BOX-JENKINS

JEN-WEN LIN PhD

Date: June 21, 2014

Important concepts in last class

- Time series: models and data
- Two types of stationary time series — strictly/weakly
- Sample autocorrelation functions
- Classical decomposition of time series

Topics of today

Autoregressive and moving average (ARMA) model

focuses & their relationship with stationarity.

- Moving average processes
- Autoregressive processes and stationarity
- Casual processes, invertible processes, and duality of ARMA models
- Casual and invertible ARMA processes

Three stages of the Box-Jenkins approach

Model identification

Model adequacy

Model selection

Irregular components and ARMA models

According to “classical decomposition”, a time series may be decomposed into trend, seasonal and irregular components.

“noise”

The first two components can be filtered out using regression or smoothing techniques.

In this class, your will learn how to model the irregular component as a linear difference equation with a stochastic forcing process, i.e., autoregressive moving-average model.

Autoregressive moving average (ARMA) model

A process $\{X_t\}$ is said to be an ARMA(p, q) process if $\{X_t\}$ is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q},$$

where $a_t \sim WN(0, \sigma^2)$.

- $\{X_t\}$ is said to be an ARMA(p,q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p,q) process.
- In the compact notation: $\Phi(B)(X_t - \mu) = \Theta(B) a_t$, where $\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$.

e.g. $X_t - 0.5 X_{t-1} = a_t + 0.6 a_{t-1}$

where $p = 1, q = 1$

Backward shift Operator

$*$ $B X_t = X_{t-1}$ $B^2 X_t = X_{t-2}, \cdots, B^k X_t = X_{t-k}$

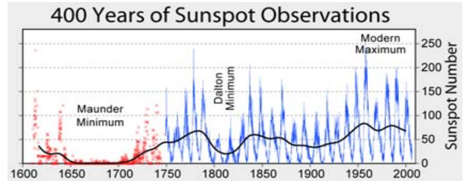
$X_t - 0.5 B X_t = a_t + 0.6 B a_t$

$(1 - 0.5B) X_t = (1 + 0.6B) a_t$

the 2 polynomials are

so $\Phi(B) = 1 - 0.5B$, $\Theta(B) = 1 + 0.6B$

The beginning of ARMA models



The origin of ARMA models can be traced back to Yule's work on Wolfer's sunspot numbers.

- See Wikipedia and Yule (1927).

ARMA models were widely accepted by researchers and professionals after the work of Box and Jenkins (1970). Therefore, analyzing time series using ARMA (or ARIMA) models is sometimes referred to as the Box-Jenkins approach.

The method of using ARMA

- The ARIMA model is an extension of an ARMA model, which allows modeling nonstationary time series.

Theoretical support of ARMA models

Wold Decomposition:

Any zero-mean process $\{X_t\}$ which is not *deterministic* can be expressed as a sum $X_t = U_t + V_t$ of an $MA(\infty)$ process $\{U_t\}$ and a deterministic $\{V_t\}$ which is uncorrelated with $\{U_t\}$.

- What is a deterministic process?
- What is an $MA(\infty)$ process?

Deterministic process

If the values $X_{n+j}, j \geq 1$ of the process $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$ were perfectly predictable in terms of $\mu_n = sp\{X_t, -\infty < t \leq n\}$. Such processes are called *deterministic*.

What does this mean?

- If X_n comes from a deterministic process, it can be predicted (or determined) by its past observations of the process, i.e., $X_t, t < n$.
- Does it sound like an autoregressive process?

Mathematical definition of Wold Decomposition

- If $\{X_t\}$ is a nondeterministic stationary time series, then

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} + V_t = \Psi(L) a_t + V_t,$$

Def (MA(∞))

Autoregressive processes

$$\boxed{\mu_n = \mathrm{sp}\{X_t, -\infty < t \le n\}.}$$

Moving average processes, linear combination of past innovations $\{a_t\}$

where

1. $\psi_0 = 1$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$,
2. $\{a_t\}$ is $WN(0, \sigma^2)$ with $\sigma^2 > 0$
3. $\mathrm{cov}(a_s, V_t) = 0, \forall\, s, t = 0, \pm 1, \pm 2, \cdots$ a_s not related with V_t term
4. V_t is deterministic

so: deterministic + MA(∞) = nondeterministic

Debates about the use of ARMA models

The Box-Jenkins and econometric traditions diverged at the outset.

Econometricians claimed that Box-Jenkins was atheoretical nonsense, a waste of time.

- And they only read the ARIMA part, which is atheoretical, and not the transfer function material.

Box-Jenkins analysts thought econometricians were dishonest in claiming all specification from theory and that in practice they took account of reality by cheating.

Reality

Things got really ugly when ARIMA models with a single parameter proved to be better forecasters than the 500 variable macro simultaneous equations produced by economists.

And then "pre-whitening analyses" demonstrated that most of textbook macroeconomics was wrong.

For the past 80 years, the ***Gaussian*** linear time series (ARMA) models have been extensively studied in academia and used by professionals.

Two special forms of ARMA models

(ignoring some parts of ARMA model)

Autoregressive model of order p, or $AR(p)$

- $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = a_t, a_t \sim WN(0, \sigma^2)$

Moving-average model of order q, or $MA(q)$

- $X_t = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, a_t \sim WN(0, \sigma^2)$

MA(q)

Moving average process of order q

$p=1$, nothing about autoregressive model.

The moving average process (or model) of order q, denoted as $MA(q)$, is given by

- $X_t = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} = \Theta(B) a_t$
- B is the backward shift operator, $B^h X_t = X_{t-h}$
- $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$
- $a_t \sim WN(0, \sigma^2)$

Question: What conditions do we need for $MA\ (p)$ processes to be stationary?

By def'n:
1. Check mean
2. Check lag. $\gamma(0)$
3. check $\gamma(1), \cdots, \gamma(q)$
4. lag higher than q is zero

a More general proof on p.

e.g.

$X_t = a_t + \theta a_{t-1}$, $a_t \sim NID(0, \sigma^2)\ \forall t$

mean constant
variance constant & exists
covariance indpt. of time
$\Rightarrow$ weekly stationary

① $E(X_t) = E(a_t + \theta a_{t-1}) = 0$

② $\gamma(0) = Var(X_t) = Cov(a_t + \theta a_{t-1}, a_t + \theta a_{t-1}) = E(a_t + \theta a_{t-1})^2 = E(a_t^2 + 2\theta a_t a_{t-1} + \theta^2 a_{t-1}^2) = E(\sigma^2 + 0 + \theta^2 \sigma^2) = (1+\theta^2)\sigma^2$

③ $\gamma(1) = Cov(a_t + \theta a_{t-1}, a_{t-1} + \theta a_{t-2}) = E[(a_t + \theta a_{t-1})(a_{t-1} + \theta a_{t-2})] = E(\theta a_{t-1}^2) = \theta \sigma^2$ $\Rightarrow$ indpt of time.

④ $\gamma(h) = \cdots$ similarly no t, indpt of t.
$(h \geq 2)$ $= 0$

Weekly Stationary Checked

MA(∞) processes

If $\{a_t\} \sim WN(0, \sigma^2)$ then we say that $\{X_t\}$ is a MA(∞) process of $\{a_t\}$ if there exists a sequence $\{\psi_j\}$ with $\sum_{j=0}^{\infty} |\psi_j| < \infty$ such that

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, t = \cdots, -1,0,1,2, \ldots$$

- Example 1: The MA(q) process is a moving average of $\{a_t\}$ with $\psi_j = \theta_j$, $j = 0,1,2, \ldots$ and $\psi_j = 0, i > q$.
- Example 2: The AR(1) process with $|\phi| < 1$ can be express as a moving average of $\{a_t\}$ with with $\psi_j = \phi^j$, $j = 0,1,2, \ldots$

MA(∞) processes cont'd

Theorem (Brockwell and Davis, 1992, p.91) The MA(∞) process is stationary with mean zero and autocovariance function

$$\boxed{\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}}$$

yep.

- We can calculate autocorrelation functions of a stochastic process $\{X_t\}$ using the aforementioned theorem as long as $\{X_t\}$ can be written in the form of a MA(∞) process.

ACF of MA(2) process

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}, a_t \sim WN(0, \sigma^2)$$

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, t = \cdots, -1,0,1,2, \ldots$$

$$\psi_j = \theta_j, j = 0,1,2 \text{ and } \psi_j = 0, j > 2$$

using thm to calculate lag $\gamma(\cdot)$ for simplification.

$\psi_0 \; \psi_1 \quad \psi_1 \; \psi_2 \quad \psi_2 \; \psi_3$

$$\gamma(1) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} = \sigma^2 (1 \cdot \theta_1 + \theta_1 \cdot \theta_2 + \theta_2 \cdot 0) = \sigma^2 (\theta_1 + \theta_1 \theta_2)$$

$\psi_0 \; \psi_2 \quad \psi_1 \; \psi_3 \quad \psi_2 \; \psi_4$

$$\gamma(2) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+2} = \sigma^2 (1 \cdot \theta_2 + \theta_1 \cdot 0 + \theta_2 \cdot 0) = \sigma^2 \theta_2$$

$$\gamma(3) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+3} = \sigma^2 (1 \cdot 0 + \theta_1 \cdot 0 + \theta_2 \cdot 0) = 0$$

Theorem: $\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$

ACF of MA(q) processes

$$X_t = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, a_t \sim WN(0, \sigma^2)$$

$$\psi_j = \theta_j, \ j = 0, 1, \cdots, q \text{ and } \psi_j = 0, \ j > q$$

$$\gamma(k) = \text{Cov}(X_t, X_{t+|k|})$$

using cov. by defn:

$$= \text{Cov}\left(\sum_{j=0}^{q} \theta_j a_{t-j}, \ \sum_{j=0}^{q} \theta_j a_{t+|k|-j}\right)$$

$$= \begin{cases} \sigma^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+|k|}, & k = 0, \pm 1, \cdots, \pm q \\ 0, & \text{otherwise} \end{cases}$$

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$$

moving average ends here.

AR(p)

Autoregressive model of order p

The autoregressive process (model) of order p, denoted as AR(p), is given by

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \Phi(B) X_t = a_t,$$

- where $a_t \sim WN(0, \sigma^2)$,
- $B^h X_t = X_{t-h}, h \in Z,$ backward operator
- $\Phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$

AR(1) processes

The autoregressive process of order one is given by $X_t - \phi X_{t-1} = a_t, \{a_t\} \sim WN(0, \sigma^2)$, where a_t is uncorrelated with X_s for all $s < t$.

AR(1) model and MA (∞) representation

b/c $X_{t-1} = \phi X_{t-2} + a_{t-1}$ iteration

$$X_t = \phi X_{t-1} + a_t = \phi(\phi X_{t-2} + a_{t-1}) + a_t$$
$$= \cdots$$
$$= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots$$
$$= \sum_{j=0}^{\infty} \phi^j a_{t-j} \quad (1)$$

current and past innovations

linear

X_t is the sum of past observation.

Stationarity condition $|\varphi| < 1$

$\mathrm{var}(X_t) = \mathrm{var}(\sum_{j=0}^{\infty} \varphi^j a_{t-j}) = \sum_{j=0}^{\infty} \varphi^{2j} \cdot \sigma^2$

The above infinite sum converges if $|\varphi| < 1$.

Need to check stationarity condition for ACFs

Convergence of a Geo. series.

MA infinite processes

if $|\Sigma \phi^j| < \infty$

AR(1) processes cont'd

- By eqn. (1), we have

$$\gamma(k) = \mathrm{cov}(X_t, X_{t+k})$$

$$= \mathrm{cov}\left(\sum_{l=0}^{\infty} \phi^l a_{t-l}, \sum_{j=0}^{\infty} \phi^j a_{t+k-j}\right)$$

$$= \mathrm{cov}\left(\sum_{l=0}^{\infty} \phi^l a_{t-l}, \sum_{s=0}^{k-1} \phi^s a_{t+k-s} + \sum_{l=0}^{\infty} \phi^{l+k} a_{t-l}\right)$$

$$= \phi^k \gamma(0)$$

- $K = 0$:

$$\gamma(0) = Var(X_t) = \sigma^2 / (1 - \phi^2)$$

We can also use the formula on page 14 to calculate ACFs.

$\boxed{\text{another perspective}}$

$$= \mathrm{Cov}\left(X_t, \phi^k X_t + \sum_{j=0}^{k-1} \phi^j a_{t+k-j}\right)$$

$$= \phi^k \gamma(0) + \underbrace{\mathrm{Cov}\left(X_t, \sum_{j=0}^{k-1} \phi^j a_{t+k-j}\right)}_{\text{no overlay!}}$$

$$= \phi^k \gamma(0) + 0$$

$$= \phi^k \gamma(0)$$

AR(1) processes cont'd

An AR(1) process with $|\phi| < 1$ is called causal or a future-independent autoregressive process. In this course, we consider only causal processes.

- By eqn. (1), we have

$$\gamma(k) = \operatorname{cov}(X_t, X_{t+k})$$

$$= \operatorname{cov}(\sum_{l=0}^{\infty} \phi^l a_{t-l}, \sum_{j=0}^{\infty} \phi^j a_{t+k-j})$$

$$= \operatorname{cov}(\boxed{\sum_{l=0}^{\infty} \phi^l a_{t-l}}, \sum_{s=0}^{k-1} \phi^s a_{t+k-s} + \boxed{\sum_{l=0}^{\infty} \phi^{l+k} a_{t-l}})$$

$$= \phi^k \gamma(0)$$

- $K = 0$:

From 1st Lec.

$$\gamma(0) = Var(X_t) = \sigma^2 / (1 - \phi^2)$$

finite if $|\phi| < 1$, why?

First order autoregression processes:

$\cdot\ X_t = \phi X_{t-1} + a_t = \cdots = \sum_{k=0}^{\infty} \phi^k a_{t-k}$

$\cdot\ Var(X_t) = \sum_0^{\infty} \phi^{2k} Var(a_t) = \frac{\sigma^2}{1-\phi^2}$

in detail:

$\gamma(0) = E(\sum_{j=0}^{\infty} \phi^j a_{t-j}, \sum_{s=0}^{\infty} \phi^s a_{t-s})$

$= \sigma^2 \sum_{j=0}^{\infty} (\phi^j)^2 = \frac{\sigma^2}{1-\phi^2}$

$X_{t+1} = \phi X_t + a_{t+1}$

$\Rightarrow X_t = \phi^{-1} X_{t+1} - \phi^{-1} a_{t+1}$

Alternate:

AR(1) processes cont'd

When we talk about weakly stat. we mean it causal.

Another $MA(\infty)$ representation:

$$X_t = \phi^{-1} X_{t+1} - \phi^{-1} a_{t+1} = \phi^{-1}(\phi^{-1} X_{t+2} - \phi^{-1} a_{t+2}) - \phi^{-1} a_{t+1}$$

$$= \cdots$$

$$= -\phi^{-1} a_{t+1} - \phi^{-2} a_{t+2} - \phi^{-3} a_{t+3} - \cdots$$

$$= -\sum_{j=1}^{\infty} \phi^{-j} a_{t+j}$$

Check expectations

future innovations

Every AR(1) process with $|\phi| > 1$ can also be written as an AR(1) process with $|\phi| < 1$ and a new white noise sequence. Nothing therefore is lost by eliminating AR(1) processes with $|\phi| > 1$ from consideration.

$$Var(X_t) = \sigma^2 (\phi^{-2} + \phi^{-4} + \cdots)$$

$$= \phi^{-2} \sigma^2 / (1 - \phi^{-2}) < \infty \quad if \ |\phi| > 1$$

convergence condition

No physical meaning b/c we don't need to be fortune teller.

Remarks

In the example of AR(1) processes, we check whether an AR(1) process is stationary via finding its MA(∞) representation and checking its autocovariance functions.

For the second-order autoregressive AR(2) process, we have $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t, a_t \sim WN(0, \sigma^2)$.

- Can we write an AR(p) process in terms of an MA(∞) process
- How to check whether an AR(p) model is stationary for $p \geq 2$?

General approach to checking stationarity

B outside unit circle

A general way of checking the stationarity condition of autoregressive processes is that the roots of the following equation $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p = 0$ must lie outside the unit circle.

Example: Check the stationarity of an AR(1) model

- $(1 - \phi B)X_t = a_t$
- $\Phi(B) = 1 - \phi B = 0$ so $B = 1/\phi$
- $|B| = |1/\phi| > 1$ so $|\phi| < 1$
- Exercise: Study the stationary conditions for an AR(2) process

Causal and invertible processes

$\boxed{\Phi(B)X_t = \Theta(B)a_t}$

check for causal (unit circle, outside) $\Phi(B)$

check for invertible, similarly $\Theta(B)=0$

Causal processes:

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

- $\{X_t\}$ can be expressed in terms of $\{a_s, s \leq t\}$. Such processes are called causal or future-independent autoregressive process (Brockwell and Davis; 1992).

. able to use ACF/PACF b/c it's stationary.

Invertible processes:

$a_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad \pi_0 = 1$

- In general, no restrictions on the $\{\theta_i\}$ are required for a (finite-order) MA process to be stationary. The imposition of the invertiblity condition ensures that there is a unique MA process for a given set of ACF.

$AR \iff MA$

Duality between AR and MA processes

A finite-order stationary AR(p) process corresponds to a MA(∞) process, and a finite-order invertible MA(q) corresponds to an AR process of infinite-order .

- Casual/stationary AR(p) ➔ MA(∞)
- Invertible MA(q)➔ AR(∞)
- This dual relationship also exists in autocorrelation and partial autocorrelation functions.

Remarks on invertibility

- Consider the following first-order MA processes:

A: $X_t = a_t + \theta a_{t-1}$

B: $X_t = a_t + \frac{1}{\theta} a_{t-1}$

A: $a_t = X_t - \theta X_{t-1} - \theta^2 X_{t-2} - \cdots$, if $|\theta| < 1$

B: $a_t = X_t - \frac{1}{\theta} X_{t-1} - \frac{1}{\theta^2} X_{t-2} - \cdots$, if $|\theta| > 1$

$$\gamma_A(0) = \text{var}(a_t + \theta a_{t-1}) = (1+\theta^2)\sigma^2$$
$$\gamma_A(1) = \text{cov}(a_t + \theta a_{t-1}, a_{t+1} + \theta a_t) = \theta\sigma^2$$
$$\rho_A(1) = \gamma_A(1)/\gamma_A(0) = \theta/(1+\theta^2)$$

$$\gamma_B(0) = \text{var}(a_t + \frac{1}{\theta} a_{t-1}) = (1+\frac{1}{\theta^2})\sigma^2$$
$$\gamma_B(1) = \text{cov}(a_t + \frac{1}{\theta} a_{t-1}, a_{t+1} + \frac{1}{\theta} a_t) = \frac{1}{\theta}\sigma^2$$
$$\rho_B(1) = \frac{\frac{1}{\theta}\sigma^2}{(1+\frac{1}{\theta^2})\sigma^2} = \theta/(1+\theta^2)$$

Two different models possess the same autocorrelation functions

- So you cannot tell 2 models by deriving ACF/PACF. ?

Causal ARMA processes

An ARMA(p,q) process, defined by the equation $\Phi(B)X_t = \Theta(B)a_t$, is said to be causal if there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty}|\psi_j| < \infty$ and $X_t = \sum_{j=0}^{\infty}\psi_j a_{t-j}$, $t = 0, \pm 1, \pm 2, \ldots$

- $$\Psi(z) = \sum_{j=0}^{\infty}\psi_j \cdot z^j = \frac{\Theta(z)}{\Phi(z)}, \quad |z| \le 1$$

Sps. $\Phi(B)X_t = \theta(B)a_t$

$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots \phi_p B^p$

$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$

$X_t = \psi(B)a_t$

$\psi(B) = \psi_0(B) + \psi_1 B + \psi_2 B^2 + \cdots$

$X_t = \frac{\theta(B)}{\Phi(B)} a_t = \psi(B)a_t$

$\theta(B) = \Phi(B)\psi(B)$

$1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$

$= (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)$

B^0: $1 = 1 * \psi_0$

B^1: $\theta_1 = -\phi_1\psi_0 + \psi_1 \Rightarrow \psi_1 = \theta_1 + \phi_1$

B^2: $\theta_2 = -\phi_1\psi_1 - \phi_2 + \psi_2 \Rightarrow \psi_2 = \cdots$

B^3: $\cdots$ (iteration)

(see next page)

Calculate ψ_j of an ARMA(p,q) process

$$\Psi(B)\Phi(B) = \Theta(B)$$

$$\Leftrightarrow \quad (\sum_{j=0}^{\infty} \psi_j B^j)(\sum_{k=0}^{p} \phi_k B^k) = \sum_{l=0}^{q} \theta_l B^l$$

$$\Leftrightarrow \quad (\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - \phi_1 B - \cdots - \phi_p B^p)$$

$$= 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$

Equate coefficients of B^j and solve for $\{\psi_j\}, j = 0,1,2, \ldots$

$$B^0 : \psi_0 = 1$$

$$B^1 : -\psi_0\phi_1 + \psi_1 = \theta_1$$

$$B^2 : -\psi_0\phi_2 - \psi_1\phi_1 + \psi_2 = \theta_2$$

$$B^3 : -\psi_0\phi_3 - \psi_1\phi_2 - \psi_2\phi_1 + \psi_3 = \theta_3$$

$$B^4 : \cdots \quad \cdots$$

Note: In the advanced time series course, psi coefficients may be calculated by solving difference equations.

Invertible ARMA models

An $ARMA(p,q)$ process is said to be invertible if there exists a sequence of constants $\{\pi_j\}$ such that $\sum_{j=0}^{\infty}|\pi_j| < \infty$ and $a_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, t = 0, \pm 1, \pm 2, \ldots$

- $$\Pi(z) = \sum_{j=0}^{\infty} \pi_j \cdot z^j = \frac{\Phi(z)}{\Theta(z)}, \quad |z| \le 1$$

$$\Pi(B)\Theta(B) = \Phi(B)$$

$$\Leftrightarrow \quad (\sum_{j=0}^{\infty} \pi_j B^j)(\sum_{l=0}^{q} \theta_l B^l) = \sum_{k=0}^{p} \phi_k B^k$$

$$\Leftrightarrow \quad (\pi_0 + \pi_1 B + \pi_2 B^2 + \cdots)(1 + \theta_1 B + \cdots + \theta_q B^q)$$

$$= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

Sps. $\Phi(B)X_t = \theta(B)a_t$

$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots \phi_p B^p$

$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$

$\Pi(B) X_t = a_t$

$\Pi(B) = \Pi_0 + \Pi_1 B + \Pi_2 B^2 + \cdots$

$\frac{\Phi(B)}{\theta(B)} X_t = a_t = \Pi(B) X_t$

$\Phi(B) = \theta(B)\Pi(B)$

特定系数法 again

(see next page)

Calculate π_j of an MA(q) processes

$$X_t = (1+\theta_1 B + \cdots + \theta_q B^q) a_t$$

$$(\pi_0 + \pi_1 B + \pi_2 B^2 + \cdots)(1+\theta_1 B + \cdots + \theta_q B^q) = 1$$

Equate coefficients of B^j, and solve $\{\pi_j\}$ for $j = 0,1,2,\ldots$

$$B^0 : \pi_0 = 1$$

$$B^1 : \pi_0\theta_1 + \pi_1 = 0 \;\Rightarrow\; \pi_1 = -\theta_1$$

$$B^2 : \pi_0\theta_2 + \pi_1\theta_1 + \pi_2 = 0 \;\Rightarrow\; \pi_2 = \theta_1^2 - \theta_2$$

$$B^3 : \pi_0\theta_3 + \pi_1\theta_2 + \pi_2\theta_1 + \pi_3 = 0$$

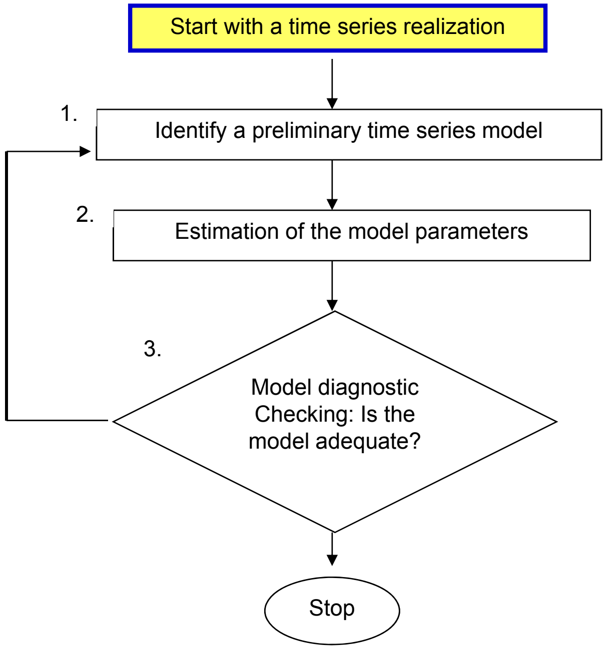
$$\Rightarrow \pi_3 = -\theta_1(\theta_1^2 - \theta_2) + \theta_1^2 - \theta_3$$

$$B^4 : \cdots \quad \cdots$$

Note: Similar, pi coefficients may be calculated by solving difference equations.

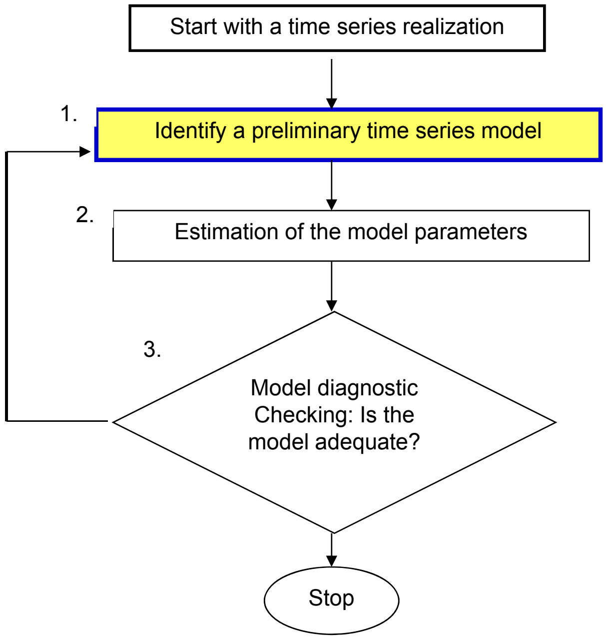
THREE STAGES OF THE BOX-JENKINS ANALYSIS

Three stages of Box-Jenkins Approach



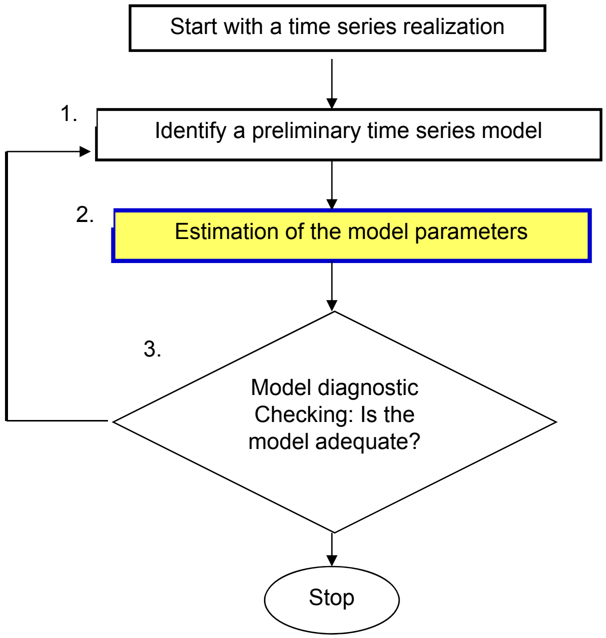
- Understand the problem of interest
- Collect data
- Plot time series data

First stage



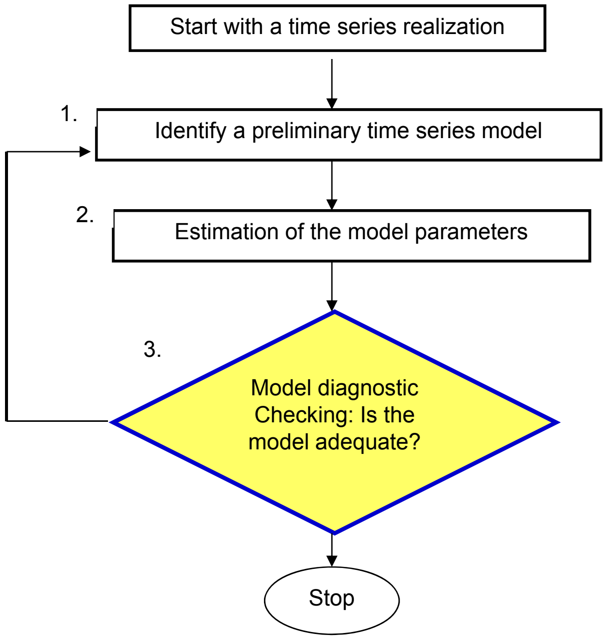
- Perform differencing and transformations to transform the data into stationarity.
- Identify preliminary ARMA(p,q) models using sample autocorrelations and sample partial autocorrelations Tools

Second stage



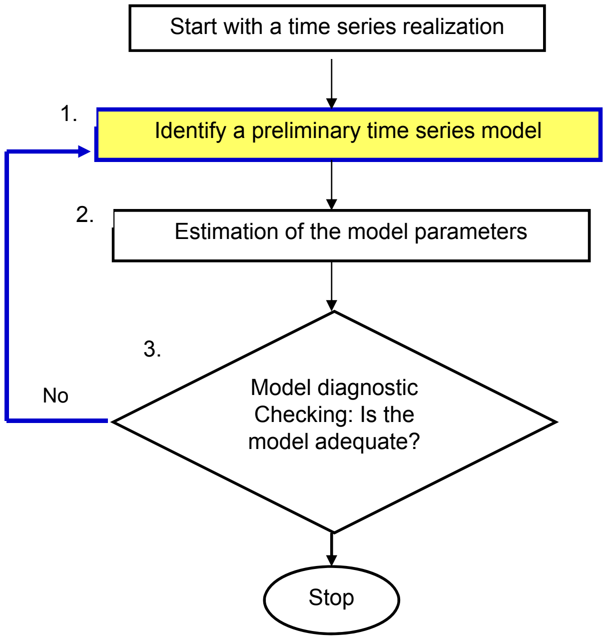
- Method of moments
- Maximum likelihood Estimation
- Kalman Filter
- Others

Third stage



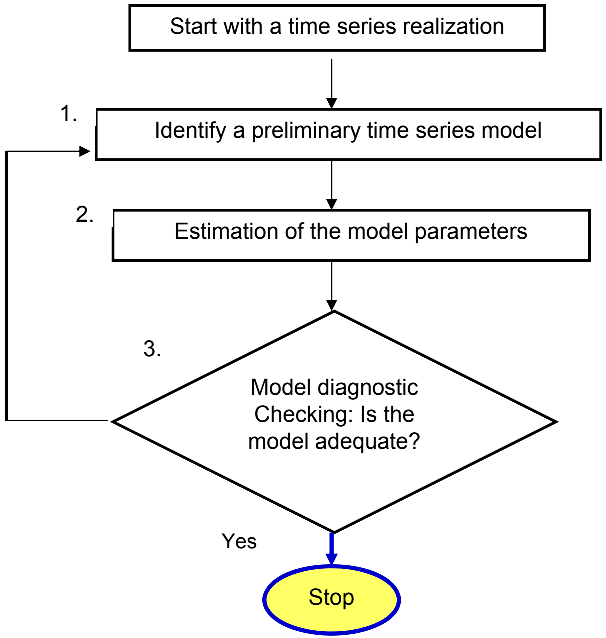
- Model adequacy is checked by examining if the residuals of the fitted model are approximately uncorrelated (after taking into account the effect of estimation)

Model is not adequate



- The fitted model fails diagnostic checks.
- Return to the first stage and identify another time series model.

Model is adequate



- If the fitted model passes diagnostic checks, we may the use the model for our analysis.

Covariance Matrix

	X_1	X_2	X_3	X_4
X_1	σ_{11}	σ_{12}	σ_{13}	σ_{14}
X_2	σ_{21}	σ_{22}	σ_{23}	σ_{24}
X_3	σ_{31}	σ_{32}	σ_{33}	σ_{34}
X_4	σ_{41}	σ_{42}	σ_{43}	σ_{44}

$$\left.\begin{aligned} \sigma_{ii} &= \gamma(0) \\ \sigma_{ij} &= cov(X_i, X_j) \\ &= \gamma(1) \quad \text{if } |i-j|=1 \\ &= 0 \quad \text{if } |i-j|>1 \end{aligned}\right\} \Rightarrow$$

First order

Distn $(X_1,X_2),(X_2,X_3),(X_3,X_4) \sim N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\gamma(0) & \gamma(1)\\ \gamma(1) & \gamma(0)\end{pmatrix}\right)$

Distn $(X_1,X_3),(X_2,X_4) \sim N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\gamma(0) & 0\\ 0 & \gamma(0)\end{pmatrix}\right)$

$(X_1,X_2,X_3),(X_2,X_3,X_4) \sim N_3(\cdot\cdot\cdot)$

MODEL IDENTIFICATION

- The first stage of Box-Jenkins analysis
- Review theoretical ACFs and PACFs of ARMA models
- Model identification using ACF and PACF with R examples

Later, we are gonna see MA(q), AR(1) & AR(p)

MA(1) processes

- Autocorrelation functions

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{1+\theta^2}$$

$$\rho_j = 0 \quad j > 1$$

(but ACF helps)

- Partial autocorrelation functions

PACF (not useful for identification)

$$\phi_{11} = \rho_1 = \frac{-\theta_1}{1+\theta_1^2} = \frac{-\theta_1(1-\theta_1^2)}{1-\theta_1^4}$$

$$\phi_{22} = -\frac{\rho_1^2}{1-\rho_1^2} = \frac{-\theta_1^2}{1+\theta_1^2+\theta_1^4} = \frac{-\theta_1^2(1-\theta_1^2)}{1-\theta_1^6}$$

$$\phi_{33} = \frac{\rho_1^3}{1-2\rho_1^2} = \frac{-\theta_1^3}{1+\theta_1^2+\theta_1^4+\theta_1^6} = \frac{-\theta_1^3(1-\theta_1^2)}{(1-\theta_1^8)}.$$

In general,

$$\phi_{kk} = \frac{-\theta_1^k(1-\theta_1^2)}{1-\theta_1^{2(k+1)}}, \quad \text{for } k \geq 1.$$

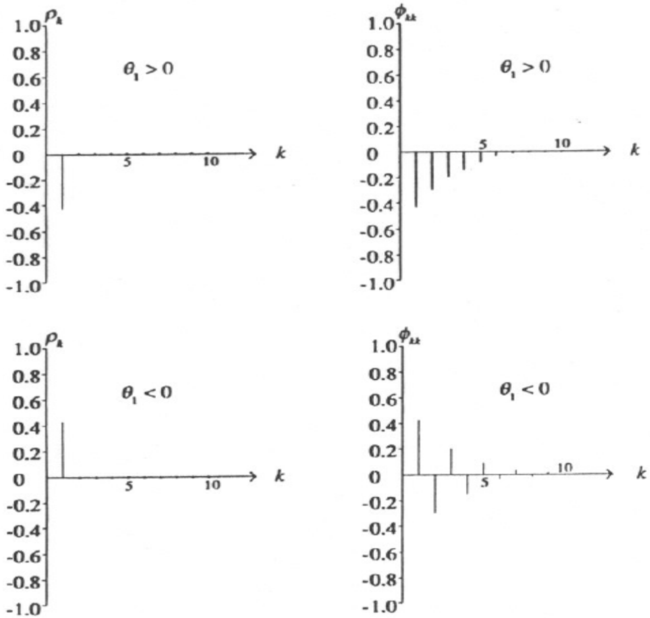


Fig. 3.10 ACF and PACF of MA(1) processes: $\dot{Z}_t = (1 - \theta B)a_t$.

Contrary to its ACF, which cuts off after lag 1, the PACF of an MA(1) model tails off exponentially in one of two forms depending on the sign of θ_1 (hence on the sign of ρ_1). If alternating in sign, it begins with a positive value; otherwise, it decays on the negative side, as shown in Figure 3.10. We also note that $|\phi_{kk}| < 1/2$.

MA(q) processes

The maximum lag of the non-zero sample autocorrelation is a good indicator of the MA(q) processes.

- The ACF of MA(q) processes, $X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q}$, cut off after lag q.
- $\rho_k = \begin{cases} \frac{\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, & k = 1, \dots, q \\ 0, & k > q \end{cases}$
- How about ACF of the $AR(p)$ processes?

AR(1) processes

- Autocovariance and Autocorrelation:

$$\gamma_j = \phi\gamma_{j-1} \quad j \geq 1$$

$$\rho_j = \frac{\gamma_j}{\gamma_o} = \phi\frac{\gamma_{j-1}}{\gamma_0} = \phi\rho_{j-1} \quad j \geq 1$$

$$\rho_j = \phi^2\rho_{j-2} = \phi^3\rho_{j-3} = \cdots = \phi^j\rho_0 = \phi^j$$

- Partial autocorrelation functions:

$$\phi_{11} = \rho_1 = \phi$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \frac{\phi^2 - \phi^2}{1-\rho_1^2} = 0$$

$$\boxed{\phi_{kk} = 0 \quad k \geq 2}$$

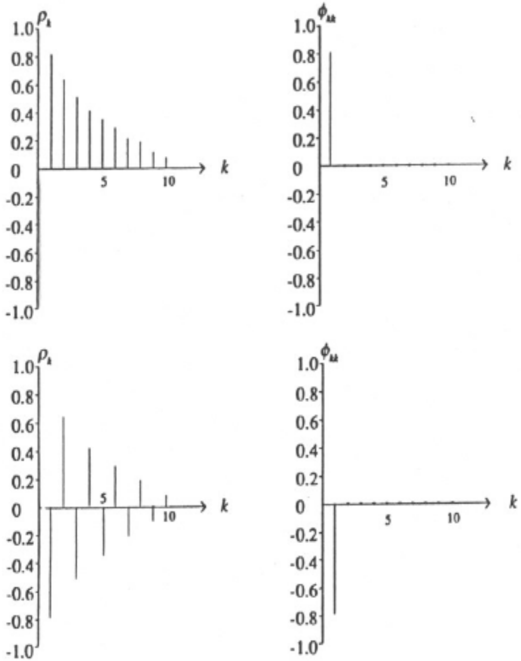


Fig. 3.1 ACF and PACF of the AR(1) process: $(1-\phi B)\dot{Z}_t = a_t$.

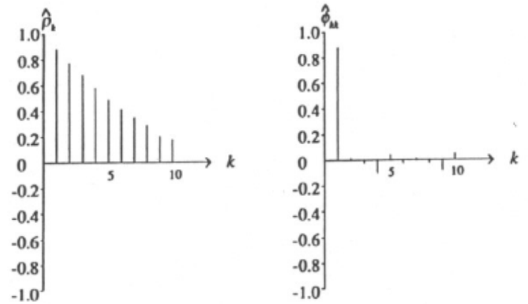


Fig. 3.3 Sample ACF and sample PACF of a simulated AR(1) series: $(1-.9B)(Z_t-10)=a_t$.

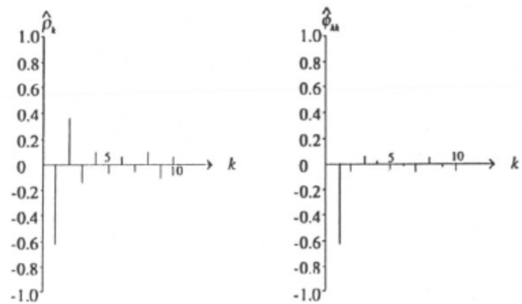


Fig. 3.5 Sample ACF and sample PACF of a simulated AR(1) series $(1+.65B)(Z_t-10)=a_t$.

AR(p) processes

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots \phi_p X_{t-p} + a_t$$

Stationarity condition of an AR(p) models are that all p roots of the characteristic equation outside of the unit circle

ACF

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots\ldots\phi_p\rho_{k-p}$$

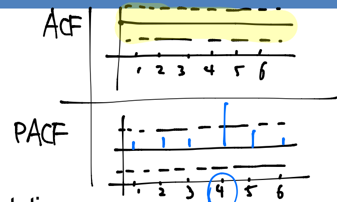
$$\left.\begin{array}{l} \rho_1 = \phi_1\rho_0 + \phi_2\rho_1 + \ldots\ldots\phi_p\rho_{p-1} \\ \rho_2 = \phi_1\rho_{11} + \phi_2\rho_0 + \ldots\ldots\phi_p\rho_{p-2} \\ \vdots \\ \rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots\ldots\phi_p\rho_0 \end{array}\right\}$$

System to solve for the first p autocorrelations: p unknowns and p equations

ACF decays as mixture of exponentials and/or damped sine waves--depending on real/complex roots

PACF

$$\phi_{kk} = 0 \text{ for } k > p$$



$\sim N(0, \frac{1}{n})$

n: # of observations

95% CI: $\pm 2\frac{1}{\sqrt{n}}$

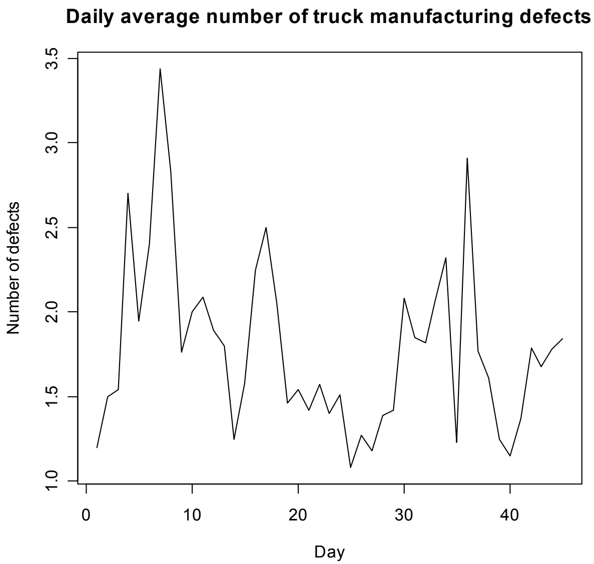
significant.

TABLE 6.1(Wei, 2006)

Characteristics of theoretical ACF and PACF for stationary processes

Process	ACF	PACF
AR(p)	Tails off as exponential decay or damped sine wave	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off as exponential decay or damped sine wave
ARMA(p,q)	Tails off after lag (q-p)	Tails off after lag (p-q)

#Example 6.1 (Series W1)
Daily average number of truck manufacturing defects



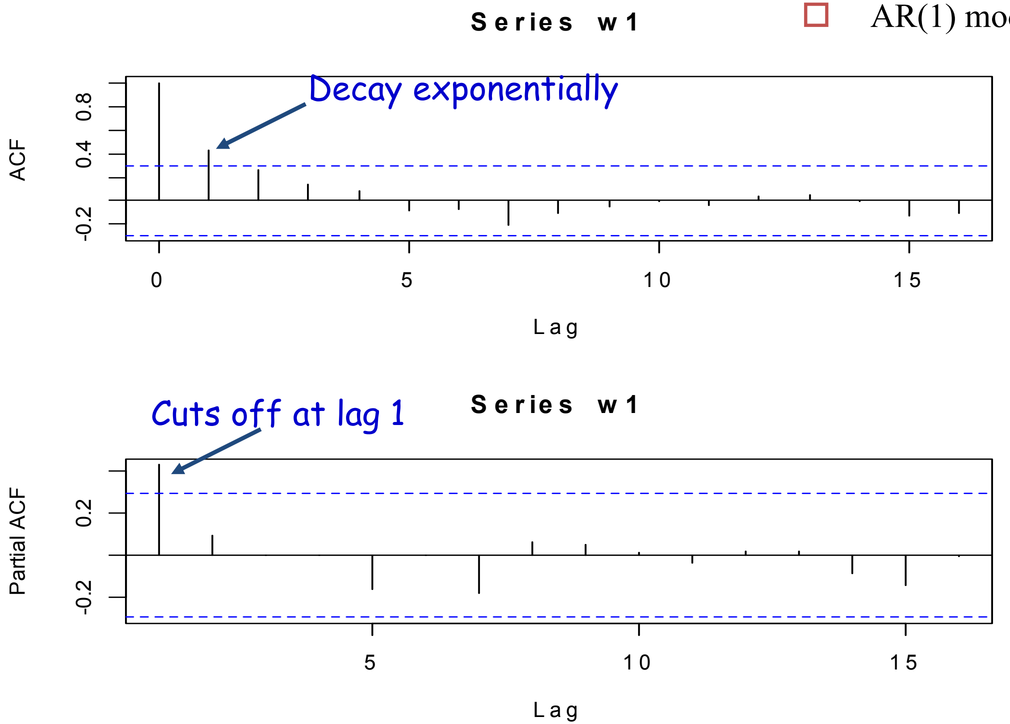
- w1<-scan("D:\\Documents and Settings\\jenwenl\\Desktop\\time series analysis\\R example\\w1.txt")
- plot(w1,xlab="Day",ylab=" Number of defects",type="l")
- title("Daily average number of truck manufacturing defects")

Characters of AR(1)

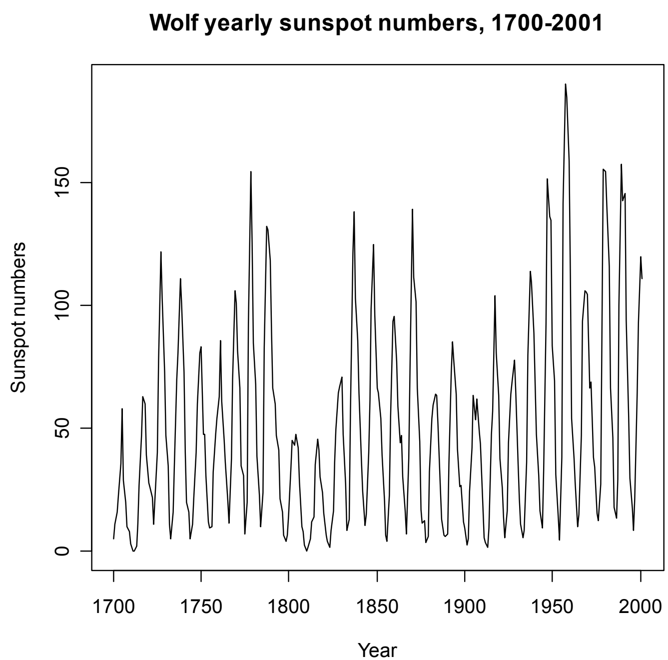
#Example 6.1 (Series W1)

par(mfrow=c(2,1)); acf(w1); pacf(w1)

- ACF decays exponentially
- PACF cuts off at lag one
- AR(1) model



#Example 6.2 (Series W2)
Wolf yearly sunspot numbers, 1700-2001

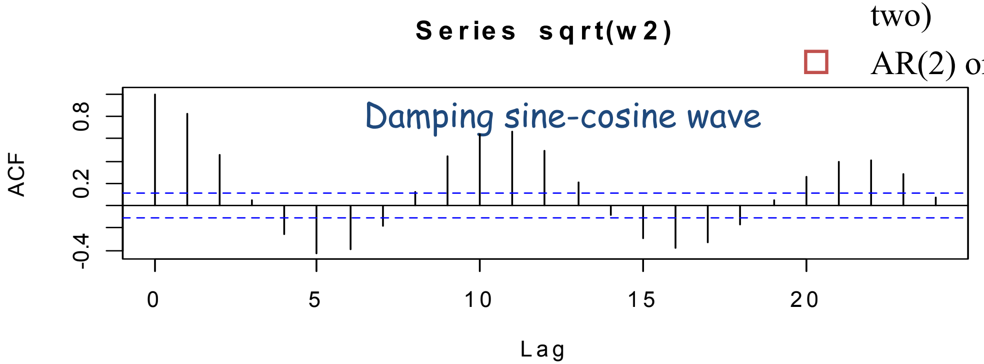


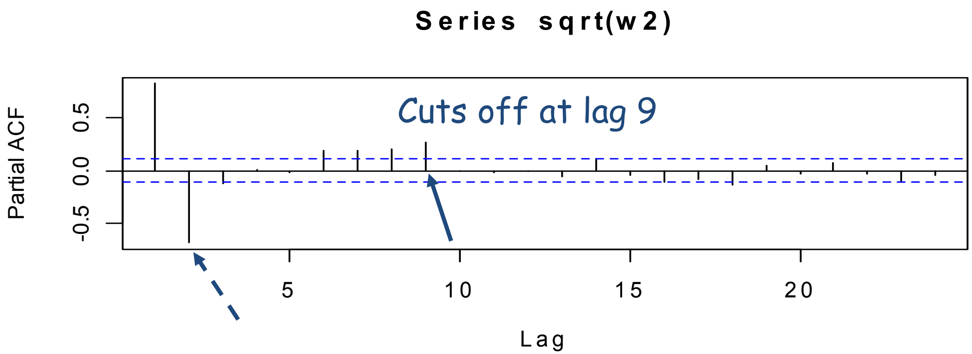
- w2<-scan("D:\\Documents and Settings\\jenwenl\\Desktop\\time series analysis\\R example\\w2.txt")
- plot(1700:(1700+301),w2,xlab="Year",ylab="Sunspot numbers",type="l")
- title("Wolf yearly sunspot numbers, 1700-2001")

#Example 6.2 (Series W2)

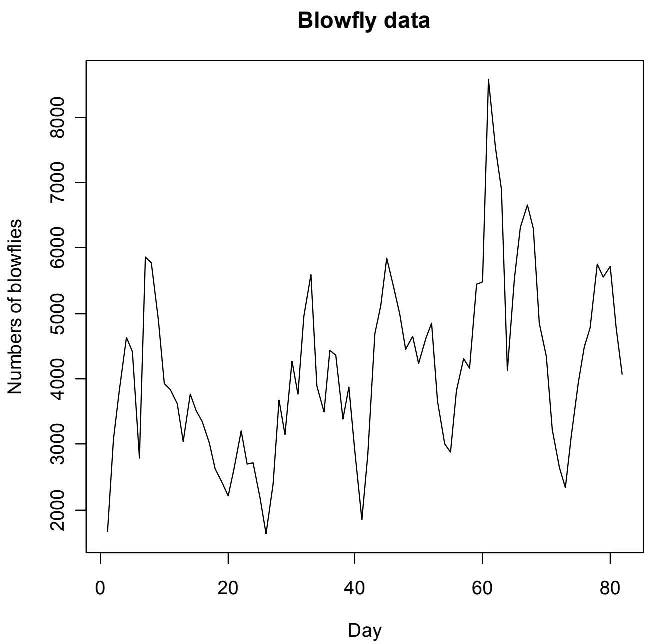
par(mfrow=c(2,1)); acf(sqrt(w2)); pacf(sqrt(w2))

- ACF damping sinusoid wave
- PACF cuts off at lag nine (or two)
- AR(2) or AR(9) model





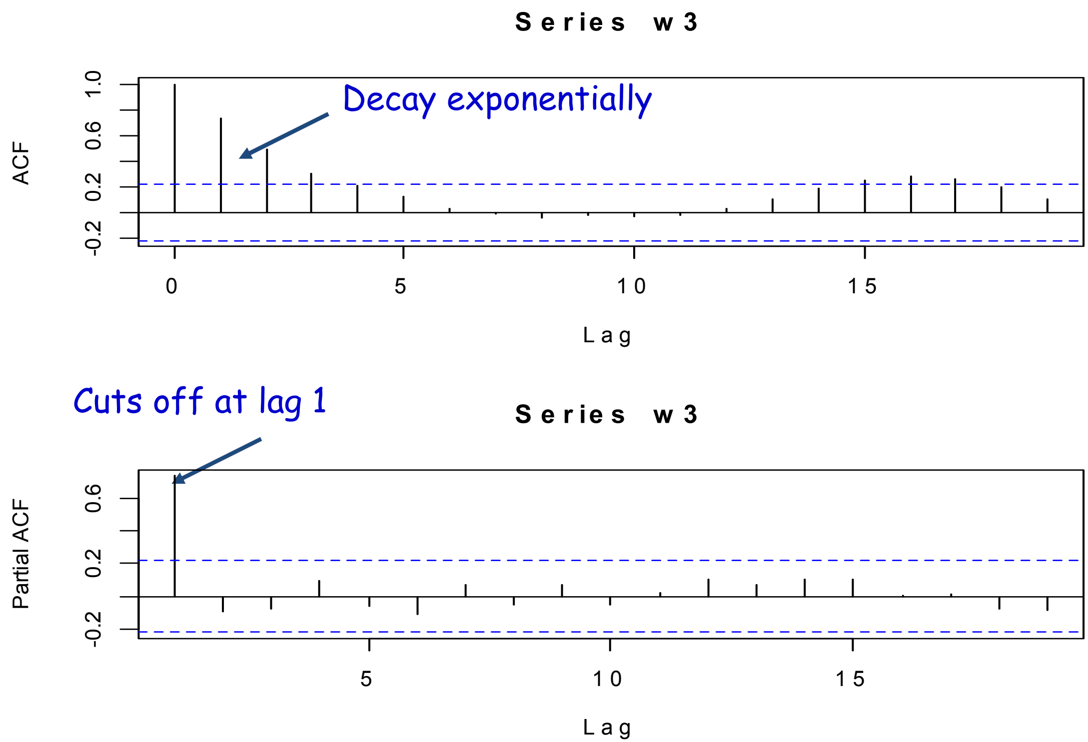
#Example 6.3 (Series W3)
Blowfly data



- w3<-scan("D:\\Documents and Settings\\jenwenl\\Desktop\\time series analysis\\R example\\w3.txt")
- plot(w3,xlab="Day",ylab="Numbers of blowflies",type="l")
- title("Blowfly data")

#Example 6.3 (Series W3)

```
par(mfrow=c(2,1)); acf(w3); pacf(w3)
```

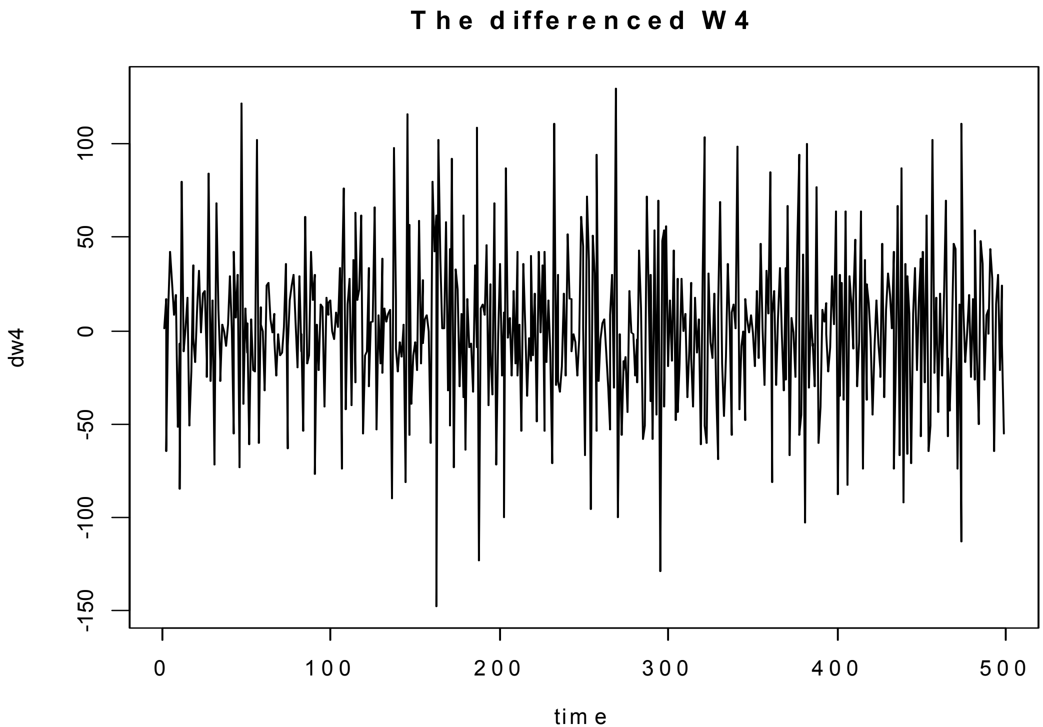


#Example 6.4 (Series W4)
US monthly series of unemployed young woman between 16 from January 1961 to August 2002

- w4<-scan("D:\\Documents and Settings\\jenwenl\\Desktop\\time series analysis\\R example\\w4.txt")
- plot(w4,xlab="time",ylab="Number of persons",type="l")
- title("US monthly series of unemployed young women")

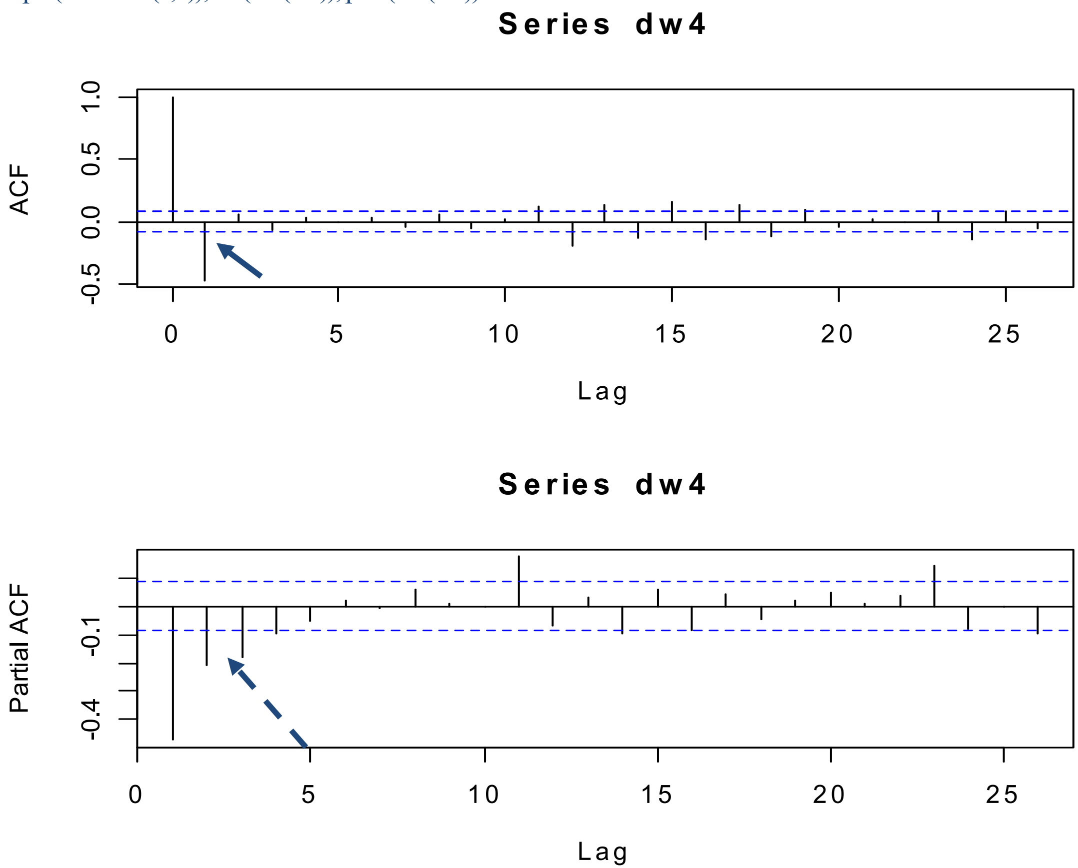
Transform data by Differencing

```
dw4<-diff(w4); plot(dw4,xlab="time",type="l")
```



#Example 6.4 (Series W4)

par(mfrow=c(2,1)); acf(diff(w4)); pacf(diff(w4))



MODEL ESTIMATION

Minimum MSE prediction and yule-walker equations

Minimum mean square error prediction

To predict future values of a time series x_{n+m}, $m = 1,2,\cdots$, based on the data collected to the present, $\boldsymbol{x} = \{x_n, x_{n-1}, \dots, x_1\}$

- Assume that x_t is stationary and the model parameters are known
- The minimum mean square error predictor of x_{n+m} is the conditional expectation $x_{n+m}^n = E(x_{n+m}|\boldsymbol{x})$
- In what follows, we restrict predictors to linear functions of data, $x_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k x_k$ and consider the best linear prediction for stationary processes

Best linear prediction for stationary processes

Given data $x_1, \ldots, x_n$, the best linear predictor (BLP),

$$x_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k x_k, \text{ of } x_{n+m}, \text{ for } m \geq 1,$$

is found by solving

$$E\left[\left(x_{n+m} - x_{n+m}^n\right)x_k\right] = 0, \quad k = 0,1, \ldots, n,$$

where $x_0 = 1$, for $\alpha_0, \alpha_1, \ldots, \alpha_n$.

$$E(X_{n+m} - \underbrace{X_{n+m}^{n}}_{\sum_{k=1}^{n} \alpha_k X_k})^2$$

$$E(2 \triangle X_k) = 0 = \frac{\partial L}{\partial \alpha_k}$$

$$\min_{\{\alpha_k : k=1, n\}} L = E(\overbrace{X_{n+m} - \sum_{k=1}^{n} \alpha_k X_k}^{\triangle})^2$$

$$\frac{\partial L}{\partial \alpha_k} = 0 \quad , k = 1, \cdots, n$$

One step ahead prediction and BLP

Best Linear Prediction

Consider the BLP of x_{n+1} as

$$x_{n+1}^n = \phi_{n1} x_n + \phi_{n2} x_{n-1} + \cdots + \phi_{nn} x_1,$$

where $\alpha_k = \phi_{n,n+1-k},\ k = 1, \dots, n$

Using the property 3.3 in the textbook, the coefficients $\{\phi_{n1}, , \dots, \phi_{nn}\}$ satisfy

$$\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n,$$

where $\Gamma_n = \{\gamma(k-j)\}_{j,k=1}^n$ is an $n \times n$ matrix, $\boldsymbol{\phi}_n = (\phi_{n1}, \phi_{n2}, \dots, \phi_{nn})'$ and $\boldsymbol{\gamma}_n = \big(\gamma(1), \dots, \gamma(n)\big)'$.

$E\{(X_{n+1} - \sum_{j=1}^n \alpha_j X_j) X_k\} = 0,\ k = 1, \cdots, n$

$k=1:\ \gamma(n) = \sum_{j=1}^n \alpha_j \gamma(1-j)$

$\vdots$

$k=n:\ \gamma(1) = \sum_{j=1}^n \alpha_j \gamma(n-j)$

More one step ahead BLP

Γ_n is nonnegative definite and nonsingular, the element $\boldsymbol{\phi}_n$ are unique and are given by

$$\boldsymbol{\phi}_n = \Gamma_n^{-1} \boldsymbol{\gamma}_n, \qquad (3.64)$$

In vector notation, we can express $x_{n+1}^n = \boldsymbol{\phi}_n' \boldsymbol{x}$

The mean square one-step-ahead prediction error is

$$P_{n+1}^n = E(x_{n+1} - x_{n+1}^n)^2 = \gamma(0) - \boldsymbol{\gamma}_n' \Gamma_n^1 \boldsymbol{\gamma}_n, \quad (3.66)$$

Durbin-Levinson algorithm

Eqn. (3.64) and (3.66) can be solved iteratively as follows:

$$\phi_{00} = 0, \qquad P_1^0 = \gamma(0), \qquad (3.68)$$

For $n \geq 1$,

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k}\, \rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(k)}, \qquad P_{n+1}^n = P_n^{n-1}(1 - \phi_{nn}^2),$$

$$(3.69)$$

where, for $n \geq 2$,

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn}\phi_{n-1,n-k}, \qquad k = 1,2, \ldots, n-1, \qquad (3.70)$$

Example 3.19 Using the Durbin-Levinson Algorithm

- $n = 1,: \phi_{11} = \rho(1),\ P_2^1 = \gamma(0)(1 - \phi_{11}^2)$
- $n = 2\ (k = 1)$:

$$\phi_{22} = \frac{\rho(2) - \phi_{11}\rho(1)}{1 - \phi_{11}\rho(1)},$$

$$P_3^2 = P_2^1(1 - \phi_{11}^2)(1 - \phi_{22}^2)$$

$$\phi_{21} = \phi_{11} - \phi_{22}\phi\_11$$

- $n = 3\ (k = 1{,}2)$:

$$\phi_{33} = \frac{\rho(3) - \phi_{21}\rho(2) - \phi_{22}\rho(1)}{1 - \phi_{21}\ \rho(1) - \phi_{22}(2)}$$

$$P_4^3 = P_3^2[1 - \phi_{33}^2] = \gamma(0)\prod_1^3(1 - \phi_{jj}^2)$$

$$\phi_{31} = \phi_{21} - \phi_{33}\phi_{22} \qquad (n = 3, k = 1)$$

$$\phi_{32} = \phi_{22} - \phi_{33}\phi_{21} \qquad (n = 3, k = 2)$$

Iterative solution for PACF

Property 3.5 (textbook)

- The PACF of a stationary x_t, can be obtained iteratively via eqn. (3.69) as ϕ_{nn} for $n = 1,2, \ldots$.

PACFs of an AR(2) process: $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t$

- $\phi_{11} = \rho(1) = \frac{\phi_1}{1-\phi_2}$
- $\phi_{22} = \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2} = \phi_2$
- $\phi_{33} = \frac{\rho(3)-\phi_1\rho(2)-\phi_2\rho(1)}{1-\phi_1\rho(1)-\phi_2\rho(2)} = 0$
- Remark: using Yule-Walker equations, we have $\rho(1) = \frac{\phi_1}{1-\phi_2}$ and $\rho(2) = \phi_1\rho(1) + \phi_2$, and $\rho(h) - \phi_1\rho(h-1) - \phi_2(h-2) = 0, h \geq 1$

Yule-Walker equations

The Yule-Walker equations of an AR(p) process
$X_t = \phi_1 X_{t-1} + \cdots + \phi_p x_{t-p} + a_t$ are given by

$$\gamma(h) = \phi_1\gamma(h-1) + \cdots + \phi_p\gamma(h-p), h = 1, \ldots, p, \qquad (3.98)$$

$$\sigma_a^2 = \gamma(0) - \phi_1\gamma(1) - \cdots - \phi_p\gamma(p), \qquad (3.99)$$

In matrix notation, we have

$$\Gamma_p\boldsymbol{\phi} = \boldsymbol{\gamma}_p,\ \sigma_a^2 = \gamma(0) - \boldsymbol{\phi}'\boldsymbol{\gamma}_p, \quad (3.100),$$

where $\Gamma_p = \{\gamma(k-j)\}_{j,k=1}^{p}$, $\boldsymbol{\phi} = \left(\phi_1, \ldots, \phi_p\right)'$ and $\boldsymbol{\gamma}_p = (\gamma(1), \cdots, \gamma(p))'$.

Yule-Walker equations

Replace $\gamma(h)$ in (3.100) by $\hat{\gamma}(h)$ and solve for

$$\widehat{\boldsymbol{\phi}} = \hat{\Gamma}_p^{-1}\widehat{\boldsymbol{\gamma}}_p, \hat{\sigma}_a^2 = \hat{\gamma}(0) - \widehat{\boldsymbol{\gamma}}'_p\hat{\Gamma}_p^{-1}\widehat{\boldsymbol{\gamma}}_p. \quad (3.101)$$

Estimators in (3.101) are typically called the Yule-Walker estimators. In practice, it is sometimes convenient to work with the sample autocorrelation function version of the Yule-Walker equations.

For AR(p) models, if the sample size is large, the wule-Walker estimators are approximately normally distributed and $\hat{\sigma}_a^2$ is close to the true value of σ_a^2.

Large sample results for Yule-Walker estimators

The asymptotic ($n \to \infty$) behavior of the Yule-Walker estimators in the case of causal AR(p) processes is as follows:

$$\sqrt{n}\left(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \sim N\left(0, \sigma_a^2 \Gamma_p^{-1}\right), \qquad \hat{\sigma}_a^2 \to \sigma_a^2$$

- In practice, we replace Γ_p^{-1} and σ_a^2 by their sample estimates while evaluating the above asymptotic distribution
- Example: see “Example 3.26” in the textbook (page. 122)

MODEL ADEQUACY

- Model adequacy is at the third stage of Box-Jenkins framework
- This stage is also called model diagnostic checking which involves techniques like over-fitting, residual plots and, more importantly, checking if residuals are approximately uncorrelated.

Why residuals are uncorrelated?

The residuals of a fitted ARMA(p,q) mode is

$$\hat{a}_t = X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_p X_{t-p} - \hat{\theta}_1 a_{t-1} - \cdots - \hat{\theta}_q a_{t-q},$$

where $\hat{\phi}_k$, $\hat{\theta}_k$ $\forall k$ are the parameter estimates obtained from the second stage, and $\{\hat{a}_t\}$ are the residuals of the fitted model

- Residuals can be seen as the sample estimates of $\{\hat{a}_t\}$ and therefore are approximately uncorrelated (white noise) because of the estimation process.
- Remark: residuals are not independent in the classical regression model

Autocorrelation among residuals

Residual autocorrelation functions at lag k

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=1}^{n} \hat{a}_t^2}$$

- How many lags are enough??
- The overall tests that check an entire group of residual autocorrelation functions (assuming that the model is adequate) are called portmanteau tests.
- In spirits, portmanteau tests may be seen as a variant of the goodness of fit tests.

Popular portmanteau tests

Box and Pierce (1970)

$$Q_{BP} = n \cdot \sum_{1}^{m} \hat{\rho}_k^2 \sim \chi^2_{m-(p+q)}$$

Ljung and Box (1978)

$$Q_{LB} = \sum_{1}^{m} \frac{n \cdot (n+2)}{(n-k)} \hat{\rho}_k^2 \sim \chi^2_{m-(p+q)}$$

identical for large sample.

Example: Li (2004), p11

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.4	.15	.07	.06	.09	.03	.05	.06	.5	.01

- $X_t = (1 - 0.4B)a_t$ was fitted to a series of 80 observations.

$$Q_{BP} = 80(.4^2 + .15^2 + \cdots + .01^2) = 16.696$$

$$Q_{LB} = 80(82)(.4^2 / 79 + .15^2 / 78 + \cdots + .01^2 / 70) = 17.488$$

- The upper *5%* critical value from the chi-squared distribution with *9*(=*10*-*1*) degrees of freedom is 16.92.

More about portmanteau tests

Pros:

- Practical purposes
- Minimal requirement for using the fitted model

Cons:

- Lack power if comparing with traditional statistical tests, such as likelihood ratio tests

Possible improvements and other applications:

- Finite sample adjustments
- Complicated functional of residual autocorrelations
- Monte Carlo test: See Lin&McLeod(2006)
- Other applications: portmanteau tests for randomness and ARMA models with infinite variance innovations

#Example 6.1 (Series W1)
Daily average number of truck manufacturing defects

AR(1) model

Model fitting in R

- > mod1<-**arima**(w1,c(1,0,0))

Coefficients:

	ar1	intercept
	0.4322	1.7799
s.e.	0.1340	0.1189

Should we put intercept here?

$$(1-0.4322\,B)(X_t-1.7799)=a_t$$

$$a_t \sim \text{NID}(0, 0.2118)$$

sigma^2 estimated as 0.2118:

log likelihood = -29.04, aic = 64.07

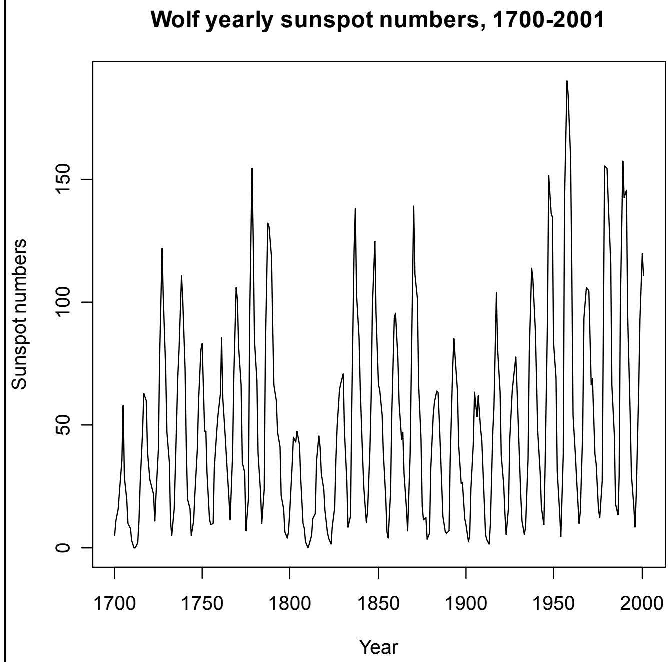
Diagnostic checks in R

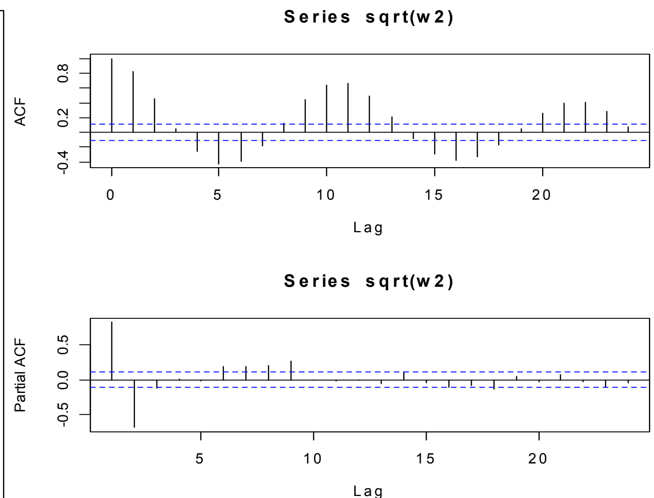
- **PRTest package in R**
- LBTest(mod1,lags=c(5,10,20))
- Lags: 5 10 20

  P-value: 0.7934879 0.8517180 0.9557307
- Large p-values—unable to reject H0

#Example 6.2 (Series W2)
Wolf yearly sunspot numbers, 1700-2001

Non stationary





AR(2) or AR(9) ??

AR(2) model for sunspot data

- > ar(sqrt(w2),**aic=F**, **order.max=2**)

Coefficients:

1	2
1.3833	-0.6796

$$(1-1.3833B+0.6796B^2)(\sqrt{X}_t-6.3476)=a_t$$

Order selected 2 sigma^2 estimated as 1.508

- > **mod2.2$x.mean**
- **[1] 6.43476**

- LBTest(arima(sqrt(w2),c(2,0,0)),c(5,10,15))

Lags:	5	10	15
P-values:	1.292324e-02	6.462316e-06	1.584744e-07

- Small p-values: reject H0

AR(9) model for sunspot data

- > LBTest(arima(sqrt(w2),c(9,0,0)),c(5,10,15))

Lags:	5	10	15
P-values:	**0.47793971**	**0.05895496**	**0.16600937**

smallest p-values greater than 5% so we won't H0

```
> arima(sqrt(w2),c(9,0,0))
Coefficients:
         ar1      ar2      ar3     ar4      ar5     ar6     ar7      ar8
      1.2050  -0.4733  -0.1282  0.2426  -0.2283  0.0171  0.1710  -0.2192
s.e.  0.0546   0.0873   0.0910  0.0917   0.0919  0.0923  0.0926   0.0887
         ar9  intercept
      0.3083     6.3511
s.e.  0.0552     0.5325

sigma^2 estimated as 1.061:  log likelihood = -439.65,  aic = 901.3
```

MODEL SELECTION

- In time series analysis, several models may adequately represent a given data set.
- How to select the best model among these candidates is called model selection or order selection.

Methods for model selection

Two of the most popular methods are Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC):

$$AIC=-2 \log ML+2 k, \qquad (a)$$

$$BIC=-2 \log ML+k \log(n), \qquad (b)$$

- where ML denotes maximum likelihood, log ML is the value of maximized log-likelihood function for a model fitted to a give data set, and k is the number of independently adjusted parameters within the model.

Methods for model selection

AIC and BIC

- A desirable attribute of AIC (and BIC) is that the modeling principles are formally incorporated into the equations.
- The first term on the R.H.S. of AIC and BIC reflects the doctrine of good statistical fit while the second entry accounts for model parsimony.

Decision criteria

- When there are several available models for modeling a given time series, the model that possesses the minimum value of the AIC (or BIC) should be selected.
- **Remark:** BIC puts more penalties on the number of parameters used by fitted models, and some empirical studies indicate that the model selected by BIC performs better in the post-sample analysis, such as forecasting.