Maximum Likelihood Estimation for Allpass Time Series Models

Richard A. Davis
Department of Statistics
Colorado State University

http://www.stat.colostate.edu/~rdavis/lectures/dekalb02.pdf

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

Estimation
- likelihood approximation
- MLE and LAD
- asymptotic results
- order selection

Empirical results
- simulation
- NZ/USA exchange rates

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
    $$\hat{\rho}_X(h) \text{ near 0 for all lags } h > 0 \text{ (MGD sequence)}$$  
  - $|X_t|$ and $X_t^2$ have slowly decaying autocorrelations
    $$\hat{\rho}_{|X|}(h) \text{ and } \hat{\rho}_{X^2}(h) \text{ converge to 0 slowly as } h \to \infty$$  
  - 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)
Motivating example
500-daily log-returns of NZ/US exchange rate
All-pass model of order 2 (t3 noise)
All-pass Models

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

Define MA polynomial:

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

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

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

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

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)

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

- 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
  - quasi-likelihood
  - least absolute deviations
  - 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)
Assume that $Z_t$ has density function $f_\sigma$ and consider the vector
\[ z = (X_{1-p}, \ldots, X_0, z_{1-p}(\phi), \ldots, z_0(\phi), z_1(\phi), \ldots, z_{n-p+1}(\phi), \ldots, z_n(\phi))' \]

**Joint density of $z$:**
\[
h(z) = h_1(X_{1-p}, \ldots, X_0, z_{1-p}(\phi), \ldots, z_0(\phi)) \cdot \left( \prod_{t=1}^{n-p} f_\sigma(\phi_q z_t(\phi)) \right) h_2(z_{n-p+1}(\phi), \ldots, z_n(\phi)),
\]

and hence the joint density of the data can be approximated by
\[
h(x) = \left( \prod_{t=1}^{n-p} f_\sigma(\phi_q z_t(\phi)) \right)
\]

where $q = \max \{0 \leq j \leq p: \phi_j \neq 0\}$. 


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

\[ f(z) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2} |z|) \]

and log-likelihood becomes

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

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

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

- \( \sigma \) a scale parameter
- mean 0, variance \( \sigma^2 \)

For \( f \) known, use maximum likelihood

- further smoothness assumptions (integrability, symmetry, etc.) on \( f \)
- Fisher information: \( \tilde{I} = \sigma^{-2} \int (f'(z))^2 / f(z) dz \)

For \( f \) unknown, use quasi-likelihood

Least absolute deviations

- assume \( f \) has median 0
- assume \( f \) continuous in neighborhood of 0
- act as if \( f = \text{Laplace} \) to get criterion function
Results

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

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

Maximum likelihood:

$$\sqrt{n}(\hat{\phi}_{\text{MLE}} - \phi_0) \overset{D}{\to} N(0, \frac{1}{2(\sigma^2 \tilde{I} - 1)\sigma^2 \Gamma_p^{-1}})$$

Least absolute deviations:

$$\sqrt{n}(\hat{\phi}_{\text{LAD}} - \phi_0) \overset{D}{\to} N(0, \frac{\text{Var}(|Z_1|)}{2\sigma^4 f_\sigma^2(0)} \sigma^2 \Gamma_p^{-1})$$
Further comments on MLE

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

Set

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

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

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

$I_f(\beta_0) = \int \frac{1}{f(z;\beta_0)} \frac{\partial f(z;\beta_0)}{\partial \beta_0} \frac{\partial f^T(z;\beta_0)}{\partial \beta_0} dz$ (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

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

Note: $\hat{\phi}_{\text{MLE}}$ is asymptotically independent of $\hat{\alpha}_{p+1,\text{MLE}}$ and $\hat{\beta}_{\text{MLE}}$
Identifiability in LAD case?

- Minimizer may not be unique.

- Gaussian case: \( \{ Z_t \} \) iid \( N(0, \sigma^2_0 \phi^{-2}_0) = N(0, \sigma^2_1 \phi^{-2}_1) \), so
  \[
  E | z_1(\phi_1) | = E \left| \frac{Z_1 \sigma_1}{\sigma_0 \phi_1} \right| = E \left| \frac{Z_1 \sigma_0}{\sigma_0 \phi_0} \right| = E | z_1(\phi_0) |
  \]

- Consider \( \{ c_j \} \) with at least two non-zero elements and
  \[
  \sum_{j=-\infty}^{\infty} | c_j | < \infty \quad \text{and} \quad \sum_{j=-\infty}^{\infty} c_j^2 = 1
  \]
  Jian and Pawitan (1998) show
  \[
  E | \sum_{j=-\infty}^{\infty} c_j Z_j | > E | Z_1 |
  \]
  holds for Laplace, Student’s t, contaminated normal, etc.

- Non-Gaussian case: \( E | z_1(\phi_1) | = E \left| \frac{\phi_0 (B^{-1}) \phi_1 (B)}{\phi_0 \phi_1 (B^{-1})} Z_t \right| > E | z_1(\phi_0) | \)
Central Limit Theorem (LAD case)

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

• Define

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

\[
= m_n(\phi_0 + n^{-1/2}u) - \sum_{t=1}^{n-p} |z_t(\phi_0)|
\]

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

\[
S(u) = \frac{f_\sigma(0)}{\phi_0 r} u' \Gamma_p u + u' N, \quad N \sim N(0, \frac{2 \text{Var}(|Z_1|)}{\phi_0^2 \sigma^2} \Gamma_p),
\]

• Hence,

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

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

\[
= -\frac{\phi_0 r}{2 f_\sigma(0)} \Gamma_p^{-1} N \sim N(0, \frac{\text{Var}(|Z_1|)}{2 \sigma^4 f_\sigma^4(0)} \phi_0^2 \Gamma_p^{-1})
\]
Asymptotic Results (LAD case):

**Theorem 1.** Let \( \{Y_t\} \) be the linear process

\[
Y_t = \sum_{j=-\infty}^{\infty} c_j z_{t-j},
\]

where \( c_0 = 0, \sum_{j=-\infty}^{\infty} |c_j| < \infty, \{z_t\} \sim \text{IID}(0, \sigma^2), \text{median}(z_1) = 0, \)

\( g(0) > 0 \) (g density of \( z_1 \)). Then

\[
S_n = \sum_{t=1}^{n-p} \left( |z_t - n^{-1/2} Y_t| - |z_t| \right)
\]

\[
\rightarrow \text{Var}(Y_1) g(0) + N
\]

where \( N \sim N(0, \gamma^*(0) + 2 \sum_{h \geq 1} \gamma^*(h)) \) and \( \gamma^*(h) \) is the covariance function for \( Y_t \, \text{sgn}(z_t) \).
Key idea:

\[
S_n = \sum_{t=1}^{n-p} \left( z_t - n^{-1/2} Y_t \mid - \mid z_t \right)
\]

\[
= -n^{-1/2} \sum_{t=1}^{n-p} Y_t \text{sgn}(z_t)
\]

\[
+ 2 \sum_{t=1}^{n-p} (n^{-1/2} Y_t - z_t) \left\{ 1_{\{0 < z_t < n^{-1/2} Y_t\}} - 1_{\{n^{-1/2} Y_t < z_t < 0\}} \right\}
\]

\rightarrow N + \text{Var}(Y_1) g(0)
Theorem 2. On $C(R^p)$,

$$S_n(u) = \sum_{t=1}^{n-p} \left( \left| z_t (\phi_0 + n^{-1/2} u) \right| - \left| z_t (\phi_0) \right| \right) \rightarrow S(u),$$

where

$$S(u) = \frac{f_{\sigma}(0)}{|\phi_{0r}|} u' \Gamma_p u + u' N,$$

$$N \sim N(0, \frac{2\text{Var}(|Z_1|)}{\phi_{0r}^2 \sigma^2} \Gamma_p),$$

and $\Gamma_p$ is the covariance matrix of a causal AR($p$).
Limit theory for LAD estimate. Note that

\[ \hat{\phi}_{LAD} = \phi_0 + \hat{u}_n / \sqrt{n} \]

so that

\[ \hat{u}_n = \sqrt{n} (\hat{\phi}_{LAD} - \phi_0) = \operatorname{arg\ min} S_n(u) \]

\[ \rightarrow \hat{u} = \operatorname{arg\ min} S(u). \]

Minimizing \( S \), we find that the minimizer or limit random variable is

\[ \hat{u}_n = \sqrt{n} (\hat{\phi}_{LAD} - \phi_0) \rightarrow - \frac{|\phi_{0r}| \Gamma_p^{-1}}{2 f_\sigma(0)} N \]

\[ - \frac{|\phi_{0r}| \Gamma_p^{-1}}{2 f_\sigma(0)} N \sim N(0, \frac{\text{Var}(|Z_1|)}{2\sigma^4 f_\sigma^2(0)} \sigma^2 \Gamma_p^{-1}) \]
Asymptotic Covariance Matrix

- For LS estimators of AR(p):

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

- For LAD estimators of AR(p):

\[
\sqrt{n}(\hat{\phi}_{LAD} - \phi_0) \xrightarrow{D} N(0, \frac{1}{4\sigma^2 f^2(0)} \sigma^2 \Gamma_p^{-1})
\]

- For LAD estimators of AP(p):

\[
\sqrt{n}(\hat{\phi}_{LAD} - \phi_0) \xrightarrow{D} N(0, \frac{\text{Var}(|Z_1|)}{2\sigma^4 f^2_\sigma(0)} \sigma^2 \Gamma_p^{-1})
\]

- For MLE estimators of AP(p):

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

\[
\frac{\text{Var}(|Z_1|)}{2\sigma^4 f^2_\sigma(0)} = \frac{1}{2} = \frac{1}{2(\sigma^2 \hat{I} - 1)}
\]

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

LAD: \[
\frac{\text{Var}(|Z_1|)}{2\sigma^4 f^2_\sigma(0)} = \frac{\Gamma^2 (\nu / 2)(\nu - 2)\pi}{2\Gamma^2 ((\nu + 1) / 2)} - \frac{2(\nu - 2)^2}{(\nu - 1)^2}
\]

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

Student’s \(t_3:\)

LAD: \(.7337\)

MLE: \(0.5\)

ARE: \(.7337 / .5 = 1.4674\)
Order Selection:

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

$$n^{1/2} \hat{\phi}_{p,LAD} \rightarrow N(0, \frac{\text{Var}(|Z|)}{2\sigma^4 f_\sigma^2(0)})$$

where $\hat{\phi}_{p,LAD}$ is the $p$th element of $\hat{\phi}_{LAD}$.

**Procedure:**

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

$$\theta^2 = \frac{\text{Var}(|Z_1|)}{2\sigma^4 f_\sigma^2(0)},$$

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.

Note: Can replace $\hat{\phi}_{p,p}$ with $\hat{\phi}_{p,MLE}$ if using MLE. In this case for $p > r$

$$n^{1/2} \hat{\phi}_{p,MLE} \rightarrow N(0, \frac{1}{2(\sigma^2 \hat{I} - 1)}).$$
AIC: 2p or not 2p?

- An approximately unbiased estimate of the Kullback-Leibler index of fitted to true model:

\[
AIC(p) := -2L_x(\hat{\phi}, \hat{\kappa}) + \frac{\text{Var}(|Z_1|)}{E|Z_1| \sigma^2 f_\sigma(0)} p
\]

- Penalty term for Laplace case:

\[
\frac{\text{Var}(|Z_1|)}{E|Z_1| \sigma^2 f_\sigma(0)} p = \frac{\sigma^2 / 2}{(\sigma / \sqrt{2})\sigma^2 (1 / \sqrt{2\sigma})} p = p
\]

- Estimated penalty term:

\[
\frac{\text{var}(|z_t(\hat{\phi})|)}{\text{ave}\{|z_t(\hat{\phi})|\} \hat{f}_{z_t(\hat{\phi})}(0)} p \xrightarrow{p} \frac{\text{Var}(|Z_1|)}{E|Z_1| \sigma^2 f_\sigma(0)} p
\]
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
Estimates:
\[ \hat{\phi}_1 = .297(.0381), \hat{\phi}_2 = .374(.0381) \]

Standard errors computed as \( \hat{\theta} \sqrt{(1 - \hat{\phi}_2^2) / 500} \)

where \( \hat{\theta} = .919 \)

Order selection:
- cut-off value for PACF is \( 1.96 \times .908 / \sqrt{500} = .0796 \)
- \( AIC(p) := -2L_x(\hat{\phi}, \hat{\kappa}) + 1.896p \)

<table>
<thead>
<tr>
<th>phi_p</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.289</td>
<td>0.374</td>
<td>0.009</td>
<td>0.011</td>
<td>0.01</td>
<td></td>
<td>0.047</td>
<td>0.034</td>
<td>-0.05</td>
<td>0.083</td>
<td>0.021</td>
</tr>
<tr>
<td>2451</td>
<td>2346</td>
<td>2347</td>
<td>2348</td>
<td>2350</td>
<td></td>
<td>2348</td>
<td>2349</td>
<td>2345</td>
<td>2343</td>
<td>2345</td>
</tr>
</tbody>
</table>
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

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

- 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</th>
<th></th>
<th>Empirical</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean std dev</td>
<td>mean std dev</td>
<td>%coverage</td>
<td>rel eff*</td>
</tr>
<tr>
<td>500</td>
<td>$\phi_1=.5$</td>
<td>.0332</td>
<td>.4979</td>
<td>.0397</td>
</tr>
<tr>
<td>5000</td>
<td>$\phi_1=.5$</td>
<td>.0105</td>
<td>.4998</td>
<td>.0109</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>Asymptotic</th>
<th></th>
<th>Empirical</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean std dev</td>
<td>mean std dev</td>
<td>%coverage</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>$\phi_1=.3$</td>
<td>.0351</td>
<td>.2990</td>
<td>.0456</td>
</tr>
<tr>
<td></td>
<td>$\phi_2=.4$</td>
<td>.0351</td>
<td>.3965</td>
<td>.0447</td>
</tr>
<tr>
<td>5000</td>
<td>$\phi_1=.3$</td>
<td>.0111</td>
<td>.3003</td>
<td>.0118</td>
</tr>
<tr>
<td></td>
<td>$\phi_2=.4$</td>
<td>.0111</td>
<td>.3990</td>
<td>.0117</td>
</tr>
</tbody>
</table>

*Efficiency relative to maximum absolute residual kurtosis:

$$
\frac{1}{n-p} \sum_{t=1}^{n-p} \left( \frac{z_t(\phi)}{v_2^{1/2}} \right)^4 \left| -3 \right|, \quad v_2 = \frac{1}{n-p} \sum_{t=1}^{n-p} \left( z_t(\phi) - \bar{z}(\phi) \right)^2
$$
### MLE Simulations Results using t-distr(3.5)

<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>φ₁=.5</td>
<td>0.0349</td>
<td>.4983</td>
<td>0.0421</td>
<td>91.7</td>
</tr>
<tr>
<td></td>
<td>ν=3.5</td>
<td>0.5853</td>
<td>3.449</td>
<td>0.4527</td>
<td>92.5</td>
</tr>
<tr>
<td></td>
<td>φ₁=.5</td>
<td>0.0110</td>
<td>.4997</td>
<td>0.0088</td>
<td>95.0</td>
</tr>
<tr>
<td></td>
<td>ν=3.5</td>
<td>0.1851</td>
<td>3.449</td>
<td>0.1341</td>
<td>96.7</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>φ₁=.3</td>
<td>0.0369</td>
<td>.2969</td>
<td>0.0451</td>
<td>89.6</td>
</tr>
<tr>
<td></td>
<td>φ₂=.4</td>
<td>0.0369</td>
<td>.3973</td>
<td>0.0446</td>
<td>90.6</td>
</tr>
<tr>
<td></td>
<td>ν=3.5</td>
<td>0.5853</td>
<td>3.556</td>
<td>0.5685</td>
<td>92.4</td>
</tr>
<tr>
<td></td>
<td>φ₁=.3</td>
<td>0.0117</td>
<td>.3002</td>
<td>0.0099</td>
<td>94.7</td>
</tr>
<tr>
<td></td>
<td>φ₂=.4</td>
<td>0.0117</td>
<td>.4001</td>
<td>0.0106</td>
<td>93.6</td>
</tr>
<tr>
<td></td>
<td>ν=3.5</td>
<td>0.1764</td>
<td>3.510</td>
<td>0.1764</td>
<td>94.7</td>
</tr>
</tbody>
</table>
**Minimum Dispersion Estimator:** 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>mean</td>
<td>.4978</td>
<td>.2988</td>
<td>mean</td>
</tr>
<tr>
<td></td>
<td>std dev</td>
<td>.0315</td>
<td>.0374</td>
<td>std dev</td>
</tr>
<tr>
<td>5000</td>
<td>mean</td>
<td>.4997</td>
<td>.2998</td>
<td>mean</td>
</tr>
<tr>
<td></td>
<td>std dev</td>
<td>.0094</td>
<td>.0374</td>
<td>std dev</td>
</tr>
</tbody>
</table>

**Empirical** | **Empirical LAD**
--- | ---
mean  | mean  | .4979 | .4997 |
std dev | std dev | .0397 | .0109 | .0456 | .0447 | .0118 | .0117
Application to financial data
500-daily log-returns of NZ/US exchange rate
All-pass model fitted to NZ-USA exchange rates (using LAD):

Order $= p = 6$, $\phi_1 = -0.367$, $\phi_2 = -0.750$, $\phi_3 = -0.391$, $\phi_4 = 0.088$, $\phi_5 = -0.193$, $\phi_6 = -0.096$

(AIC had local minima at $p=6$ and $10$)
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 \( \implies \) invertible

Some zero(s) inside the unit circle \( \implies \) 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

$\theta_1(B) = (1 + 1/2B)(1 + 1/3B)$ 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).

\[ X_t = (1 + .513B + .277B^2 + .270B^3 + .202B^4) U_t \quad \text{(invertible MA(4))} \]

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

\[
(1 + .184B + .132B^2 - .833B^3 - .314B^4)U_t \\
= (1 + 2.65B - .418B^2 - .586B^3 - 3.18B^4)Z_t.
\]

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 and LAD
  • order selection

❖ Empirical results
  • simulation study
  • AP(6) for NZ/USA exchange rates

❖ 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