# Modeling Noncausal Autoregressions Using All-Pass Filters

F. Jay Breidt

Department of Statistics
Colorado State University

http://www.stat.colostate.edu/~jbreidt/

June 2001

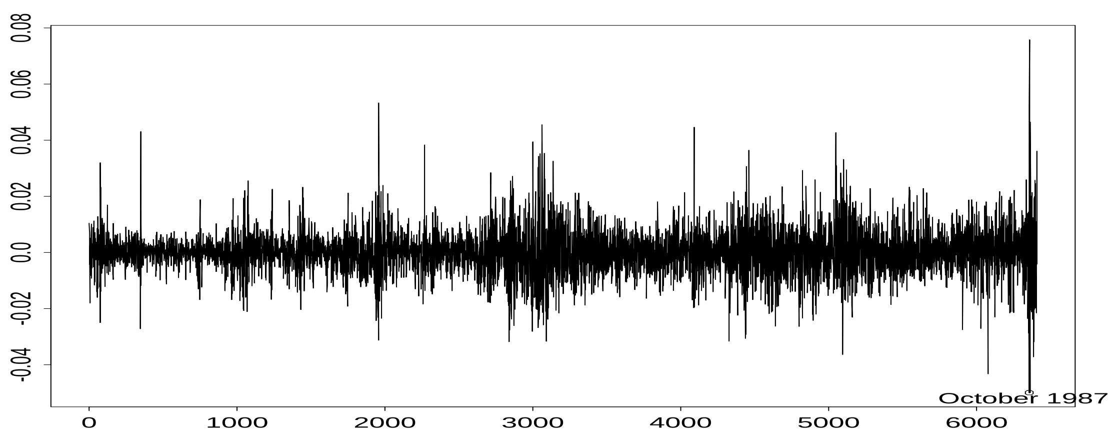
Joint work with Richard Davis and Beth Andrews,
Colorado State University,
and Alex Trindade, University of Florida

## Outline

- Introduction
  - 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
- Noncausal autoregressive processes
  - preliminaries
  - a two-step estimation procedure
  - Microsoft trading volume
- Summary

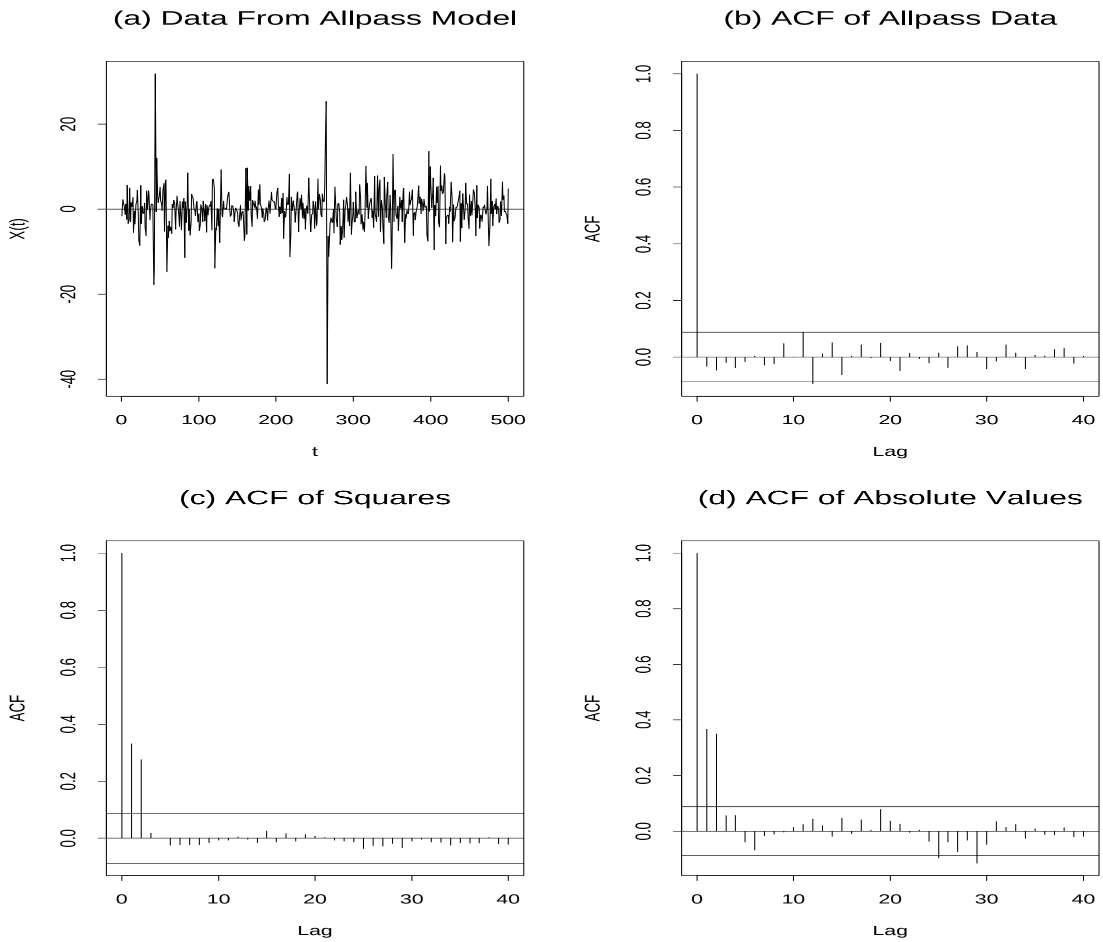
## Financial Time Series

- Returns on financial assets often exhibit:
  - lack of serial correlation
  - heavy-tailed marginal distributions
  - bursts of outliers / volatility clustering
  - dependence outside 2nd-order moment structure
- Nonlinear models: $X_t = \sigma_t Z_t$
  - ARCH and its variants (Engle 1982; Bollerslev, Chou, and Kroner 1992)
  - Stochastic volatility (Clark 1973; Taylor 1986)



# A Simulated Example

- White noise with "volatility clustering"



# Autoregressive Moving Average Models

- ARMA$(p, q)$ model:

$$\begin{aligned} X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} &= Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \\ \phi(B) X_t &= \theta(B) Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2) \end{aligned}$$

  where

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p \\ \theta(B) &= 1 + \theta_1 B + \cdots + \theta_q B^q \\ B^k X_t &= X_{t-k} \end{aligned}$$

- A broad class of linear models
  - *stationary:* no roots of $\phi(z)$ on unit circle
  - *causal:* no roots of $\phi(z)$ inside unit circle

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

  - *invertible:* no roots of $\theta(z)$ inside unit circle

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

## A Special ARMA Model

- All-pass model of order 1:

$$\begin{aligned}(1-\phi B)X_t &= (1-\phi^{-1}B)Z_t \\ &= -\phi^{-1}B(1-\phi B^{-1})Z_t\end{aligned}$$

  where $|\phi| < 1$ and $\{Z_t\}$ iid $f_\sigma$, mean 0, variance $\sigma^2$

- Causal, noninvertible ARMA(1,1)

## All-Pass Models

- Set up for later order selection:

$$\begin{aligned} r &= \text{unknown } r\text{eal model order} \\ &\leq s = \text{known } s\text{ufficiently large model order} \\ p &= p\text{roposed model order} \leq s \end{aligned}$$

- Causal AR($s$) polynomial:

$$\phi_0(z) = 1 - \phi_{01}z - \cdots - \phi_{0s}z^s$$

where $\phi_{0r} \neq 0$ and $\phi_{0j} = 0$ for $r < j \leq s$

- Causal AP($r$) is $\{X_t\}$ satisfying

$$\phi_0(B)X_t = \frac{B^s\phi_0(B^{-1})}{-\phi_{0r}}Z_t,\ \{Z_t\} \text{ iid} f_\sigma$$

- Causal, noninvertible ARMA($r, r$)

## Properties of All-Pass Models

- Infinite moving average:

$$X_t = \frac{B^s \phi_0(B^{-1})}{-\phi_{0r}\phi_0(B)} Z_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

- Spectral density of AP($r$) is

$$\frac{|e^{-is\omega}|^2 |\phi_0(e^{i\omega})|^2}{\phi_{0r}^2 |\phi_0(e^{-i\omega})|^2} \frac{\sigma^2}{2\pi} = \frac{\sigma^2}{\phi_{0r}^2 2\pi}$$

- $\{X_t\}$ ∼AP($r$) is
  - zero mean
  - serially uncorrelated (flat spectrum)
  - dependent if $\{Z_t\}$ is non-Gaussian
  - heavy-tailed if $\{Z_t\}$ is heavy-tailed
- Linear time series with “non-linear” behavior
  - illustration of general result (Bickel and Bühlmann, 1996)

## Gaussian Case

- $\{X_t\} \sim$AP$(r)$ is WN$(0, \sigma_0^2\phi_{0r}^{-2})$
  - zero mean
  - serially uncorrelated
- If $\{Z_t\}$ iid N$(0, \sigma_0^2)$ then

$$\begin{aligned} \{X_t\} \;\; & \text{iid} \;\; \mathrm{N}\left(0, \sigma_0^2\phi_{0r}^{-2}\right) \\ & = \;\; \mathrm{N}\left(0, \sigma_1^2\phi_{1p}^{-2}\right) \end{aligned}$$

  provided $\sigma_0^2\phi_{0r}^{-2} = \sigma_1^2\phi_{1p}^{-2}$
- Non-identifiable

## 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

## Estimation Preliminaries

- Write $z_t = Z_t \phi_{0r}^{-1}$ so

$$\phi_0(B)X_t = -\phi_0(B^{-1})z_{t-s}$$

- Note the backward recursion

$$z_{t-s} = \phi_{01}z_{t-s+1} + \cdots + \phi_{0s}z_t - \phi_0(B)X_t$$

- For causal $\phi(z) = 1 - \phi_1 z - \cdots - \phi_s z^s$ set

$$z_n(\boldsymbol{\phi}) = z_{n-1}(\boldsymbol{\phi}) = \cdots = z_{n-s+1}(\boldsymbol{\phi}) = 0$$

  and use the backward recursion

$$z_{t-s}(\boldsymbol{\phi}) = \phi_1 z_{t-s+1}(\boldsymbol{\phi}) + \cdots + \phi_s z_t(\boldsymbol{\phi}) - \phi(B)X_t$$

  for $t = n, \ldots, s+1$

## Linear Transformation

- Define "data" vector:

$$\mathbf{x} := (X_{1-s}, \ldots, X_0, X_1, \ldots, X_n, z_{n-s+1}(\boldsymbol{\phi}), \ldots, z_n(\boldsymbol{\phi}))'$$

  and "noise" vector

$$\mathbf{z} := (X_{1-s}, \ldots, X_0, z_{1-s}(\boldsymbol{\phi}), \ldots, z_{n-s+1}(\boldsymbol{\phi}), \ldots, z_n(\boldsymbol{\phi}))'$$

- Transformation:

$$A\mathbf{x} = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ -\phi_s & \cdots & -\phi_1 & 1 & & & & & \\ & \ddots & & \ddots & \ddots & & & & \\ & & -\phi_s & \cdots & -\phi_1 & 1 & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{bmatrix} \begin{bmatrix} X_{1-s} \\ \vdots \\ X_0 \\ X_1 \\ \vdots \\ X_n \\ z_{n-s+1}(\boldsymbol{\phi}) \\ \vdots \\ z_n(\boldsymbol{\phi}) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ \phi_s & \cdots & \phi_1 & -1 & & & & & \\ & \ddots & & \ddots & \ddots & & & & \\ & & \phi_s & \cdots & \phi_1 & -1 & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{bmatrix} \begin{bmatrix} X_{1-s} \\ \vdots \\ X_0 \\ z_{1-s}(\boldsymbol{\phi}) \\ \vdots \\ z_{n-s}(\boldsymbol{\phi}) \\ z_{n-s+1}(\boldsymbol{\phi}) \\ \vdots \\ z_n(\boldsymbol{\phi}) \end{bmatrix} = B\mathbf{z}$$

  where $|A| = 1$ and $|B| = 1$

## Approximating the Likelihood

- Joint distribution of $\mathbf{z}$ under $\boldsymbol{\phi}$:

$$\begin{aligned} h(\mathbf{z}) &= h_1(X_{1-s}, \ldots, X_0, z_{1-s}(\boldsymbol{\phi}), \ldots, z_0(\boldsymbol{\phi})) \\ &\quad \times \left( \prod_{t=1}^{n-s} f_\sigma \left( \phi_p z_t(\boldsymbol{\phi}) \right) |\phi_p| \right) h_2(z_{n-s+1}(\boldsymbol{\phi}), \ldots, z_n(\boldsymbol{\phi})) \end{aligned}$$

  where $p = \max\{0 \le j \le s : \phi_j \ne 0\}$

- Joint distribution of $\mathbf{x}$ under $\boldsymbol{\phi}$:

$$h(\mathbf{x}) = h_1 \times \left( \prod_{t=1}^{n-s} f_\sigma \left( \phi_p z_t(\boldsymbol{\phi}) \right) |\phi_p| \right) \times h_2$$

- Log-likelihood approximation:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\phi}, \sigma) &= \sum_{t=1}^{n-s} \ln f_\sigma \left( \phi_p z_t(\boldsymbol{\phi}) \right) + (n-s) \ln |\phi_p| \\ &= (n-s) \ln \sigma^{-1} |\phi_p| + \sum_{t=1}^{n-s} \ln f(\sigma^{-1} |\phi_p| z_t(\boldsymbol{\phi})) \end{aligned}$$

## Assumptions

- Assume $\{Z_t\}$ 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 assumptions on $f$
  - Fisher information: $\tilde{I} = \sigma^{-2} \int (f'(w))^2 / f(w)\, dw$
- For $f$ unknown, use quasi-likelihood
- Least absolute deviations
  - assume $f$ has median 0
  - assume $f$ continuous in neighborhood of zero
  - act as if $f$ = Laplace to get criterion function

## Results

- Let $\gamma(h) =$ ACVF of AR $\phi_0(\cdot)$ and

$$\mathbf{\Gamma}_s = [\gamma(j-k)]_{j,k=1}^{s}$$

- Maximum likelihood:

$$n^{1/2}(\hat{\boldsymbol{\phi}}_{MLE} - \boldsymbol{\phi}_0) \xrightarrow{\mathcal{L}} \mathrm{N}\left(\mathbf{0}, \frac{\sigma^{-2}}{2(\tilde{I} - \sigma^{-2})}\sigma^2 \mathbf{\Gamma}_s^{-1}\right)$$

- Least absolute deviations:

$$n^{1/2}(\hat{\boldsymbol{\phi}}_{LAD} - \boldsymbol{\phi}_0) \xrightarrow{\mathcal{L}} \mathrm{N}\left(\mathbf{0}, \frac{\mathrm{Var}\,(|Z_1|)}{2\sigma^4 f_\sigma^2(0)}\sigma^2 \mathbf{\Gamma}_s^{-1}\right)$$

# Least Absolute Deviations

- Laplacian noise with variance $\sigma^2$:

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

- Log-likelihood:

$$\text{constant} - (n-s)\ln\kappa - \sum_{t=1}^{n-s} \frac{\sqrt{2}|z_t(\boldsymbol{\phi})|}{\kappa}$$

  where $\kappa = \sigma|\phi_p|^{-1}$

- LAD estimator of $\kappa$:

$$\hat{\kappa} = \frac{\sqrt{2}}{n-s} \sum_{t=1}^{n-s} |z_t(\hat{\boldsymbol{\phi}})|$$

- Concentrated Laplacian likelihood:

$$\ell(\boldsymbol{\phi}) = \text{constant} - (n-s)\ln \sum_{t=1}^{n-s} |z_t(\boldsymbol{\phi})|$$

- Equivalently, minimize absolute deviations

$$m_n(\boldsymbol{\phi}) = \sum_{t=1}^{n-s} |z_t(\boldsymbol{\phi})|$$

## Identifiability?

- Minimizer may not be unique
- Gaussian case: $\{X_t\}$ iid $\mathrm{N}(0, \sigma_0^2 \phi_{0r}^{-2}) = \mathrm{N}(0, \sigma_1^2 \phi_{1p}^{-2})$, so

$$\mathrm{E}|z_1(\boldsymbol{\phi}_1)| = \mathrm{E}\Big|\frac{Z_1 \sigma_1}{\sigma_0 \phi_{1p}}\Big| = \mathrm{E}\Big|\frac{Z_1 \sigma_0}{\sigma_0 \phi_{0r}}\Big| = \mathrm{E}|z_1(\boldsymbol{\phi}_0)|$$

- Consider $\{c_j\}$ with at least two non-zero elements and

$$\sum_j |c_j| < \infty \text{ and } \sum_j c_j^2 = 1$$

Jian and Pawitan (1998) show

$$\mathrm{E}\Big|\sum_{j=-\infty}^{\infty} c_j Z_{t-j}\Big| > \mathrm{E}|Z_1|$$

holds for Laplace, Student's $t$, contaminated normal, etc.

- Non-Gaussian case:

$$\mathrm{E}|z_1(\boldsymbol{\phi}_1)| = \mathrm{E}\Big|\frac{\phi_0(B^{-1})\phi_1(B)}{\phi_{0r}\phi_1(B^{-1})\phi_0(B)} Z_t\Big| > \mathrm{E}|z_1(\boldsymbol{\phi}_0)|$$

# Central Limit Theorem

- Think of $\mathbf{u} = n^{1/2}(\boldsymbol{\phi} - \boldsymbol{\phi}_0) \in \mathbb{R}^s$
- Define

$$S_n(\mathbf{u}) \;=\; m_n(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u}) - \sum_{t=1}^{n-s} |z_t(\boldsymbol{\phi}_0)|$$

- Then $S_n \xrightarrow{\mathcal{L}} S$ on $C(\mathbb{R}^s)$ where

$$S(\mathbf{u}) = \frac{f_\sigma(0)}{|\phi_{0r}|}\mathbf{u}'\boldsymbol{\Gamma}_s\mathbf{u} + \mathbf{u}'\mathbf{N},\;\; \mathbf{N} \sim \mathrm{N}\left(\mathbf{0}, \frac{2\mathrm{Var}\left(|Z_1|\right)}{\phi_{0r}^2\sigma^2}\boldsymbol{\Gamma}_s\right)$$

- Hence,

$$\begin{aligned}
\operatorname{argmin} S_n(\mathbf{u}) \;&=\; n^{1/2}(\hat{\boldsymbol{\phi}}_{LAD} - \boldsymbol{\phi}_0) \\
&\xrightarrow{\mathcal{L}}\; \operatorname{argmin} S(\mathbf{u}) \\
&=\; -\frac{|\phi_{0r}|\boldsymbol{\Gamma}_s^{-1}}{2f_\sigma(0)}\mathbf{N} \\
&\sim\; \mathrm{N}\left(\mathbf{0}, \frac{\mathrm{Var}\left(|Z_1|\right)}{2\sigma^4 f_\sigma^2(0)}\sigma^2\boldsymbol{\Gamma}_s^{-1}\right)
\end{aligned}$$

## Sketch of Proof

- For $z \neq 0$,

$$|z-y|-|z| = -y\,\mathrm{sgn}\,(z) + 2(y-z)\left\{\mathbf{1}_{\{0<z<y\}} - \mathbf{1}_{\{y<z<0\}}\right\}$$

- Define

$$Y_t = \mathbf{u}'\left[-\frac{z_{t-j}}{\phi_0(B)} + \frac{z_{t+j}}{\phi_0(B^{-1})}\right]_{j=1}^{s}$$

- Then $z_t(\boldsymbol{\phi}) = z_t(\boldsymbol{\phi}_0 + n^{-1/2}\mathbf{u}) \simeq z_t - n^{-1/2}Y_t$, so

$$\begin{aligned} S_n(\mathbf{u}) &\simeq \sum_{t=1}^{n-s}\left\{|z_t - n^{-1/2}Y_t| - |z_t|\right\} \\ &= -n^{-1/2}\sum_{t=1}^{n-s} Y_t\,\mathrm{sgn}\,(z_t) \\ &\quad +2\sum_{t=1}^{n-s}\left(n^{-1/2}Y_t - z_t\right)\left\{\mathbf{1}_{\{0<z_t<n^{-1/2}Y_t\}} - \mathbf{1}_{\{n^{-1/2}Y_t<z_t<0\}}\right\} \\ &\xrightarrow{\mathcal{L}} N + \text{constant}, \end{aligned}$$

where

$$\begin{aligned} N &\sim \mathrm{N}\left(0, \gamma^*(0) + 2\sum_{h=1}^{\infty}\gamma^*(h)\right) \\ &= \mathrm{N}\left(0, \frac{2\mathrm{Var}\,(|Z_1|)}{\phi_{0r}^2\sigma^2}\mathbf{u}'\boldsymbol{\Gamma}_s\mathbf{u}\right) \end{aligned}$$

## Asymptotic Covariance Matrix

- For LS estimators of AR($r$),

$$n^{1/2}(\hat{\boldsymbol{\phi}}_{LS} - \boldsymbol{\phi}_0) \xrightarrow{\mathcal{L}} \mathrm{N}\left(\mathbf{0}, \sigma^2 \mathbf{\Gamma}_r^{-1}\right)$$

- For LAD estimators of AP($r$),

$$n^{1/2}(\hat{\boldsymbol{\phi}}_{LAD} - \boldsymbol{\phi}_0) \xrightarrow{\mathcal{L}} \mathrm{N}\left(\mathbf{0}, \frac{\mathrm{Var}\,(|Z_1|)}{2\sigma^4 f_\sigma^2(0)} \sigma^2 \mathbf{\Gamma}_r^{-1}\right)$$

  – Laplace:

$$\frac{\mathrm{Var}\,(|Z_1|)}{2\sigma^4 f_\sigma^2(0)} = \frac{1}{2}$$

  – Student's $t_\nu$, $\nu > 2$:

$$\frac{\mathrm{Var}\,(|Z_1|)}{2\sigma^4 f_\sigma^2(0)} = \frac{\Gamma^2(\nu/2)(\nu-2)\pi}{2\Gamma^2((\nu+1)/2)} - \frac{2(\nu-2)^2}{(\nu-1)^2}$$

  – Student's $t_3$: 0.7337

## Order Selection

- True model is AP($r$) and fitted model is AP($s$), $s > r$:

$$n^{1/2}\hat{\phi}_{s,LAD} \xrightarrow{\mathcal{L}} N\left(0, \frac{\text{Var}\,(|Z_1|)}{2\sigma^4 f_\sigma^2(0)}\right)$$

- Model selection procedure:

  1. Fit AP($s$), $s$ large and obtain residuals $\{z_t(\hat{\boldsymbol{\phi}})\}$

$$\begin{aligned}\hat{\theta}^2 &= \frac{\text{var}\{|z_t(\hat{\boldsymbol{\phi}})|\}}{2\left[\text{var}\{z_t(\hat{\boldsymbol{\phi}})\}\right]^2\left\{\hat{f}_{z_t(\hat{\phi})}(0)\right\}^2} \\ &\xrightarrow{P} \frac{\text{Var}\,(|Z_1|)\,\phi_{0r}^{-2}}{2(\sigma^2\phi_{0r}^{-2})^2\left\{|\phi_{0r}|f_\sigma(0)\right\}^2} \\ &= \frac{\text{Var}\,(|Z_1|)}{2\sigma^4 f_\sigma^2(0)}\end{aligned}$$

  2. Fit AP($p$) $p = 1, 2, \ldots, s$ via LAD and obtain $\hat{\phi}_{pp}$
  3. Choose the model order $\hat{r}$:

$$\hat{r} = \min\{0 \le p \le s : |\hat{\phi}_{jj}| < 1.96\hat{\theta}n^{-1/2} \text{ for } j > p\}$$

## AIC: $2p$ or not $2p$?

- Approximately unbiased estimator of the Kullback-Leibler index of fitted to true model:

$$\text{AIC}(p) := -2\mathcal{L}_X(\hat{\boldsymbol{\phi}}, \hat{\kappa}) + \frac{\text{Var}|Z_1|}{\text{E}|Z_1|\sigma^2 f_\sigma(0)} p$$

- Penalty term for Laplace case:

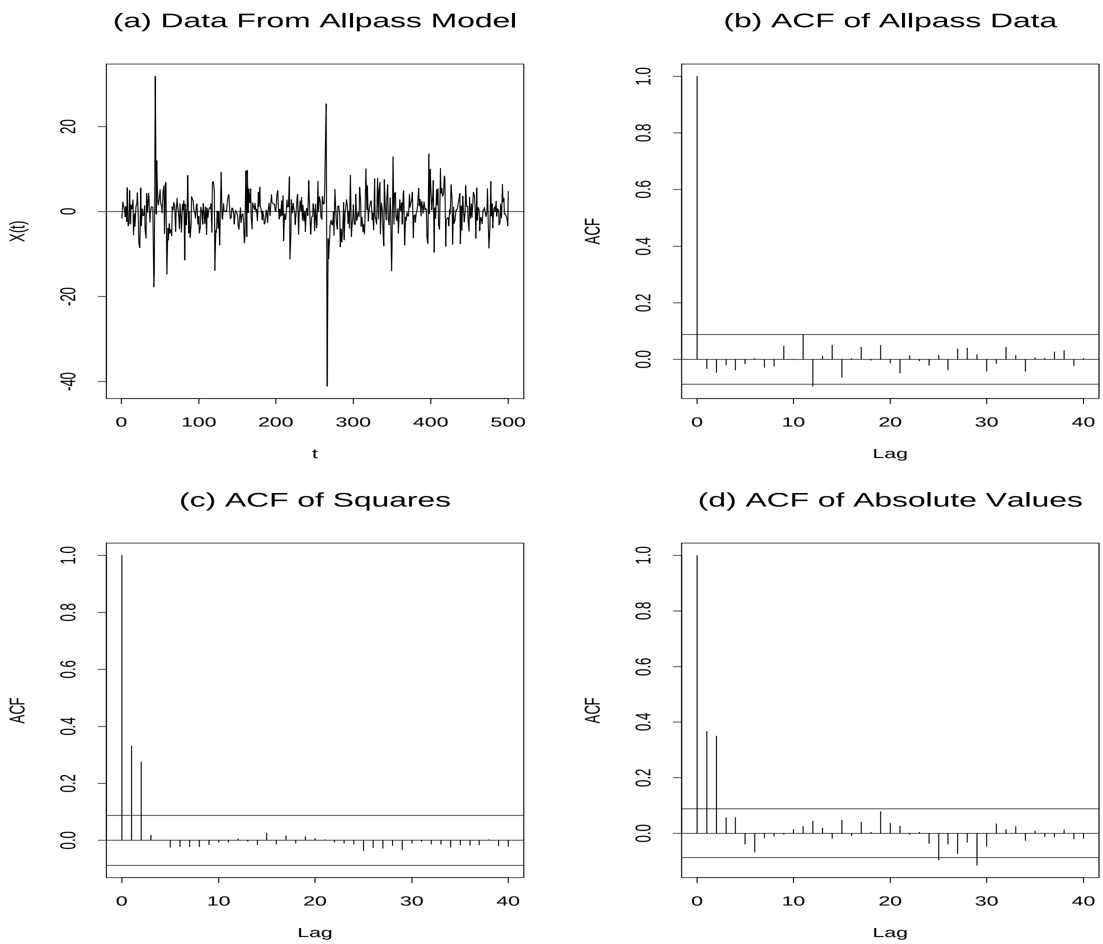
$$\frac{\text{Var}|Z_1|}{\text{E}|Z_1|\sigma^2 f_\sigma(0)} p = \frac{\sigma^2/2}{(\sigma/\sqrt{2})\sigma^2(1/\sqrt{2}\sigma)} p = p$$

- Estimated penalty term:

$$\frac{\text{var}\{|z_t(\hat{\boldsymbol{\phi}})|\}}{\text{ave}\{|z_t(\hat{\boldsymbol{\phi}})|\}\text{var}\{z_t(\hat{\boldsymbol{\phi}})\}\hat{f}_{z_t(\hat{\phi})}(0)} p \xrightarrow{P} \frac{\text{Var}|Z_1|}{\text{E}|Z_1|\sigma^2 f_\sigma(0)} p$$

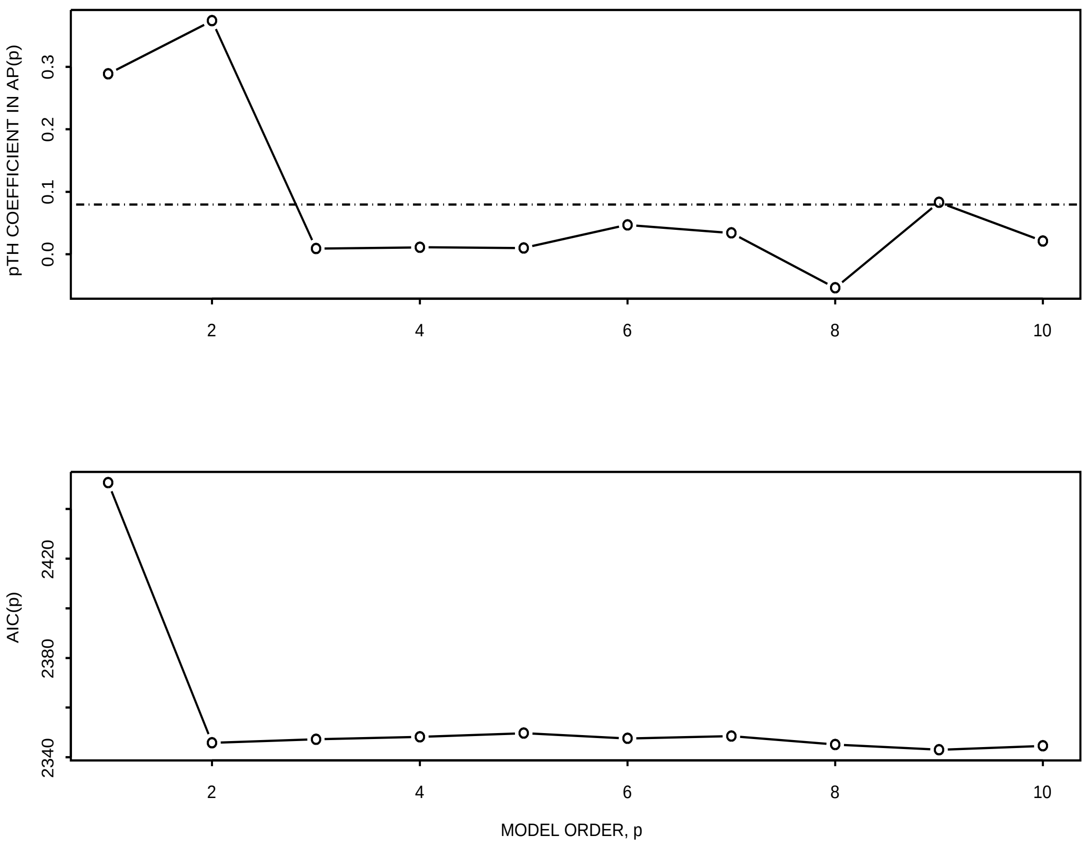
# Example: Simulated AP(2)

- $\phi_1 = 0.3$, $\phi_2 = 0.4$, $n = 500$, noise is iid $t_3$



## Example, Continued: Simulated AP(2)

- Truth: $\phi_1 = 0.3$, $\phi_2 = 0.4$, $n = 500$, noise is iid $t_3$
- Order selection:




- Estimates: $\hat{\phi}_1 = 0.297\ (0.0381)$; $\hat{\phi}_2 = 0.374\ (0.0381)$

## Simulation Study

- AP(1):

$$(1-\phi_{01}B)X_t = \frac{B}{-\phi_{01}}(1-\phi_{01}B^{-1})Z_t$$

in which

$$\hat{\phi}_{LAD} \text{ is AN}\left(\phi_{01}, \frac{\text{Var}\,(|Z_1|)}{2\sigma^4 f_\sigma^2(0)}\frac{\sigma^2}{n}(1-\phi_{01}^2)\right)$$

- AP(2):

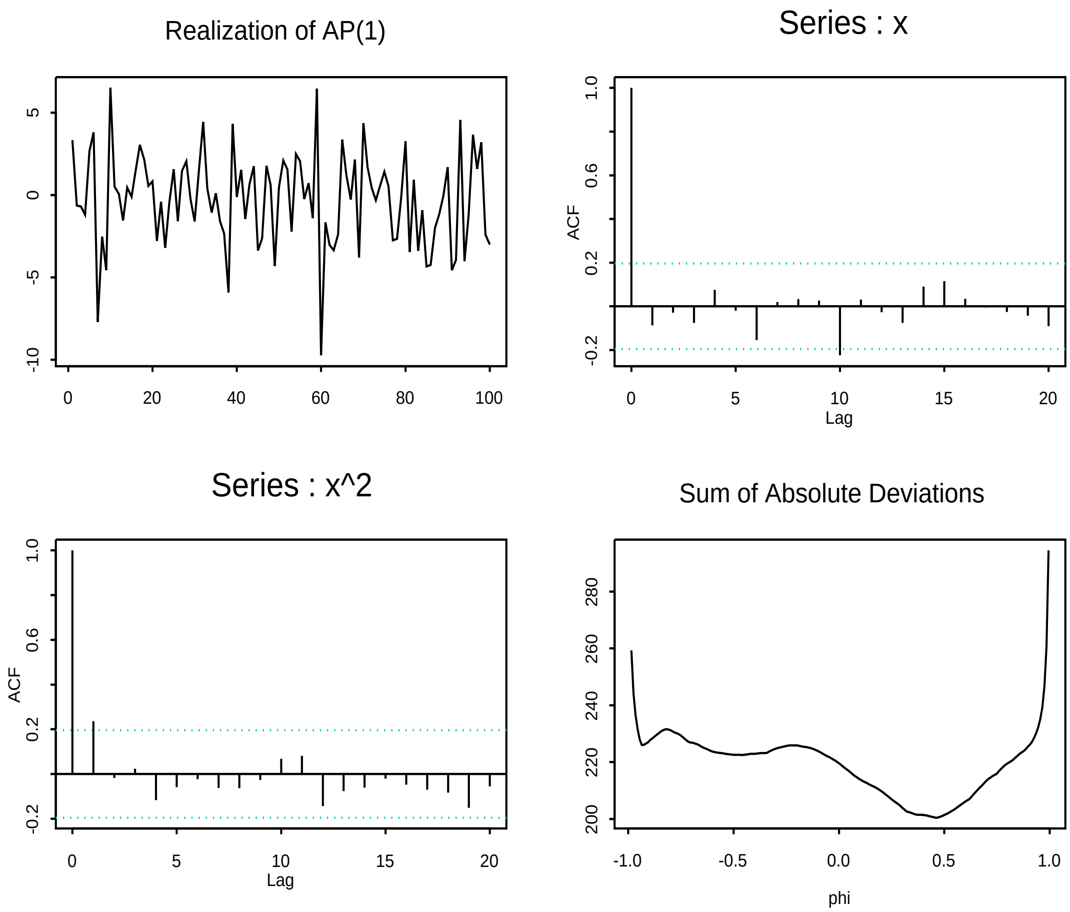
$$(1-\phi_{01}B-\phi_{02}B^2)X_t = \frac{B^2}{-\phi_{02}}(1-\phi_{01}B^{-1}-\phi_{02}B^{-2})Z_t$$

in which

$$\hat{\boldsymbol{\phi}}_{LAD} \text{ is AN}\left(\boldsymbol{\phi}_0, \frac{\text{Var}\,(|Z_1|)}{2\sigma^4 f_\sigma^2(0)}\frac{\sigma^2}{n}\begin{bmatrix} 1-\phi_{02}^2 & -\phi_{01}(1+\phi_{02}) \\ -\phi_{01}(1+\phi_{02}) & 1-\phi_{02}^2 \end{bmatrix}\right)$$

# Simulation Worries

- Local minima in criterion function



## Simulation Details

- Starting values: distributed uniformly in PACF space
- PACF mapped to AR coefficients via Durbin-Levinson:
  1. Draw $\phi_{11}^{(k)}, \phi_{22}^{(k)}, \ldots, \phi_{rr}^{(k)}$ iid uniform$(-1, 1)$
  2. For $j = 2, \ldots, r$, compute

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

- Obtain values for $k = 1, 2, \ldots, 250$; then pare to 10 best
- Use Hooke and Jeeves starting from 10 best and choose overall best

# Simulation Results: AP(1)

- Noise distribution is $t_3$; 1000 replications

| | Asymptotic | | Empirical | | | |
|---|---|---|---|---|---|---|
| $n$ | mean | std.dev. | mean | std.dev. | % coverage | rel. effic. |
| 500 | $\phi_{01} = 0.5$ | 0.0332 | 0.4979 | 0.0397 | 94.2 | 11.4 |
| 5000 | $\phi_{01} = 0.5$ | 0.0105 | 0.4998 | 0.0109 | 95.4 | 9.3 |
| 500 | $\phi_{01} = 0.9$ | 0.0167 | 0.8834 | 0.1027 | 91.2 | 1.2 |
| 5000 | $\phi_{01} = 0.9$ | 0.0053 | 0.8993 | 0.0056 | 95.7 | 67.6 |

- Efficiency relative to maximum absolute residual kurtosis

$$\left|\frac{1}{n-s}\sum_{t=1}^{n-s}\left(\frac{z_t(\boldsymbol{\phi})}{\hat{v}_2^{1/2}}\right)^4 - 3\right|$$

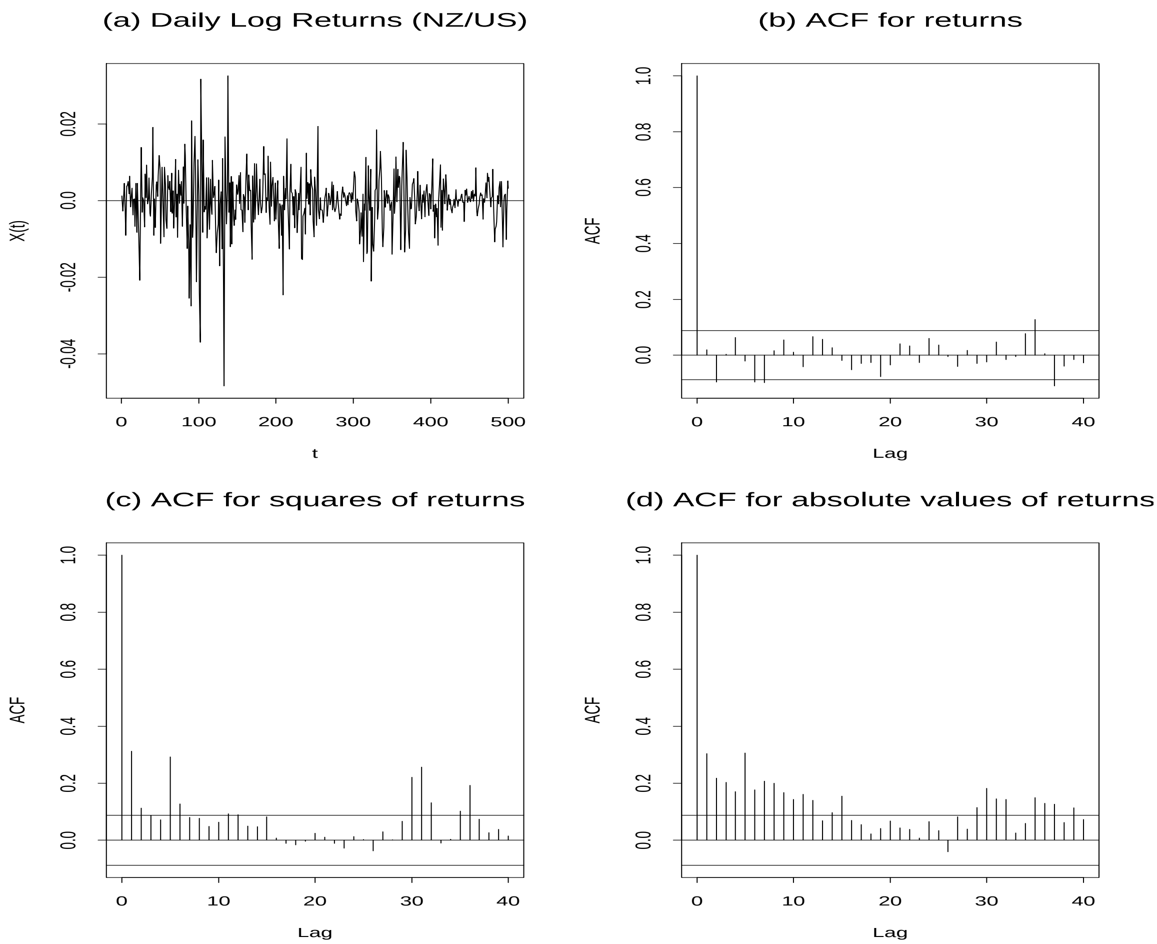
## Simulation Results: AP(2)

- Noise distribution is $t_3$; 1000 replications

| | Asymptotic | | Empirical | | |
|---|---|---|---|---|---|
| $n$ | mean | std.dev. | mean | std.dev. | % coverage |
| 500 | $\phi_{01} = 0.3$ | 0.0351 | 0.2990 | 0.0456 | 92.5 |
| | $\phi_{02} = 0.4$ | 0.0351 | 0.3965 | 0.0447 | 92.1 |
| 5000 | $\phi_{01} = 0.3$ | 0.0111 | 0.3003 | 0.0118 | 95.5 |
| | $\phi_{02} = 0.4$ | 0.0111 | 0.3990 | 0.0117 | 94.7 |

## Linear Time Series with "Nonlinear" Behavior

- 500 daily returns of New Zealand/US exchange rate
  - serially uncorrelated
  - heavy-tailed marginal
  - volatility clustering
- Try all-pass as linear alternative to ARCH, stochastic volatility

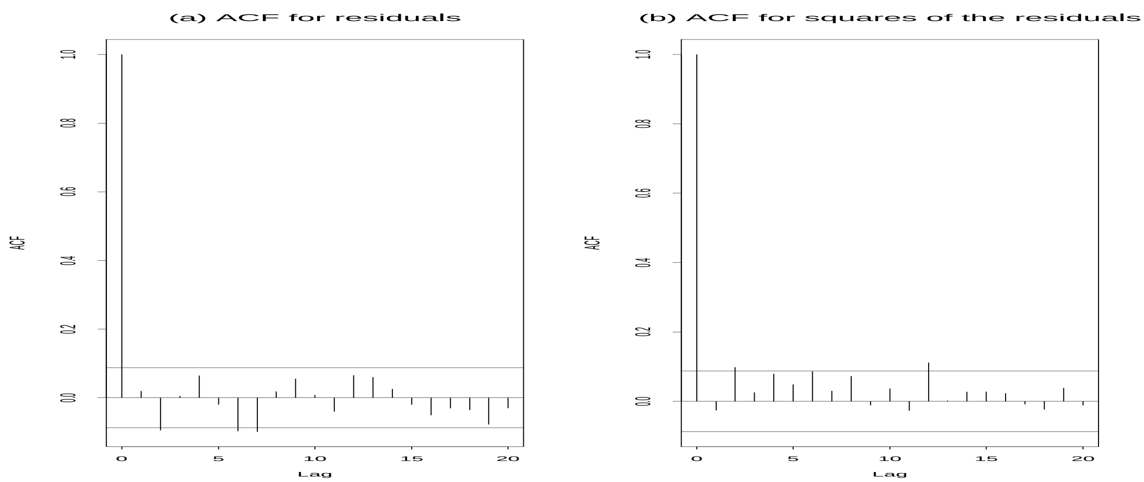


## New Zealand/US Exchange Rate

- Select AP(6):

  $$1 + 0.367B + 0.75B^2 + 0.391B^3 - .088B^4 + 0.193B^5 + 0.096B^6$$

- Linear model can mimic “nonlinear” behavior

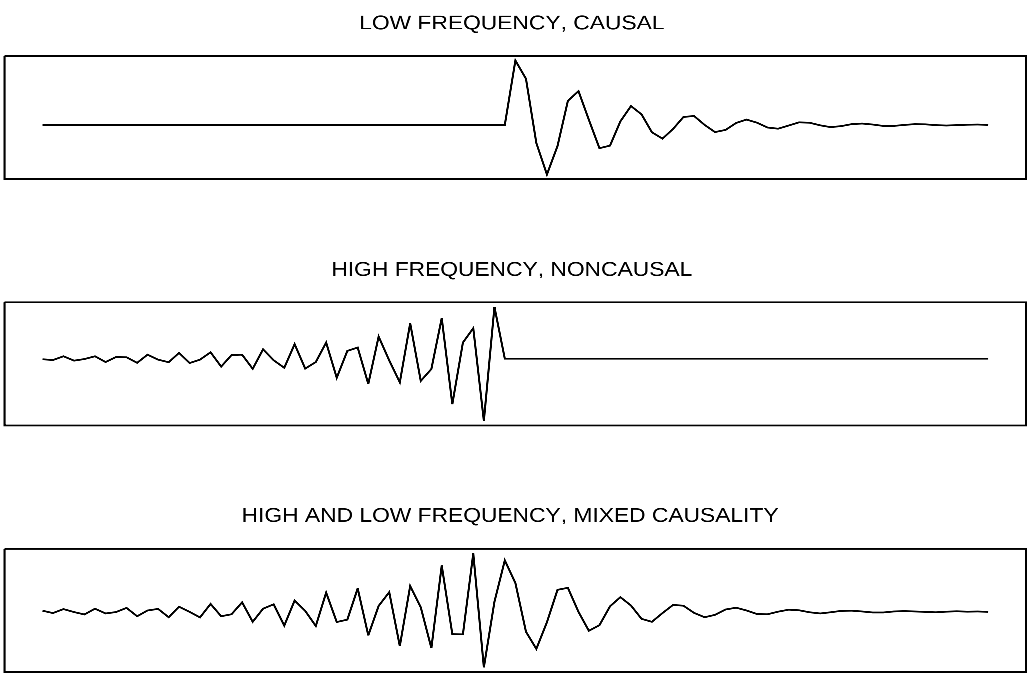


## Estimation for Noncausal Autoregressive Processes

- Introduction and definitions
- Maximum likelihood estimation
- Two-step procedure using all-pass
  - order selection
  - preliminary estimation
- Application to Microsoft trading volume

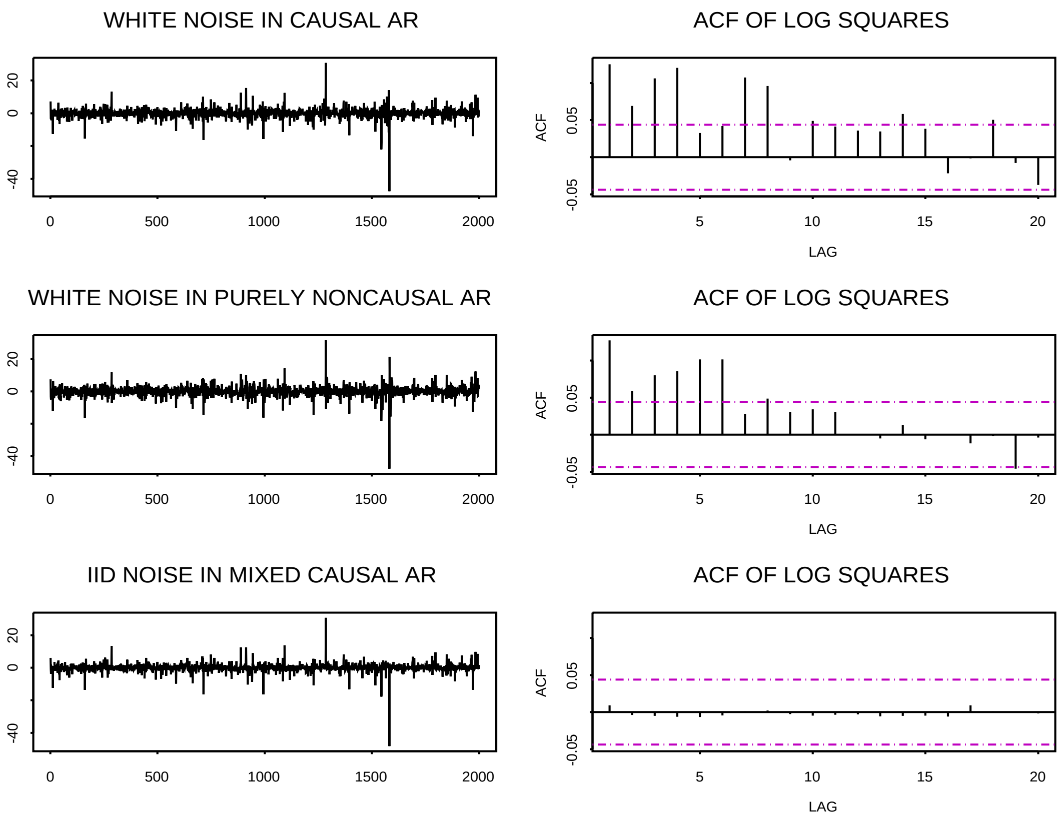
## Noncausal Autoregressive Processes

- Definitions: AR with polynomial $\phi(\cdot)$ is
  - *stationary* if no roots on unit circle
  - *causal* if no roots inside unit circle
  - *purely noncausal* if all roots inside unit circle
  - *mixed noncausal* if some roots inside unit circle
- Deconvolution problems: seismic, astronomical, speech
- Impulse response functions:



# Second-Order Equivalent Representations

- $2^{\#\text{ roots}}$ AR's with differing causality
  - different white noise sequences
  - only one is iid in non-Gaussian case



## Estimation for Noncausal Autoregressions

- Mixed AR($s$) = AR($q$)AR($r$):

$$\phi(B)X_t = \phi_c(B)\phi_{nc}(B)X_t = Z_t, \quad \{Z_t\} \text{ iid}$$

- Maximum likelihood: Breidt, Davis, Lii, and Rosenblatt (1991)
  - maximize criterion function over all $2^s$ possible root configurations
  - cumbersome for higher-order models
- Alternative: note second-order equivalent causal representation:

$$\begin{aligned} U_t &= \phi_c(B)\phi_{nc}^{(c)}(B)X_t \\ &= \phi_c(B)\phi_{nc}^{(c)}(B)\frac{Z_t}{\phi_c(B)\phi_{nc}(B)} \\ &= \frac{\phi_{nc}^{(c)}(B)}{-\phi_{nc,r}B^r\phi_{nc}^{(c)}(B^{-1})}Z_t \end{aligned}$$

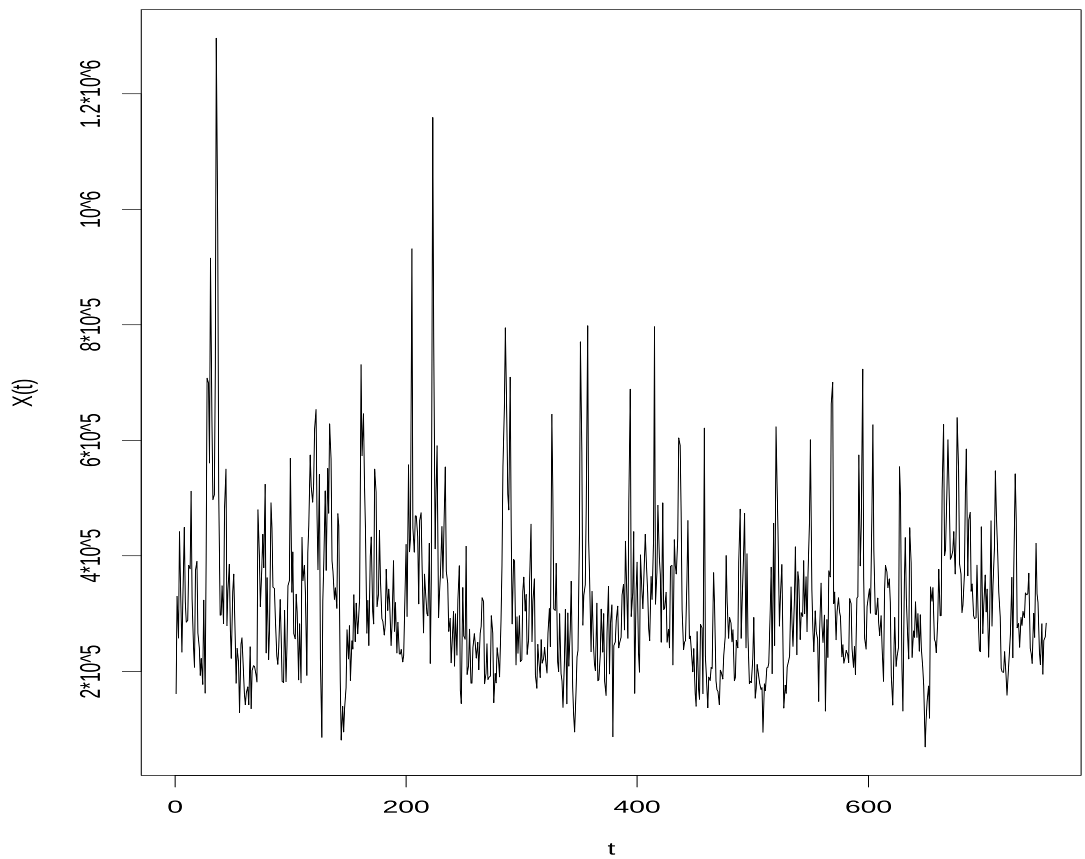
- $\{U_t\}$ is purely noncausal AP($r$), hence WN

## Two-Step Fitting Procedure

1. Fit causal AR($\hat{s}$):
   - use standard order selection and estimation
   - obtain residuals $\{\hat{U}_t\}$
2. Fit purely noncausal All-Pass to residuals
   - select order $\hat{r}$
   - look for iid noise (not merely white)
   - obtain purely noncausal AR($\hat{r}$), $\phi_{nc}(\cdot)$
   - find inverse roots of AR($\hat{r}$)
   - cancel corresponding roots in causal AR($\hat{s}$) to obtain causal AR($\hat{q}$), $\phi_c(\cdot)$

## Example: Microsoft Trading Volume

- Volumes of MSFT stock traded over 754 transaction days from 06/03/96 to 05/27/99



## Two-Step Fitting: Microsoft Trading Volume

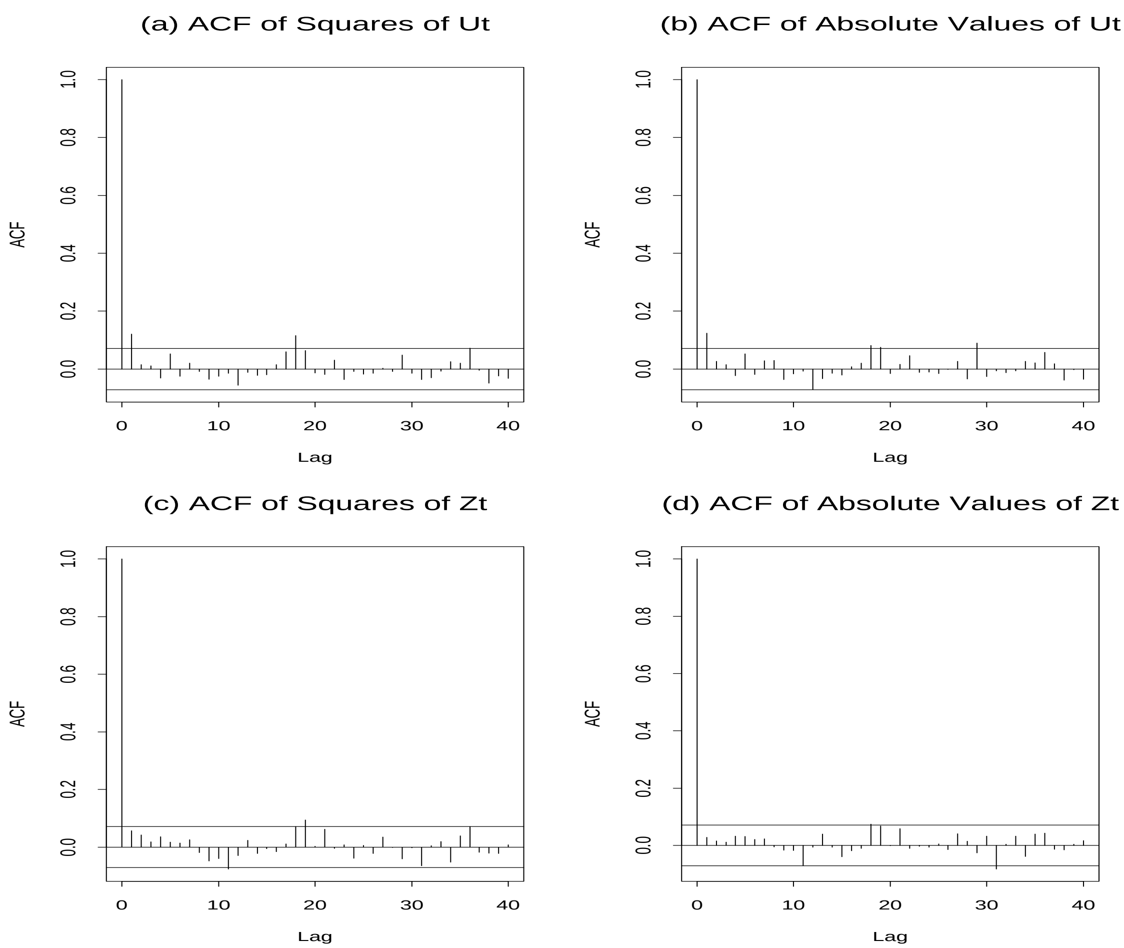
1. Fit causal AR($s$) to log volume
   - $s = 1$; $\hat{\phi}_c(B)\hat{\phi}_{nc}^{(c)}(B) = 1 - 0.5834B$
   - residuals $\{\hat{U}_t\}$ do not appear iid
2. Fit purely noncausal All-Pass to residuals
   - to get iid noise, choose $r = 1$
   - obtain purely noncausal AR(1): $\tilde{\phi}_{nc}(B) = 1 - 1.7522B$
   - find inverse root and cancel corresponding root in causal model:

$$\frac{\hat{\phi}_c(B)\hat{\phi}_{nc}^{(c)}(B)}{\tilde{\phi}_{nc}^{(c)}(B)} = \frac{1 - 0.5834B}{1 - (1.7522)^{-1}B} = \frac{1 - 0.5834B}{1 - 0.5707B} \simeq 1$$

   - in this simple case, fitted AR model is purely noncausal

## Summary: Microsoft Trading Volume

- Two-step fit of noncausal AR(1): $1 - 1.7522B$
  - causal AR(1); residuals not iid
  - purely noncausal AP(1); residuals iid
- Direct fit of noncausal AR(1): $1 - 1.7141B$
- For ATML and MCHP, causal AR models fit



## Summary

- All-pass models and their properties
  - linear time series with “nonlinear” behavior
- Estimation
  - likelihood approximation
  - MLE and LAD
  - central limit theorems
  - order selection
- Empirical results
  - simulation study
  - AP(6) for NZ/USA exchange rates
- Noncausal autoregressive processes
  - two-step estimation procedure using all-pass
  - noncausal AR(1) for Microsoft trading volume

## Further Work

- Least absolute deviations
  - further simulations
  - order selection
  - heavy-tailed case?
- Maximum likelihood
  - Gaussian mixtures
  - simulation studies
  - applications
- Noncausal autoregressive modeling
  - initial estimates from two-step all-pass procedure
  - adaptive procedures
  - comparisons with cumulant methods

# Heavy Tails

- Pareto-like tails:

$$x^\alpha \Pr[|Z_t| > x] \to \text{ constant as } x \to \infty$$

- Diverging moments: $\mathrm{E}|X_t|^k = \infty$ for $k \geq \alpha$
- "ACF" of $|X_t|^\delta$:

$$\hat{\rho}_\delta(h) \xrightarrow{P} \rho_\delta(h) = \frac{\sum_{j=0}^{\infty} |\psi_j|^\delta |\psi_{j+h}|^\delta}{\sum_{j=0}^{\infty} |\psi_j|^{2\delta}}$$