(t)}(\phi)\} \) are the ordered \( \{z_t(\phi)\} \).

<table>
<thead>
<tr>
<th>N</th>
<th>( \phi_1 = .5 )</th>
<th>( \phi_2 = .4 )</th>
<th>( \phi_1 = .3 )</th>
<th>( \phi_2 = .4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.4978</td>
<td>0.3957</td>
<td>0.2988</td>
<td>0.3957</td>
</tr>
<tr>
<td></td>
<td>0.0315</td>
<td>0.0360</td>
<td>0.0374</td>
<td>0.0360</td>
</tr>
<tr>
<td>5000</td>
<td>0.4997</td>
<td>0.3997</td>
<td>0.2988</td>
<td>0.3997</td>
</tr>
<tr>
<td></td>
<td>0.0094</td>
<td>0.0360</td>
<td>0.0374</td>
<td>0.0360</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>( \phi_1 = .5 )</th>
<th>( \phi_2 = .4 )</th>
<th>( \phi_1 = .3 )</th>
<th>( \phi_2 = .4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.4979</td>
<td>0.3965</td>
<td>0.2990</td>
<td>0.3965</td>
</tr>
<tr>
<td></td>
<td>0.0397</td>
<td>0.0447</td>
<td>0.0456</td>
<td>0.0447</td>
</tr>
<tr>
<td>5000</td>
<td>0.4998</td>
<td>0.3990</td>
<td>0.2990</td>
<td>0.3990</td>
</tr>
<tr>
<td></td>
<td>0.0109</td>
<td>0.0117</td>
<td>0.0118</td>
<td>0.0117</td>
</tr>
</tbody>
</table>
Noninvertible MA models with heavy tailed noise

\[ X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \]

a. \( \{Z_t\} \sim \text{IID}(\alpha) \) with Pareto tails

b. \( \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \)

No zeros inside the unit circle \( \Rightarrow \) invertible

Some zero(s) inside the unit circle \( \Rightarrow \) noninvertible
Realizations of an invertible and noninvertible MA(2) processes

Model: $X_t = \theta_*(B) Z_t$, $\{Z_t\} \sim \text{IID}(\alpha = 1)$, where

$\theta_i(B) = (1 + 1/2B)(1 + 1/3B)$ and $\theta_{ni}(B) = (1 + 2B)(1 + 3B)$
Application of all-pass to noninvertible MA model fitting

Suppose \( \{X_t\} \) follows the noninvertible MA model

\[
X_t = \theta_i(B) \theta_{ni}(B) Z_t, \quad \{Z_t\} \sim \text{IID}.
\]

Step 1: Let \( \{U_t\} \) be the residuals obtained by fitting a purely invertible MA model, i.e.,

\[
X_t = \hat{\theta}(B) U_t \\
\approx \theta_i(B) \tilde{\theta}_{ni}(B) U_t, \quad (\tilde{\theta}_{ni} \text{ is the invertible version of } \theta_{ni}).
\]

So

\[
U_t \approx \frac{\theta_{ni}(B)}{\tilde{\theta}_{ni}(B)} Z_t
\]

Step 2: Fit a purely causal AP model to \( \{U_t\} \)

\[
\tilde{\theta}_{ni}(B) U_t = \theta_{ni}(B) Z_t.
\]
Volumes of Microsoft (MSFT) stock traded over 755 transaction days (6/3/96 to 5/28/99)
Analysis of MSFT:

Step 1: Log(volume) follows MA(4).

\[ X_t = (1 + 0.513B + 0.277B^2 + 0.270B^3 + 0.202B^4) U_t \quad \text{(invertible MA(4))} \]

Step 2: All-pass model of order 4 fitted to \( \{U_t\} \) using MLE (t-dist):

\[
(1 - 0.628B - 0.229B^2 + 0.131B^3 - 0.202B^4)U_t = (1 - 0.649B + 1.135B^2 + 3.116B^3 - 4.960B^4)Z_{\hat{t}} \quad \text{for } \hat{\nu} = 6.26
\]

(Model using R-estimation is nearly the same.)

Conclude that \( \{X_t\} \) follows a noninvertible MA(4) which after refitting has the form:

\[ X_t = (1 + 1.34B + 1.374B^2 + 2.54B^3 + 4.96B^4) Z_t, \quad \{Z_t\} \sim \text{IID } t(6.3) \]
(a) ACF of Squares of Ut

(b) ACF of Absolute Values of Ut

(c) ACF of Squares of Zt

(d) ACF of Absolute Values of Zt
Summary: Microsoft Trading Volume

- Two-step fit of noninvertible MA(4):
  - invertible MA(4): residuals not iid
  - causal AP(4); residuals iid

- Direct fit of purely noninvertible MA(4):
  \((1+1.34B+1.374B^2+2.54B^3+4.96B^4)\)

- For MCHP, invertible MA(4) fits.
Summary

- All-pass models and their properties
  - linear time series with “nonlinear” behavior

- Estimation
  - likelihood approximation
  - MLE, LAD, R-estimation
  - order selection

- Empirical results
  - simulation study

- Noninvertible moving average processes
  - two-step estimation procedure using all-pass
  - noninvertible MA(4) for Microsoft trading volume
Further Work

- Least absolute deviations
  - further simulations
  - order selection
  - heavy-tailed case
  - other smooth objective functions (e.g., min dispersion)

- Maximum likelihood
  - Gaussian mixtures
  - simulation studies
  - applications

- Noninvertible moving average modeling
  - initial estimates from two-step all-pass procedure
  - adaptive procedures