Nonlinear Time Series Modeling

Richard A. Davis
Colorado State University

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

MaPhySto Workshop
Copenhagen
September 27 — 30, 2004
Part I: Introduction to Linear and Nonlinear Time Series

1. Introduction

2. Examples

3. Linear processes
   3.1 Preliminaries
   3.2 Wold Decomposition
   3.3 Reversibility
   3.4 Identifiability
   3.5 Linear tests
   3.6 Prediction

4. Allpass models
   4.1 Application of allpass
      • Noninvertible MA model fitting
      • Microsoft
      • Muddy Creek
      • Seisomogram deconvolution
   4.2 Estimation
Part II: Time Series Models in Finance

1. Classification of white noise
2. Examples
3. “Stylized facts” concerning financial time series
4. ARCH and GARCH models
5. Forecasting with GARCH
6. IGARCH
7. Stochastic volatility models
8. Regular variation and application to financial TS
   8.1 univariate case
   8.2 multivariate case
   8.3 applications of multivariate regular variation
   8.4 application of multivariate RV equivalence
   8.5 examples
   8.6 Extremes for GARCH and SV models
   8.7 Summary of results for ACF of GARCH & SV models
Part III: Nonlinear and NonGaussian State-Space Models

1. Introduction
   1.1 Motivation examples
   1.2 Linear state-space models
   1.3 Generalized state-space models

2. Observation-driven models
   2.1 GLARMA models for TS of counts
   2.2 GLARMA extensions
   3.3 Other

3. Parameter-driven models
   3.1 Estimation
   3.2 Simulation and Application
   3.3 How good is the posterior approximation
Part IV: Structural Break Detection in Time Series

1. Piecewise AR models
2. Minimum description length (MDL)
3. Genetic algorithm (GA)
4. Simulation examples
5. Applications (EEG and speech examples)
6. Application to nonlinear models
References:


• Brockwell and Davis (2001). *Introduction to Time Series and Forecasting.*


• Tong (2000). *Nonlinear Time Series Models; a dynamical systems approach.*
1. Introduction

Why nonlinear time series models?

What are the limitations of linear time series models?

What key features in data cannot be captured by linear time series models?

What diagnostic tools (visual or statistical) suggest incompatibility of a linear model with the data?
Example: $Z_1, \ldots, Z_n \sim \text{IID}(0, \sigma^2)$

Sample autocorrelation function (ACF):

$$\hat{\rho}_Z(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad \text{where} \quad \hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-|h|} (Z_t - \bar{Z})(Z_{t+|h|} - \bar{Z})$$

is the sample autocovariance function (ACVF).
Theorem. If \( \{Z_t\} \sim \text{IID}(0, \sigma^2) \), then
\[
(\hat{\rho}_Z(1), \ldots, \hat{\rho}_Z(h))' \text{ is approximately IID } \text{N}(0, 1/n).
\]

Proof: (see problem 6.24 TSTM)
Cor. If \( \{Z_i\} \sim \text{IID}(0, \sigma^2) \) and \( E|Z_1|^4 < \infty \), then

\[
(\hat{\rho}_{Z^2}(1), \ldots, \hat{\rho}_{Z^2}(h))' \quad \text{is approximately IID } N(0, 1/n).
\]
What if $E|Z_1|^2 = \infty$? For example, suppose $\{Z_t\} \sim$ IID Cauchy.

Result (see TSTM 13.3): If $\{Z_t\} \sim$ IID Cauchy, then

$$\frac{n}{\ln n} \hat{\rho}_Z(h) \Rightarrow \frac{S_1}{S_5},$$

$S_1$ and $S_5$ are independent stable random variables.
How about the ACF of the squares?

Result: If \( \{Z_i\} \sim \text{IID Cauchy} \), then

\[
\left( \frac{n}{\ln n} \right)^2 \hat{\rho}_{Z^2}(h) \Rightarrow \frac{S_{1/2}}{S_{.25}},
\]

\( S_{.5} \) and \( S_{.25} \) are independent stable random variables.
Reversibility. The stationary sequence of random variables \( \{X_t\} \) is **time-reversible** if \((X_1, \ldots, X_n) \sim_d (X_n, \ldots, X_1)\).

**Result:** IID sequences \( \{Z_t\} \) *are* time-reversible.

**Application:** If plot of time series does not look time-reversible, then it *cannot* be modeled as an IID sequence. Use the “flip and compare” inspection test!
Reversibility. *Does the following series look time-reversible?*
2. Examples

Closing Price for IBM 1/2/62-11/3/00
Sample ACF IBM (a) 1962-1981, (b) 1982-2000

(a) ACF of IBM (1st half)

(b) ACF of IBM (2nd half)

**Remark:** Both halves look like white noise?
Sample ACF of abs values for IBM (a) 1961-1981, (b) 1982-2000

Remark: Series are not independent white noise?
Remark: Series are not independent white noise? Try GARCH or a stochastic volatility model.
Example: Pound-Dollar Exchange Rates

Remark: Usually marginal distribution of a linear process is continuous.
Muddy Creek- tributary to Sun River in Central Montana

Muddy Creek: surveyed every 15.24 meters, total of 5456m; 358 measurements

<table>
<thead>
<tr>
<th>Degree</th>
<th>$\Delta \text{IC}_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1455</td>
</tr>
<tr>
<td>1</td>
<td>294.3</td>
</tr>
<tr>
<td>2</td>
<td>251.3</td>
</tr>
<tr>
<td>3</td>
<td>47.1</td>
</tr>
<tr>
<td>4</td>
<td>34.0</td>
</tr>
<tr>
<td>5</td>
<td>35.5</td>
</tr>
</tbody>
</table>
**Muddy Creek: residuals from poly(d=4) fit**

Minimum AICc ARMA model:

\[
ARMA(1,1) \quad Y_t = 0.574 \, Y_{t-1} + \epsilon_t - 0.311 \, \epsilon_{t-1},
\]

\[\{\epsilon_t\} \sim WN(0,0.0564)\]

Some theory:

- LS estimates of trend parameters are asymptotically efficient.
- LS estimates are asymptotically independent of cov parameter estimates.

Noncausal ARMA(1,1) model:

\[
Y_t = 1.743 \, Y_{t-1} + \epsilon_t - 0.311 \, \epsilon_{t-1}
\]

Blue = sample

Red = model
Muddy Creek (cont)

Summary of models fitted to Muddy Creek bed elevation:

<table>
<thead>
<tr>
<th>Degree</th>
<th>$AIC_c$</th>
<th>ARMA</th>
<th>$AIC_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1455</td>
<td>(1,2)</td>
<td>59.67</td>
</tr>
<tr>
<td>1</td>
<td>294.3</td>
<td>(2,1)</td>
<td>26.98</td>
</tr>
<tr>
<td>2</td>
<td>251.3</td>
<td>(2,1)</td>
<td>26.30</td>
</tr>
<tr>
<td>3</td>
<td>47.1</td>
<td>(1,1)</td>
<td>7.12</td>
</tr>
<tr>
<td>4</td>
<td>34.0</td>
<td>(1,1)</td>
<td>2.78</td>
</tr>
<tr>
<td>5</td>
<td>35.5</td>
<td>(1,1)</td>
<td>4.68</td>
</tr>
</tbody>
</table>
Example: NEE=Net Ecosystem Exchange in Harvard Forest

- About half of the CO2 emitted by humans accumulates in the atmosphere
- Other half is absorbed by “sink” processes on land and in the oceans

\[ \text{NEE} = (\text{Rh} + \text{Ra}) - \text{GPP} \] (carbon flux)

\[ \text{GPP} = \text{Gross Primary Production (photosynthesis)} \]

\[ \text{Rh} = \text{Heterotrophic (microbial) respiration} \]

\[ \text{Ra} = \text{autotrophic (plant) respiration}. \]

The NEE data from the Harvard Forest consists of hourly measurements. We will aggregate over the day and consider daily data from Jan 1, 1992 to Dec 31, 2001.
3. Linear Processes

3.1 Preliminaries

**Def:** The stochastic process \( \{X_t, t=0, \pm 1, \pm 2, \ldots \} \) defined on a probability space is called a *discrete-time time series*.

**Def:** \( \{X_t\} \) is *stationary* or *weakly stationary* if

i. \( E|X_t|^2 < \infty \), for all \( t \).

ii. \( EX_t = m \), for all \( t \).

iii. \( \text{Cov}(X_t, X_{t+h}) = \gamma(h) \) depends on \( h \) only.

**Def:** \( \{X_t\} \) is *strictly stationary* if \( (X_1, \ldots, X_n) =_d (X_{1+h}, \ldots, X_{n+h}) \) for all \( n \geq 1 \) and \( h=0, \pm 1, \pm 2, \ldots \).

**Remarks:**

i. SS + \( (E|X_t|^2 < \infty) \) ⇒ weak stationarity

ii. WS \( \not\Rightarrow \) SS (think of an example)

iii. WS + Gaussian ⇒ SS (why?)
3.1 Preliminaries (cont)

**Def:** \( \{X_t\} \) is a **Gaussian time series** if

\[(X_m, \ldots, X_n) \text{ is multivariate normal}\]

for all integers \( m < n \), i.e., all finite dimensional distributions are normal.

**Remark:** A Gaussian time series is completely determined by the mean function and covariance functions,

\[m(t) = EX_t \text{ and } \gamma(s,t) = \text{Cov}(X_s, X_t).\]

It follows that a **Gaussian TS** is **stationary** (SS or WS) if and only if

\[m(t) = m \text{ and } \gamma(s,t) = \gamma(t-s) \text{ depends only on the time lag } t-s.\]
3.1 Preliminaries (cont)

**Def:** \( \{X_t\} \) is a *linear time series* with mean 0 if

\[
X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},
\]

where \( \{Z_t\} \sim WN(0, \sigma^2) \) and \( \sum_{j=-\infty}^{\infty} \psi_j^2 < \infty \).

**Important remark:** As a reminder WN means uncorrelated random variables and not necessarily independent noise nor independent Gaussian noise.

**Proposition:** A linear TS is stationary with

i. \( EX_t = 0 \), for all \( t \).

ii. \( \gamma(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \) and \( \rho(h) = \frac{\sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}}{\sum_{j=-\infty}^{\infty} \psi_j^2} \)

If \( \{Z_t\} \sim IID(0, \sigma^2) \), then the linear TS is strictly stationary.
Is the converse to the previous proposition true? That is, are all stationary processes linear?

**Answer:** Almost.

3.2 Wold Decomposition (TSTM Section 5.7)

**Example:** Set

\[ X_t = A \cos(\omega t) + B \sin(\omega t), \quad \omega \in (0, \pi), \]

where \( A, B \sim WN(0, \sigma^2) \). Then \( \{X_t\} \) is stationary since

- \( E X_t = 0 \),
- \( \gamma(h) = \sigma^2 \cos(\omega h) \)

**Def:** Let \( \widetilde{P}_n(\cdot) \) be the *best linear predictor operator* onto the linear span of the observations \( X_n, X_{n-1}, \ldots \).

For this example,

\[ \widetilde{P}_{n-1}(X_n) = X_n. \]

Such processes with this property are called *deterministic*.
3.2 Wold Decomposition (cont)

**The Wold Decomposition.** If \( \{X_t\} \) is a nondeterministic stationary time series with mean zero, then

\[
X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t,
\]

where

i. \( \psi_0 = 1, \Sigma \psi_j^2 < \infty \).

ii. \( \{Z_t\} \sim WN(0, \sigma^2) \)

iii. \( \text{cov}(Z_s, V_t) = 0 \) for all \( s \) and \( t \)

iv. \( \tilde{P}_t(Z_t) = Z_t \) for all \( t \).

v. \( \tilde{P}_s(V_t) = V_t \) for all \( s \) and \( t \).

vi. \( \{V_t\} \) is deterministic.

The sequences \( \{Z_t\}, \{V_t\}, \) and \( \{\psi_t\} \) are unique and can be written as

\[
Z_t = X_t - \tilde{P}_{t-1}(X_t), \quad \psi_j = E(X_t Z_{t-j}) / E(Z_t^2), \quad V_t = X_t - \sum_{j=0}^{\infty} \psi_j Z_{t-j}.
\]
3.2 Wold Decomposition (cont)

Remark. For many time series (in particular for all ARMA processes) the deterministic component $V_t$ is 0 for all $t$ and the series is then said to be purely nondeterministic.

Example. Let

$$X_t = U_t + Y,$$

where $\{U_t\} \sim WN(0, \sigma^2)$ and is independent of $Y \sim (0, \tau^2)$. Then, in this case, $Z_t = U_t$ and $V_t = Y$ (see TSTM, problem 5.24).

Remarks:

- If $\{X_t\}$ is purely nondeterministic, then $\{X_t\}$ is a linear process.
- Spectral distribution for nondeterministic processes has the form

$$F_X = F_U + F_V,$$

where $U_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ which has spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \psi_j e^{ij\lambda} \right|^2 = \frac{\sigma^2}{2\pi} \left| \psi(e^{i\lambda}) \right|^2$$
3.2 Wold Decomposition (cont)

- If $\sigma^2 = E(X_t - \hat{P}_{t-1}(X_t))^2 > 0$, then
  $$F_X = F_U + F_V,$$
  is the Lebesque decomposition of the spectral distribution function; $F_U$ is the absolutely continuous part and $F_V$ is the singular part.

**Example.** Let

$$X_t = U_t + Y,$$
where $\{U_t\} \sim WN(0, \sigma^2)$ and is independent of $Y \sim (0, \tau^2)$. Then

$$F_X(d\lambda) = \frac{\sigma^2}{2\pi}(d\lambda) + \tau^2\delta_0(d\lambda)$$

**Kolmogorov’s Formula.**

$$\sigma^2 = 2\pi \exp\left\{ (2\pi)^{-1} \int_{-\pi}^{\pi} \ln f(\lambda)d\lambda \right\},$$
where $\sigma^2 = E(X_t - \hat{P}_{t-1}(X_t))^2$.

Clearly $\sigma^2 > 0$ iff $\int_{-\pi}^{\pi} \ln f(\lambda)d\lambda > -\infty$. 
3.2 Wold Decomposition (cont)

Example (TSTM problem 5.23).

\[ X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim WN(0, \tau^2), \quad \psi_j = \left( \frac{1}{\pi} \right) \left( \frac{\sin j}{j} \right). \]

This process has a spectral density function but is deterministic!!

Example (see TSTM problem 5.20). Let

\[ X_t = \varepsilon_t - 2\varepsilon_{t-1}, \quad \{\varepsilon_t\} \sim WN(0, \tau^2), \]

and set

\[ Z_t = (1 - .5B)^{-1} X_t \]

\[ = \sum_{j=0}^{\infty} .5^j X_{t-j} = \varepsilon_t - 2\varepsilon_{t-1} + .5(\varepsilon_{t-1} - 2\varepsilon_{t-2}) + .5^2(\varepsilon_{t-2} - \varepsilon_{t-3}) + \cdots \]

\[ = \varepsilon_t - 3 \sum_{j=1}^{\infty} .5^j \varepsilon_{t-j} \]

It follows that \( \{Z_t\} \sim WN(0, \sigma^2) \) and \( X_t = Z_t - .5Z_{t-1} \) is the WD for \( \{X_t\} \).

a) If \( \{\varepsilon_t\} \sim \text{IID } N(0, \sigma^2) \), is \( \{Z_t\} \) IID? Answer?

b) If \( \{\varepsilon_t\} \sim \text{IID } (0, \sigma^2) \), is \( \{Z_t\} \) IID? Answer?
3.2 Wold Decomposition (cont)

Remark: In this last example, the process \( \{Z_t\} \) is called an allpass model of order 1. More on this type of process later.

3.3 Reversibility

Recall that the stationary time series \( \{X_t\} \) is time-reversible if
\[
(X_1, \ldots, X_n) =_d (X_n, \ldots, X_1)
\]
for all \( n \).
3.3 Reversibility

The stationary time series \{X_t\} is time-reversible if \((X_1, \ldots , X_n) =_d (X_n, \ldots , X_1)\) for all \(n\).

**Theorem (Breidt & Davis 1991).** Consider the linear time series \{X_t\}

\[ X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim IID, \]

where \(\psi(z) \neq \pm z^r \psi(z^{-1})\) for any integer \(r\). Assume either

(a) \(Z_0\) has mean 0 and finite variance and \{\(X_t\)\} has a spectral density positive almost everywhere.

or

(b) \(1/\psi(z) = \pi(z) = \sum_j \pi_j z^j\), the series converging absolutely in some annulus \(D\) containing the unit circle and

\[ \pi(B)X_t = \sum_j \pi_j X_{t-j} = Z_t. \]

Then \{\(X_t\)\} is time-reversible if and only if \(Z_0\) is Gaussian.
3.3 Reversibility (cont)

**Remark:** The condition $\psi(z) \neq \pm z^r \psi(z^{-1})$ on the filter precludes the filter from being symmetric about one of the coefficients. In this case, the time series would be time-reversible for non-Gaussian noise. For example, consider the series

$$X_t = Z_t - 0.5Z_{t-1} + Z_{t-2}, \quad \{Z_t\} \sim IID$$

Here $\psi(z) = 1 - 0.5z + z^2 = z^2 (1 - 0.5z^{-1} + z^2) = z^2 \psi(z^{-1})$ and the series is time-reversible.

**Proof of Theorem:** Clearly any stationary Gaussian time series is time-reversible (why?). So suppose $Z_0$ is non-Gaussian and assume (a). If $\{X_t\}$ time-reversible, then

$$Z_t = \frac{1}{\psi(B)} X_t = d \frac{1}{\psi(B^{-1})} X_t = \frac{\psi(B)}{\psi(B^{-1})} Z_t = \sum_{j=-\infty}^{\infty} a_j Z_{t-j}.$$
3.3 Reversibility (cont)

The first equality takes a bit of argument and relies on the spectral representation of \( \{X_t\} \) given by

\[
X_t = \int_{(-\pi, \pi]} e^{i\lambda t} dZ(\lambda),
\]

where \( Z(\lambda) \) is a process of orthogonal increments (see TSTM, Chapter 4). It follows, by the assumptions on the spectral density of \( \{X_t\} \) that

\[
\frac{1}{\psi(B^\pm)} X_t = \int_{(-\pi, \pi]} \frac{1}{\psi(e^{\pm i\lambda})} e^{i\lambda t} dZ(\lambda),
\]

is well defined. So

\[
Z_t = d \quad \frac{\psi(B)}{\psi(B^{-1})} Z_t = \sum_{j=-\infty}^{\infty} a_j Z_{t-j}.
\]

and, by the assumption on \( \psi(z) \), the rhs is a non-trivial sum. Note that

\[
\sum_{j=-\infty}^{\infty} a_j^2 = 1 \quad \text{Why?}
\]

The above relation is a characterization of a Gaussian distribution (see Kagan, Linnik, and Rao (1973).) \( \square \)
3.3 Reversibility (cont)

Example: Recall for the example

\[ X_t = \varepsilon_t - 2\varepsilon_{t-1}, \quad \{\varepsilon_t\} \sim IID(0, \tau^2), \]

and non-normal, the Wold decomposition is given by

\[ X_t = Z_t - .5Z_{t-1}, \]

where

\[ Z_t = \varepsilon_t - 3 \sum_{j=1}^{\infty} .5^j \varepsilon_{t-j}. \]

By previous result, \{Z_t\} cannot be time-reversible and hence is not IID.

Remark: This theorem can be used to show identifiability of the parameters and noise sequence for an ARMA process.
3.4 Identifiability

Motivating example: The *invertible* MA(1) process

\[ X_t = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim IID(0, \sigma^2), \quad |\theta| < 1, \]

has a *non-invertible* MA(1) representation,

\[ X_t = \varepsilon_t + \theta^{-1} \varepsilon_{t-1}, \quad \{\varepsilon_t\} \sim WN(0, \theta^2 \sigma^2), \quad |\theta| < 1. \]

Question: Can the \{\varepsilon_t\} also be IID?

Answer: Only if the \{Z_t\} are Gaussian.

If the \{Z_t\} are Gaussian, then there is an *identifiability* problem,

\[ (\theta, \sigma^2) \leftrightarrow (\theta^{-1}, \theta^2 \sigma^2), \quad |\theta| < 1, \]

give the same model.
3.4 Identifiability (cont)

For ARMA processes \( \{X_t\} \) satisfying the recursions,

\[
X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \quad \{Z_t\} \sim \text{IID}(0, \sigma^2), \\
\phi(B) X_t = \theta(B) Z_t
\]

casuality and invertibility are typically assumed, i.e.,

\[
\phi(z) \neq 0 \quad \text{and} \quad \theta(z) \neq 0 \quad \text{for} \quad |z| \leq 1.
\]

By flipping roots of the AR and MA polynomials from outside the unit circle to inside the unit circle, there are approximately \( 2^{p+q} \) equivalent ARMA representations of \( X_t \) driven with noise that is white (not IID). For each of these equivalent representations, the noise is only IID in the Gaussian case.

**Bottom line:** For nonGaussian ARMA, there is a distinction between causal and noncausal; and invertible and non-invertible models.
3.4 Identifiability (cont)

**Theorem (Cheng 1992):** Suppose the linear time series

\[ X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim \text{IID}(0, \sigma^2), \sum_j \psi_j^2 < \infty, \]

has a positive spectral density a.e. and can also be represented as

\[ X_t = \sum_{j=-\infty}^{\infty} \eta_j Y_{t-j}, \quad \{Y_t\} \sim \text{IID}(0, \tau^2), \sum_j \eta_j^2 < \infty. \]

Then if \( \{X_t\} \) is nonGaussian, it follows that

\[ Y_t = c Z_{t-t_0}, \quad \eta_j = \frac{1}{c} \psi_{j+t_0}, \]

for some positive constant \( c \).

**Proof of Theorem:** As in the proof of the reversibility result, we can write

\[ Z_t = \frac{1}{\psi(B)} X_t = \frac{\eta(B)}{\psi(B)} Y_t = \sum_{j=-\infty}^{\infty} a_j Y_{t-j} \quad \text{and} \quad Y_t = \sum_{j=-\infty}^{\infty} b_j Z_{t-j} \]
3.4 Identifiability (cont)

Now let \( \{Y(s,t)\} \sim \text{IID}, \ Y(s,t) \equiv_d Y_1 \) and set

\[
U_t = \sum_{s=-\infty}^{\infty} a_s Y(s,t).
\]

Clearly, \( \{U_t\} \) is IID with same distribution as \( Z_1 \). Consequently,

\[
Y_1 = \sum_{t=-\infty}^{\infty} b_t U_t = \sum_{t=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} b_t a_s Y(s,t).
\]

Since

\[
\sum_{t=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} b_t^2 a_s^2 = 1,
\]

Which by applying Theorems 5.6.1 and 3.3.1 in Kagan, Linnik, and Rao (1973), the sum above is trivial, i.e., there exists integers \( m \) and \( n \) such that \( a_m \) and \( b_n \) are the only two nonzero coefficients. It follows that

\[
Y_t = b_n Z_{t-n}, \quad \eta_j = \frac{1}{b_n} \psi_{j+n}.
\]
3.5 Linear Tests

**Cumulants and Polyspectra.** We cannot base tests for linearity on second moments. A direct approach is to consider moments of higher order and corresponding generalizations of spectral analysis.

Suppose that \( \{X_t\} \) satisfies \( \sup_t E|X_t|^k < \infty \) for some \( k \geq 3 \) and

\[
E(X_{t_0}X_{t_1}\cdots X_{t_j}) = E(X_{t_0+h}X_{t_1+h}\cdots X_{t_j+h})
\]

for all \( t_0, t_1, \ldots, t_j, h=0, \) and \( j = 0, \ldots, k-1. \)

**\( k \)th order cumulant.** Coefficient, \( C_k(r_1, \ldots, r_{k-1}) \), of \( i^{k}z_1z_2\cdots z_k \) in the Taylor series expansion about \((0,0,\ldots,0)\) of

\[
\chi(z_1,\ldots,z_k) = \ln E \exp(iz_1X_t + iz_2X_{t+r_1} + \cdots + iz_kX_{t+r_{k-1}})
\]
3.5 Linear Tests (cont)

3rd order cumulant.

\[ C_3(r, s) = E((X_t - \mu)(X_{t+r} - \mu)(X_{t+s} - \mu)) \]

If

\[ \sum_r \sum_s |C_3(r, s)| < \infty \]

then we define the **bispectral density** or **(3rd – order polyspectral density)**

To be the Fourier transform,

\[ f_3(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \sum_{r=\infty}^{\infty} \sum_{s=\infty}^{\infty} C_3(r, s)e^{-i\omega_1r - i\omega_2s}, \]

\[-\pi \leq \omega_1, \omega_2 \leq \pi.\]
3.5 Linear Tests (cont)

$k^{th}$-order polyspectral density.

Provided

$$\sum_{r_1} \sum_{r_2} \cdots \sum_{r_{k-1}} |C_k(r_1, \ldots, r_{k-1})| < \infty,$$

$$f_k (\omega_1, \ldots, \omega_{k-1}) := \frac{1}{(2\pi)^{k-1}} \sum_{r_1 = -\infty}^{\infty} \sum_{r_2 = -\infty}^{\infty} \cdots \sum_{r_{k-1} = -\infty}^{\infty} C_k (r_1, \ldots, r_{k-1}) e^{-ir_1\omega_1 \cdots ir_{k-1}\omega_{k-1}},$$

$$-\pi \leq \omega_1, \ldots, \omega_{k-1} \leq \pi. \quad \text{(See Rosenblatt (1985) Stationary Sequences and Random Fields for more details.)}$$
3.5 Linear Tests (cont)

Applied to a linear process. If \( \{X_t\} \) has the Wold decomposition

\[
X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim IID(0, \sigma^2),
\]

with \( E|Z_t|^3 < \infty, EZ_t^3 = \eta \), and \( \Sigma_j |\psi_j| < \infty \), then

\[
C_3(r, s) = \eta \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+r} \psi_{j+s}
\]

where \( \psi_j := 0 \) for \( j < 0 \). Hence

\[
f_3(\omega_1, \omega_2) = \frac{\eta}{4\pi^2} \psi(e^{i\omega_1+i\omega_2})\psi(e^{-i\omega_1})\psi(e^{-i\omega_2}).
\]
3.5 Linear Tests (cont)

The spectral density of \( \{X_t\} \) is

\[ f(\omega) = \frac{\sigma^2}{2\pi} |\psi(e^{i\omega})|^2. \]

Hence, defining

\[ \phi(\omega_1, \omega_2) = \frac{|f_3(\omega_1, \omega_2)|^2}{f(\omega_1) f(\omega_2) f(\omega_1 + \omega_2)}, \]

we find that

\[ \phi(\omega_1, \omega_2) = \frac{\eta^2}{2\pi \sigma^6}. \]

Testing for constancy of \( \phi(\cdot) \) thus provides a test for linearity of \( \{X_t\} \) (see Subba Rao and Gabr (1980)).
3.5 Linear Tests (cont)

**Gaussian linear process.** If \( \{X_t\} \) is Gaussian, then \( EZ^3=0 \), and the third order cumulant is zero (why?). In fact \( C_k \equiv 0 \) for all \( k > 2 \).

It follows that \( f_3(\omega_1, \omega_2) \equiv 0 \) for all \( \omega_1, \omega_2 \in [0, \pi] \). A test for linear Gaussianity can therefore be obtained by estimating \( f_3(\omega_1, \omega_2) \) and testing the hypothesis that \( f_3 \equiv 0 \) (see Subba Rao and Gabr (1980)).
3.6 Prediction

Suppose \( \{X_t\} \) is a purely nondeterministic process with WD given by

\[
X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim WN(0, \sigma^2).
\]

Then

\[
Z_t = X_t - \tilde{P}_{t-1}(X_t)
\]

so that

\[
\tilde{P}_{t-1} X_t = \sum_{j=1}^{\infty} \psi_j Z_{t-j}.
\]

**Question.** When does the best linear predictor equal the best predictor?

That is, when does

\[
\tilde{P}_{t-1} X_t = E( X_t \mid X_{t-1}, X_{t-2} \ldots )?
\]
3.6 Prediction (cont)

\[ \tilde{P}_{t-1}X_t = E( X_t | X_{t-1}, X_{t-2}, \ldots )? \]

**Answer.** Need

\[ Z_t = X_t - \tilde{P}_{t-1}X_t \perp \sigma(X_{t-1}, X_{t-2}, \ldots) \]

or, equivalently,

\[ E(Z_t | X_{t-1}, X_{t-2}, \ldots) = 0. \]

That is,

BLP = BP

if and only if \( \{Z_t\} \) is a *Martingale-difference sequence*.

**Def.** \( \{Z_t\} \) is a *Martingale-difference sequence wrt a filtration \( F_t \) (an increasing sequence of sigma fields) if \( E|Z_t| < \infty \) for all \( t \) and

a) \( Z_t \) is \( F_t \) measurable

b) \( E(Z_t | F_{t-1})=0 \) a.s.
3.6 Prediction (cont)

Remarks.
1) An IID sequence with mean zero is a MG difference sequence.
2) A purely nondeterministic Gaussian process is a Gaussian linear process. This follows by the Wold decomposition and the fact that the resulting \( \{Z_t\} \) sequence must be IID \( N(0, \sigma^2) \).

Example (Whittle): Consider the noncausal AR(1) process given by

\[ X_t = 2X_{t-1} + Z_t, \]

where \( \{Z_t\} \sim \text{IID } P(Z_t = -1) = P(Z_t = 0) = .5 \). Iterating backwards in time, we find that

\[
X_{t-1} = .5X_t - .5Z_t \\
= .5^2 X_{t+1} - .5^2 Z_{t+1} - .5Z_t \\
\vdots \\
= .5(-Z_t - .5Z_{t+1} - .5^2 Z_{t+2} - \cdots).
\]
3.6 Prediction (cont)

\[ X_t = .5(-Z_{t+1} - .5Z_{t+2} - .5^2 Z_{t+3} - \cdots) \]

\[ = \frac{Z^*_t + Z^*_t}{2} + \frac{Z^*_t}{2^2} + \frac{Z^*_t}{2^3} + \cdots, \quad Z^*_1 = -Z_{t+1} \]

is a binary expansion of a uniform (0,1) random variable. Notice that from \( X_t \), we can find \( X_{t+1} \), by lopping off the first term in the binary expansion. This operation is exactly,

\[ X_{t+1} = 2X_t \mod 1 \]

\[ = \begin{cases} 
2X_t, & \text{if } X_t < .5, \\
2X_t - 1, & \text{if } X_t > .5.
\end{cases} \]

**Properties:**

1. \( EX_t = \frac{1}{2} \).
2. \( \rho_X(h) = (0.5)^{|h|} \).
3. \( P(X_t|1, X_s, s < t) = \frac{1}{4} + \frac{1}{2}X_{t-1} \).
4. \( E(X_t|X_s, s < t) = 2X_{t-1}(\mod 1) = X_t \).
5. \( X_t - \frac{1}{2} = \frac{1}{2}(X_{t-1} - \frac{1}{2}) + \epsilon_t, \quad \{\epsilon_t\} \sim \text{WN}(0, \sigma^2). \)
4. Allpass models

Realization from an all-pass model of order 2
(t3 noise )
4. Allpass models (cont)

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

**Define MA polynomial:**

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

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

**Model for data** \( \{X_t\} \):

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

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

**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 assuming known density
  • quasi-likelihood

❖ Other
  • LAD- least absolute deviation
  • R-estimation (minimum dispersion)
4.1 Application of Allpass models

Noninvertible MA models with heavy tailed noise

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

a. \( \{Z_t\} \sim \text{IID nonnormal} \)

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):

\[
(1 - 0.628B + 0.229B^2 + 1.31B^3 - 0.202B^4)U_t \\
= (1 - 0.649B + 1.135B^2 + 3.116B^3 - 4.960B^4)Z_t. \quad (\hat{\nu} = 6.26)
\]

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

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

\[ X_t = (1 + 1.34B + 1.374B^2 + 2.54B^3 + 4.96B^4) Z_t, \quad \{Z_t\} \sim \text{IID } \text{t}(6.3) \]
(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.
Muddy Creek: residuals from poly(d=4) fit

Minimum AICc ARMA model:
ARMA(1,1)

\[ Y_t = 0.574 Y_{t-1} + \varepsilon_t - 0.311 \varepsilon_{t-1}, \{\varepsilon_t\} \sim WN(0, 0.0564) \]
Causal ARMA(1,1) model

\[ Y_t = 0.574 Y_{t-1} + \varepsilon_t - 0.311 \varepsilon_{t-1}, \]

\( \{\varepsilon_t\} \sim WN(0, 0.0564) \)

Noncausal ARMA(1,1) model:

\[ Y_t = 1.743 Y_{t-1} + \varepsilon_t - 0.311 \varepsilon_{t-1} \]
Example: Seismogram Deconvolution

Simulated water gun seismogram

\[ X_t = \sum_k \beta_k Z_{t-k} \]

- \( \{\beta_k\} \) = wavelet sequence (Lii and Rosenblatt, 1988)
- \( \{Z_t\} \) IID reflectivity sequence
Water Gun Seismogram Fit

Step 1: AICC suggests ARMA (12,13) fit

- fit invertible ARMA(12,13) via Gaussian MLE
- residuals \( \{ \hat{W}_t \} \) not IID

Step 2: fit all-pass to \( \{ \hat{W}_t \} \) residuals

- order selected is \( r = 2 \).
- residuals \( \{ \hat{Z}_t \} \) appear IID

Step 3: Conclude that \( \{ X_t \} \) follows a non-invertible ARMA
ACF of $W_t^2$

ACF of $Z_t^2$
Water Gun Seismogram Fit (cont)

Recorded water gun wavelet and its estimate

- True
- Estimate
- Estimate (invertible)
Water Gun Seismogram Fit (cont)

Simulated reflectivity sequence and its estimates

![Simulated Reflectivity Sequence](image1)

![Estimated Reflectivity Sequence](image2)
4.2 Estimation for Allpass Models: 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,

\[
z_{t-p}(\phi) = \begin{cases} 0, & \text{if } t = n + p, \ldots, n+1, \\ \phi_1z_{t-p+1}(\phi) + \cdots + \phi_pz_t(\phi) - \phi(B)X_t, & \text{if } t = n, \ldots, p+1. \end{cases}
\]

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 $z = (X_{1-p},...,X_0,z_{1-p}(\phi),...,z_0(\phi),z_1(\phi),...,z_{n-p+1}(\phi),...,z_n(\phi))'$

Joint density of $z$:

$$h(z) = h_1(X_{1-p},...,X_0,z_{1-p}(\phi),...,z_0(\phi))$$

$$\cdot \left( \prod_{t=1}^{n-p} f_\sigma(\phi_q z_t(\phi)) \mid \phi_q \right) h_2(z_{n-p+1}(\phi),...,z_n(\phi)),$$

and hence the joint density of the data can be approximated by

$$h(x) = \left( \prod_{t=1}^{n-p} f_\sigma(\phi_q z_t(\phi)) \mid \phi_q \right)$$

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

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

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 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

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

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

\[ \hat{I} = \sigma_0^{-2} \int (f'(z;\beta_0))^2 / f(z;\beta_0) dz \]
\[ \hat{K} = \alpha_{0,p+1}^{-2} \left\{ \int z^2 (f'(z;\beta_0))^2 / f(z;\beta_0) dz - 1 \right\} \]
\[ L = -\alpha_{0,p+1}^{-1} \int z f''(z;\beta_0) \frac{\partial f(z;\beta_0)}{\partial \beta_0} f(z;\beta_0) d\beta_0 \]
\[ I_f(\beta_0) = \int \frac{1}{f(z;\beta_0)} \frac{\partial f(z;\beta_0)}{\partial \beta_0} \frac{\partial f^T(z;\beta_0)}{\partial \beta_0} dz \]  

(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

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

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

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

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

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

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

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

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

Student’s \( t_3: \)

LAD: \(0.7337\)

MLE: \(0.5\)

ARE: \(0.7337/0.5 = 1.4674\)
R-Estimation:

Minimize the objective function

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

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

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

Remarks:

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

Assume \( \{Z_i\} \) iid with density function \( f \) (distr \( F \))

- mean 0, variance \( \sigma^2 \)

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

Results

Set

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

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

\[ \phi(x) = x - 1/2 \] is called the Wilcoxon weight function

By formally choosing

\[ \phi(x) = \begin{cases} 
-1, & 0 < x < 1/2, \\
1, & 1/2 < x < 1.
\end{cases} \]

we obtain

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

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

The R-estimation objective function is smoother than the LAD-objective function and hence easier to minimize.
Objective Functions

R-estimation
LAD
Summary of asymptotics

Maximum likelihood:

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

R-estimation

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

Least absolute deviations:

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

\[ R: \frac{\sigma^2 \tilde{\mathcal{J}} - \tilde{K}^2}{2(\sigma^2 \tilde{L} - \tilde{K})^2} = \frac{5}{6} \]

(\text{using } \varphi(x) = x-1/2, \text{ Wilcoxon})

LAD=MLE: 1/2

Students \( t_v \):

<table>
<thead>
<tr>
<th>( v )</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>.733</td>
<td>.520</td>
<td>.500</td>
<td>1.411</td>
<td>.962</td>
</tr>
<tr>
<td>6</td>
<td>6.22</td>
<td>3.01</td>
<td>3.00</td>
<td>2.068</td>
<td>.997</td>
</tr>
<tr>
<td>9</td>
<td>16.8</td>
<td>7.15</td>
<td>7.00</td>
<td>2.354</td>
<td>.980</td>
</tr>
<tr>
<td>12</td>
<td>32.6</td>
<td>13.0</td>
<td>12.5</td>
<td>2.510</td>
<td>.964</td>
</tr>
<tr>
<td>15</td>
<td>53.4</td>
<td>20.5</td>
<td>19.5</td>
<td>2.607</td>
<td>.952</td>
</tr>
<tr>
<td>20</td>
<td>99.6</td>
<td>36.8</td>
<td>34.5</td>
<td>2.707</td>
<td>.937</td>
</tr>
<tr>
<td>30</td>
<td>234</td>
<td>83.6</td>
<td>77.0</td>
<td>2.810</td>
<td>.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

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

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

- Then \( S_n(u) \rightarrow S(u) \) 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,

\[
\arg\min S_n(u) = n^{1/2} (\hat{\phi}_R - \phi_0) \\
\rightarrow \arg\min D \arg\min S(u) \\
= -\frac{|\phi_{0r}|}{2(\sigma^2 \tilde{L} - \tilde{K})} \sigma^2 \Gamma_p^{-1} N \sim N(0, \frac{\sigma^2 \tilde{J} - \tilde{K}^2}{2(\sigma^2 \tilde{L} - \tilde{K})^2 |\phi_{0r}|^2} \sigma^2 \Gamma_p^{-1})
\]
Main ideas (R-estimation)

- Define

\[ \tilde{S}_n(u) = \sum_{t=1}^{n-p} \varphi(F_z(z_t))z_t(\phi_0 + n^{-1/2}u) - \sum_{t=1}^{n-p} \varphi(F_z(z_t))z_t(\phi_0), \]

where \( F_z \) is the df of \( z_t \).

- Using a Taylor series, we have

\[
\tilde{S}_n(u) \sim n^{-1/2}u' \sum_{t=1}^{n-p} \left( \varphi(F_z(z_t)) \frac{\partial z_t(\phi_0)}{\partial \phi} \right) + 2^{-1} n^{-1} u' \sum_{t=1}^{n-p} \left( \varphi(F_z(z_t)) \frac{\partial^2 z_t(\phi_0)}{\partial \phi \partial \phi'} \right) u \rightarrow u'N - u'\tilde{K} | \phi_0 |^{-1} \sigma^{-2} \Gamma_p u
\]

- Also,

\[
S_n(u) - \tilde{S}_n(u) = u'\sigma^2 \tilde{L} \sigma^{-2} \Gamma_p u / | \phi_0 | + o_p(1).
\]

- Hence

\[
S_n(u) \rightarrow_D | \phi_0 |^{-1} (\sigma^2 \tilde{L} - \tilde{K}) u' \sigma^{-2} \Gamma_p u + u'N, \quad N \sim N(0, 2(\sigma^2 \tilde{J} - \tilde{K}^2) | \phi_0 |^{-2} \sigma^{-2} \Gamma_p).
\]
**Order Selection:**

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

$$n^{1/2} \hat{\phi}_{p,LAD} \rightarrow N(0, \frac{\text{Var}(|Z_1|)}{2(2\sigma^2 f_\sigma(0) - E|Z_1|)^2})$$

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,

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

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.

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

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

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

\[ AIC(p) := -2L_x(\hat{\phi}, \hat{\kappa}) + \frac{\text{Var}(|Z_1|)}{(2\sigma^2 f_\sigma(0) - E|Z_1|)^2} \left( \frac{2\sigma^2 f_\sigma(0)}{E|Z_1|} - 1 \right)p \]

• Penalty term for Laplace case:

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

• Penalty term can be estimated from the data.
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 $\phi^{(k)}_{11}, \phi^{(k)}_{22}, \ldots, \phi^{(k)}_{pp}$ iid uniform $(-1,1)$.
2. For $j=2, \ldots, p$, compute

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

- Select top 10 based on minimum function evaluation.
- Run Hooke and Jeeves with each of the 10 starting values and choose best optimized value.
### Asymptotic Empirical

<table>
<thead>
<tr>
<th>N</th>
<th>mean</th>
<th>std dev</th>
<th>mean</th>
<th>std dev</th>
<th>%coverage</th>
<th>rel eff*</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>$\phi_1 = .5$</td>
<td><strong>.0332</strong></td>
<td>.4979</td>
<td><strong>.0397</strong></td>
<td>94.2</td>
<td>11.8</td>
</tr>
<tr>
<td>5000</td>
<td>$\phi_1 = .5$</td>
<td><strong>.0105</strong></td>
<td>.4998</td>
<td><strong>.0109</strong></td>
<td>95.4</td>
<td>9.3</td>
</tr>
</tbody>
</table>

### Asymptotic Empirical

<table>
<thead>
<tr>
<th>N</th>
<th>mean</th>
<th>std dev</th>
<th>mean</th>
<th>std dev</th>
<th>%coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>$\phi_1 = .3$</td>
<td><strong>.0351</strong></td>
<td>.2990</td>
<td><strong>.0456</strong></td>
<td>92.5</td>
</tr>
<tr>
<td>5000</td>
<td>$\phi_1 = .3$</td>
<td><strong>.0111</strong></td>
<td>.3003</td>
<td><strong>.0118</strong></td>
<td>95.5</td>
</tr>
<tr>
<td>500</td>
<td>$\phi_2 = .4$</td>
<td><strong>.0351</strong></td>
<td>.3965</td>
<td><strong>.0447</strong></td>
<td>92.1</td>
</tr>
<tr>
<td>5000</td>
<td>$\phi_2 = .4$</td>
<td><strong>.0111</strong></td>
<td>.3990</td>
<td><strong>.0117</strong></td>
<td>94.7</td>
</tr>
</tbody>
</table>

*Efficiency relative to maximum absolute residual kurtosis:

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

$$v_2 = \frac{1}{n-p} \sum_{t=1}^{n-p} (z_t(\phi) - z(\phi))^2$$
**R-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>mean</th>
<th>std dev</th>
<th>mean</th>
<th>std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>(\phi_1=.5)</td>
<td>.4978</td>
<td>.0315</td>
<td>.4979</td>
</tr>
<tr>
<td>5000</td>
<td>(\phi_1=.5)</td>
<td>.4997</td>
<td>.0094</td>
<td>.4998</td>
</tr>
<tr>
<td>500</td>
<td>(\phi_1=.3)</td>
<td>.2988</td>
<td>.0374</td>
<td>.2990</td>
</tr>
<tr>
<td></td>
<td>(\phi_2=.4)</td>
<td>.3957</td>
<td>.0360</td>
<td>.3965</td>
</tr>
<tr>
<td>5000</td>
<td>(\phi_1=.3)</td>
<td>.3007</td>
<td>.0101</td>
<td>.3003</td>
</tr>
<tr>
<td></td>
<td>(\phi_2=.4)</td>
<td>.3993</td>
<td>.0104</td>
<td>.3990</td>
</tr>
</tbody>
</table>