ADVANCED ANALYTICS

Time Series (ARIMA)

Ms Catherine Khaw
Institute of Systems Science
National University of Singapore
Email: issv155@nus.edu.sg

Topics

- What is a time series
- Autogressive Models
- Moving Average Models
- Integrated models
- ARMA, ARIMA, SARIMA

What is a Time Series?

- The observations from a discrete time series, made at fixed interval h, at times $\tau_1, \tau_2, \ldots, \tau_N$ may be denoted by $x(\tau_1), x(\tau_2), \ldots, x(\tau_N)$

- Discrete time series may arise in two ways:
  - 1- By sampling a continuous time series
  - 2- By accumulating a variable over a period of time

- Characteristics of time series
  - Time periods are of **equal length**
  - No missing values

Example of a Time Series

- Daily sales
- Hour clicks on website
- Daily stock price
- Daily produces by a factory line
- ….

Goal of Time Series Forecasting

Class discussion

Time Series application

- For a long time there has been very little communication between econometricians and time-series analysts.

- Theories were imposed on the data even when the temporal structure of the data was not in conformity with the theories.

Time Series application (Cont.)

- Econometricians have emphasized economic theory and a study of contemporaneous relationships. Lagged variables were introduced in unsystematic way, and no serious attempts were made to study the temporal structure of the data

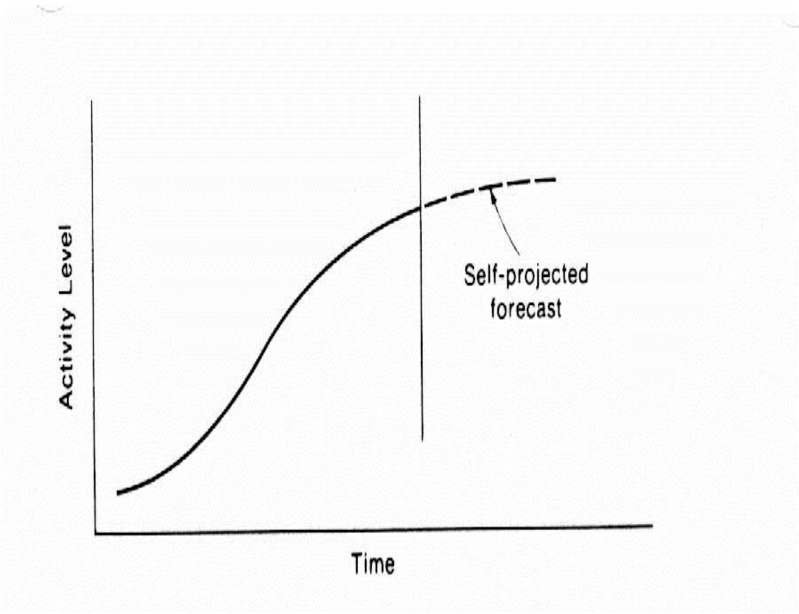
- The time-series analysts, on the other hand, did not believe in economic theories and thought that they were better off allowing the data to determine the model

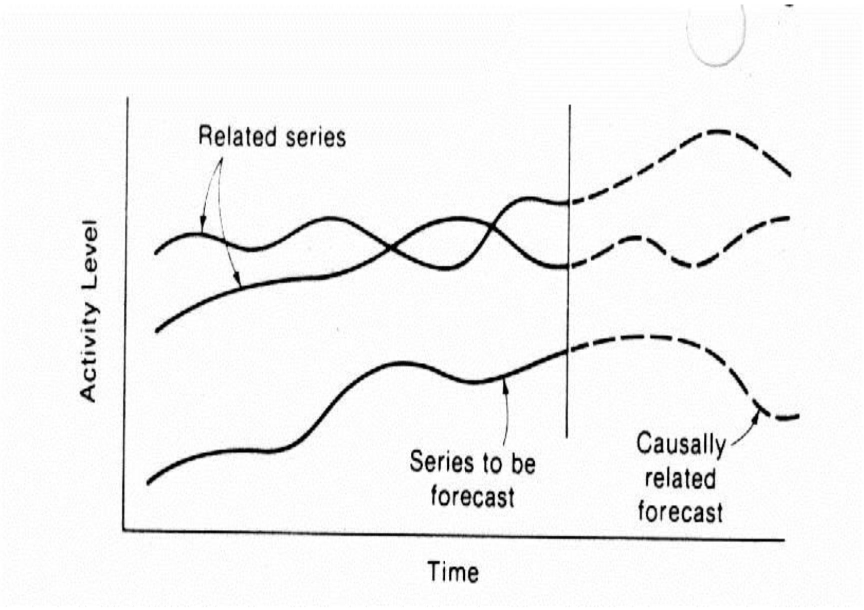
- Since the mid-1970s these two approaches—the time-series approach and the econometric approach—have been converging

- Econometricians now use some of the basic elements of time-series analysis in checking the specification of their econometric models, and some economic theories have influenced the direction of time-series work.

Time Series Approach

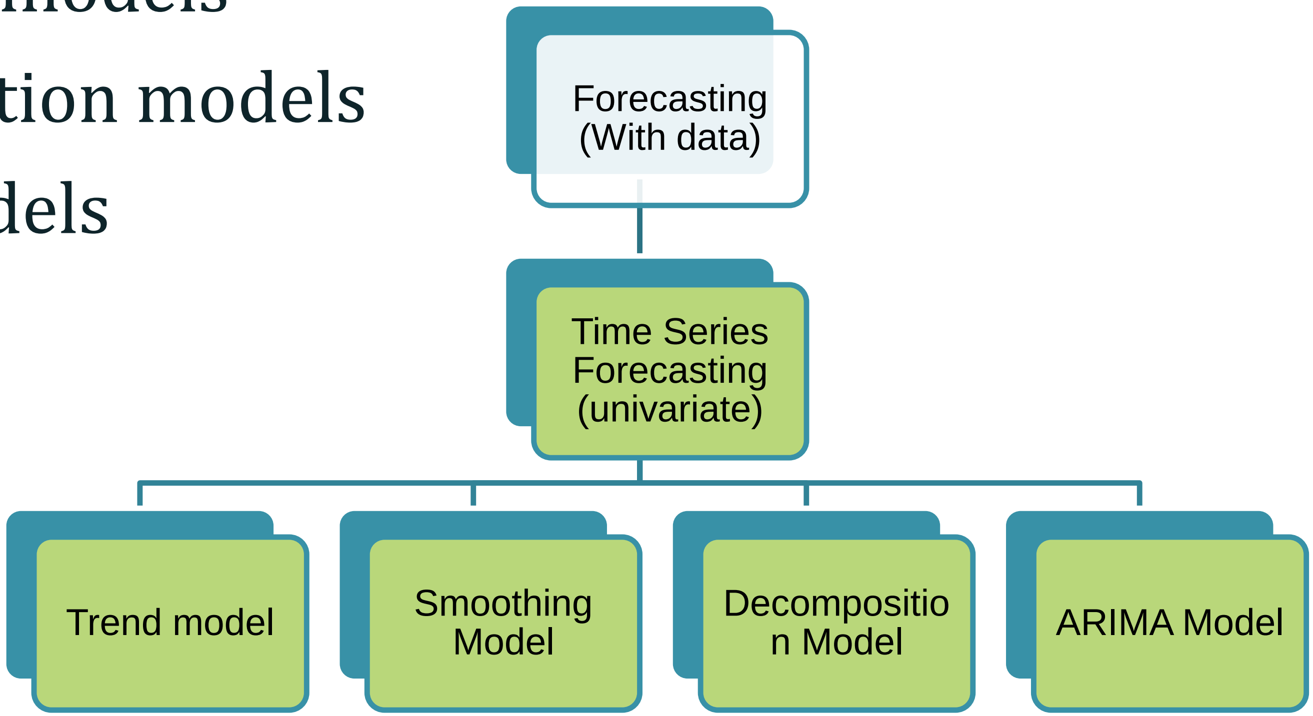
- Self-projecting approach (univariate)
- Cause-and-effect approach (multivariate)





Common Self Projecting Models

- Overall Trend models
- Smoothing models
- Decomposition models
- ARIMA models



AUTOREGRESSIVE MODELS

Autoregressive (AR) Models

- Recall in line regression, auto-correlation must not exist when applying SLR.

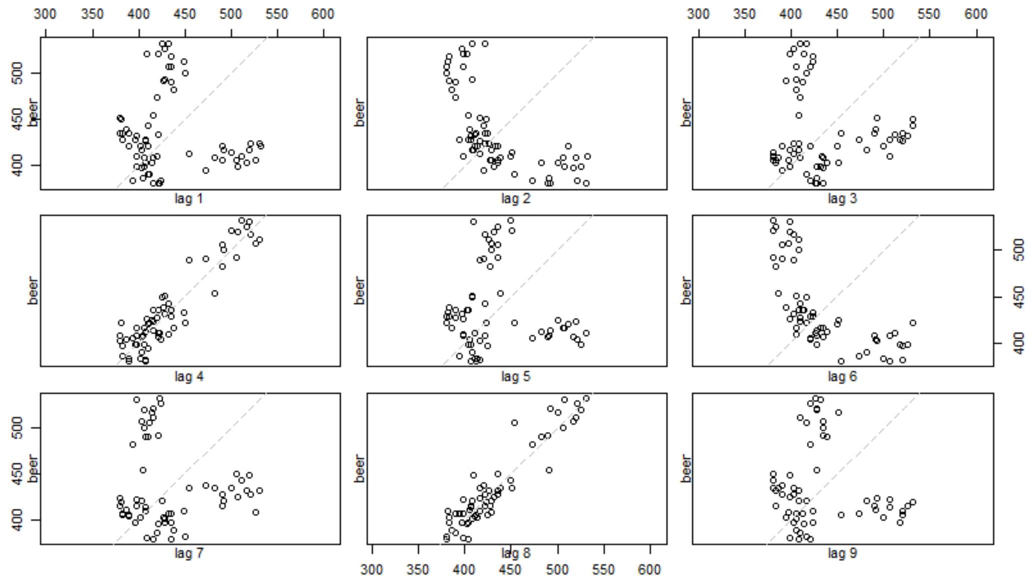
- In the case of AR Model, the target variable is express as a linear regression of its history.

$$\hat{x}_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + a_3 x_{t-3} + \dots .$$

Auto-correlation

Just as correlation (r) measures the extent of linear relationship between two variables; **Auto-correlation** measures the linear relationship between lagged values of a time series.

- For example :
  - $r_1 \sim$ relationship between x_t and x_{t-1}
  - $r_2 \sim$ relationship between x_t and x_{t-2}



AR Models (cont'd)

$$\hat{x}_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + a_3 x_{t-3} + \dots .$$

- Auto-Regressive ~ Regression on itself (x on x history)

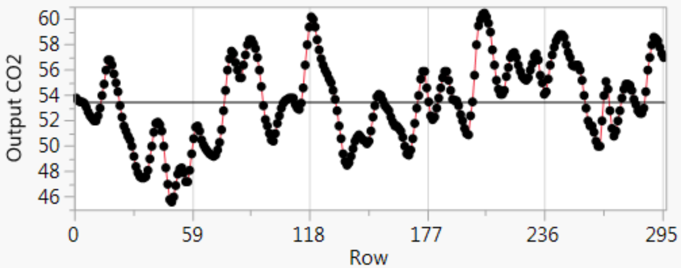
- Lets assume that x_t can be predicted base on x_{t-1}

- Error = e_t= actual - expected = $x_t - \hat{x}_t = x_t - a_0 - a_1 x_{t-1}$
- Full model : $x_t = a_0 + a_1 x_{t-1} + e_t$

- a_0 & a_1 are estimated base minimizing SSE

Example

Time Series Output CO2

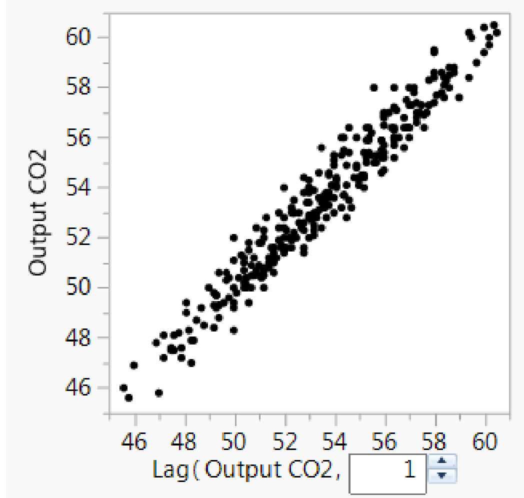


Mean	53.509122
Std	3.1967072
N	296
Zero Mean ADF	0.137862
Single Mean ADF	-1.863957
Trend ADF	-2.257128

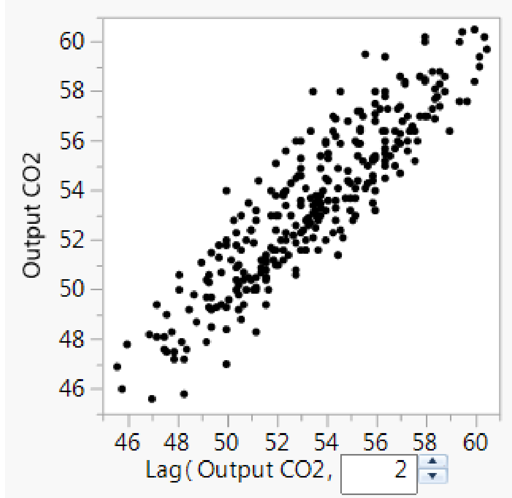
Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.9708	281.778	<.0001*	1	0.9708
2	0.8960	522.666	<.0001*	2	-0.8039
3	0.7925	711.766	<.0001*	3	0.1883
4	0.6800	851.434	<.0001*	4	0.2600
5	0.5745	951.471	<.0001*	5	0.0595
6	0.4854	1023.15	<.0001*	6	-0.0626
7	0.4161	1075.99	<.0001*	7	-0.0144
8	0.3656	1116.93	<.0001*	8	0.0549
9	0.3304	1150.48	<.0001*	9	0.0055
10	0.3065	1179.46	<.0001*	10	0.0314
11	0.2880	1205.12	<.0001*	11	-0.1166
12	0.2693	1227.65	<.0001*	12	-0.0430
13	0.2473	1246.71	<.0001*	13	0.0511
14	0.2215	1262.05	<.0001*	14	0.0538
15	0.1930	1273.74	<.0001*	15	-0.0461
16	0.1649	1282.30	<.0001*	16	0.0339
17	0.1398	1288.48	<.0001*	17	-0.0033
18	0.1210	1293.12	<.0001*	18	0.0855
19	0.1103	1297.00	<.0001*	19	0.0165
20	0.1078	1300.72	<.0001*	20	-0.0258
21	0.1112	1304.68	<.0001*	21	-0.0492
22	0.1171	1309.10	<.0001*	22	0.0097
23	0.1228	1313.96	<.0001*	23	0.0491
24	0.1259	1319.10	<.0001*	24	-0.0012
25	0.1259	1324.26	<.0001*	25	0.0043

Lag Plot



Lag Plot



Example (Cont'd)

Model: AR(3)

Model Summary

DF	292	Stable	Yes
Sum of Squared Errors	34.2952484	Invertible	Yes
Variance Estimate	0.11744948		
Standard Deviation	0.34270903		
Akaike's 'A' Information Criterion	216.272318		
Schwarz's Bayesian Criterion	231.033756		
RSquare	0.98864058		
RSquare Adj	0.98852387		
MAPE	0.46466563		
MAE	0.24796464		
-2LogLikelihood	208.272318		

Parameter Estimates

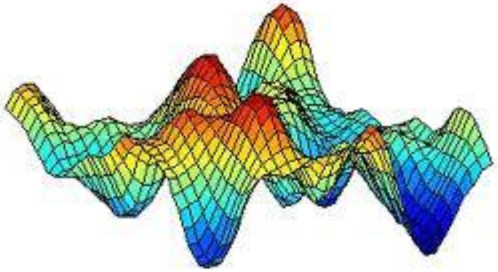
Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
AR1	1	2.19621	0.0512770	42.83	<.0001*	1.46625292
AR2	2	-1.68410	0.0963793	-17.47	<.0001*	
AR3	3	0.46054	0.0513752	8.96	<.0001*	
Intercept	0	53.61491	0.7062130	75.92	<.0001*	

Important Assumptions

- Linear relationship between successive values
- X_t is a Stationary process
- Errors are normal and independent

A Normal process (A Gaussian process)

- The Box-Jenkins methodology analyze a time series as a realization of a stochastic process.
  - The observation z_t at a given time t can be regarded as a realization of a random variable $\mathbf{z_t}$ with probability density function $\mathbf{p(z_t)}$
  - The observations at any two times t_1 and t_2 may be regarded as realizations of two random variables z_{t_1}, z_{t_2} and with joint probability density function $p(z_{t_1}, z_{t_2})$
  - If the probability distribution associated with any set of times is multivariate Normal distribution, the process is called a **normal or Gaussian process**



Stochastic Process

- A stochastic process is a collection $\{X_t : t = 1, 2, \ldots, T\}$ of random variables ordered in time. Example : the error term in a linear regression model is assumed to be a stochastic process.
- A stochastic process is weakly stationary if for all t values

$$\begin{aligned} E[X_t] &= \mu \\ \text{var}(X_t) &= \sigma^2 \\ \text{cov}(X_t X_{t-k}) &= \gamma_k \qquad \forall \qquad t \end{aligned}$$

i.e. its statistical properties do not change over time.

Stationary stochastic processes

- In order to model a time series with the Box-Jenkins approach, the series **has to be stationary**

- If an AR model is not stationary, this implies that previous values of the error term will have a non-declining effect on the current value of the dependent variable.

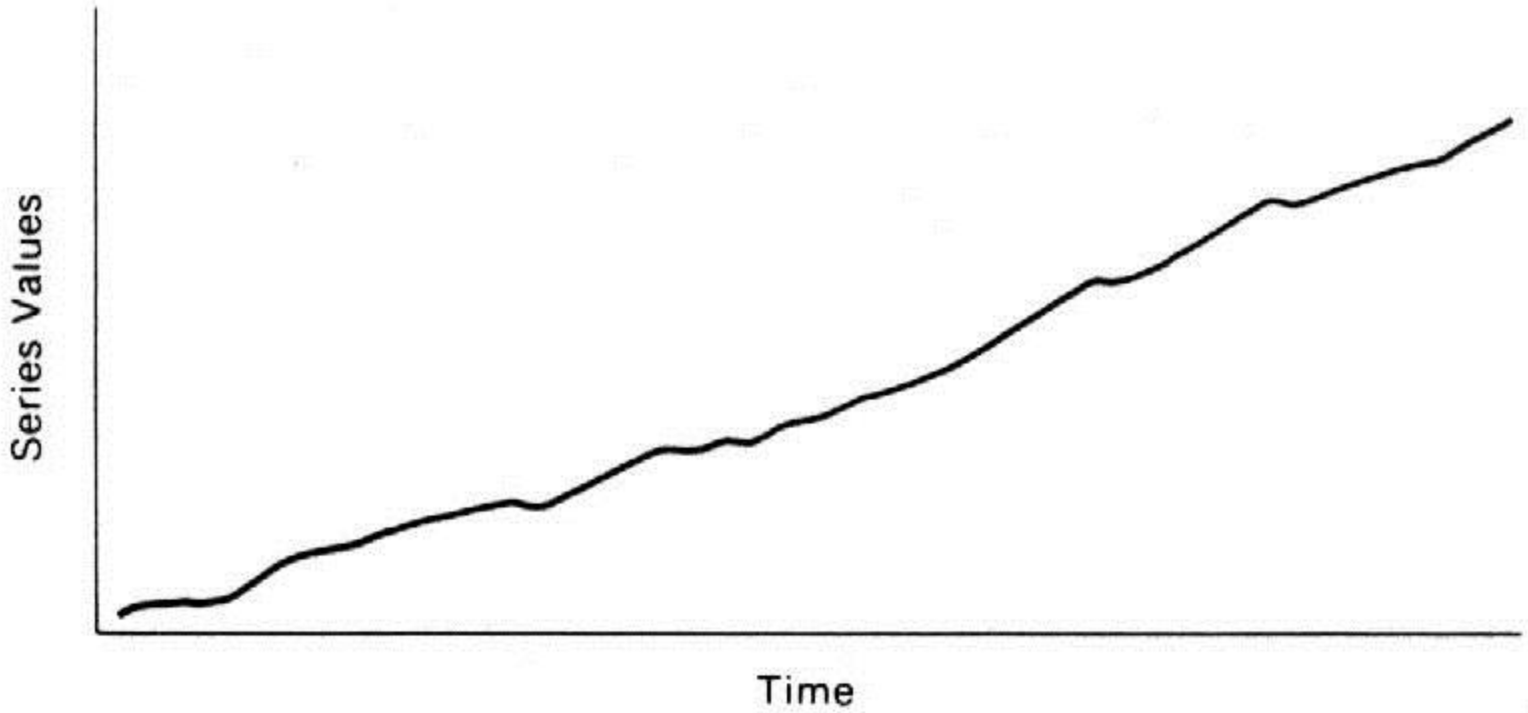
- This implies that the coefficients on the MA process would not converge to zero as the lag length increases.

- For an AR model to be stationary, the coefficients on the corresponding MA process decline with lag length, **converging on 0.**

Stationary stochastic processes (cont.)

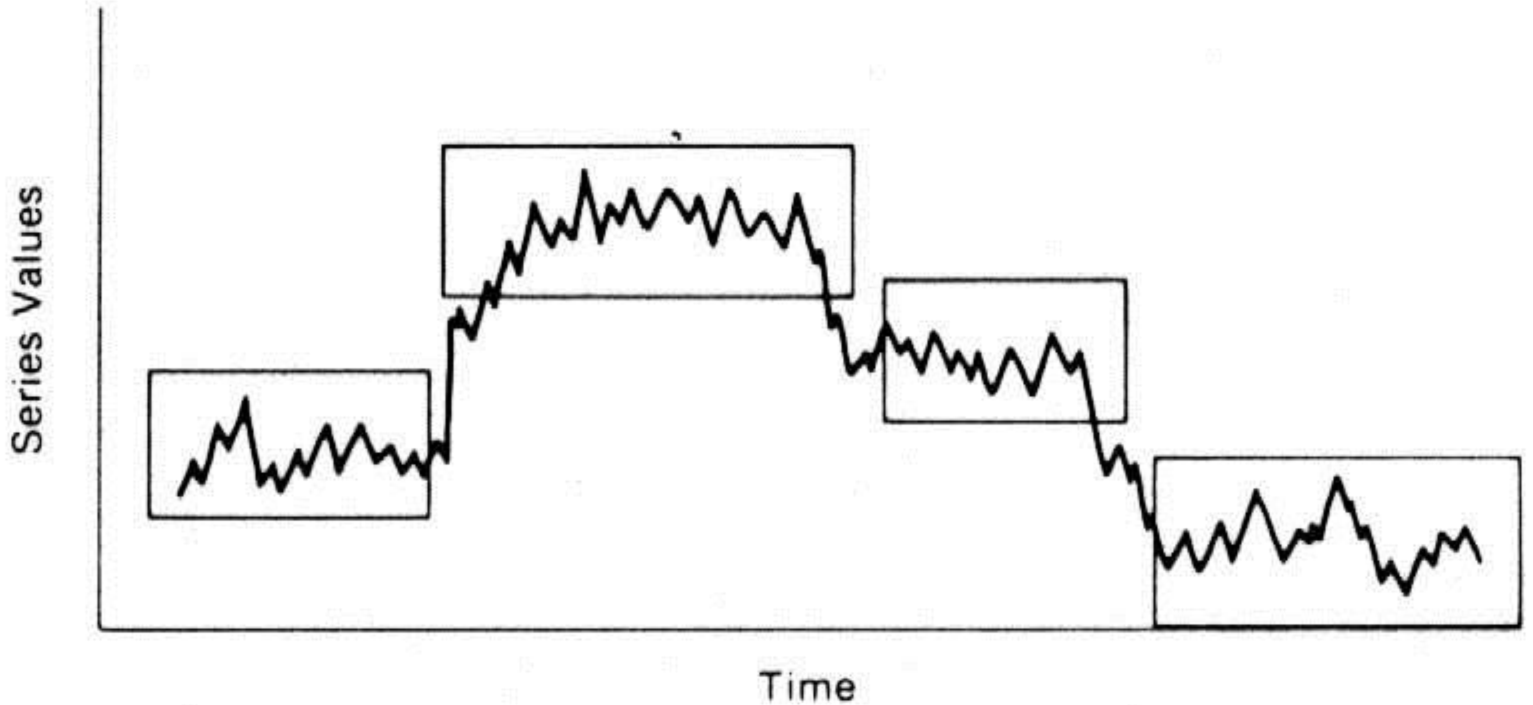
- a stationary series is one for which **the mean** and **variance** are **constant across time** and the covariance between current and lagged values of the series (autocovariances) depends only on the distance between the time points.
- Most time series data are nonstationary

Some nonstationary series



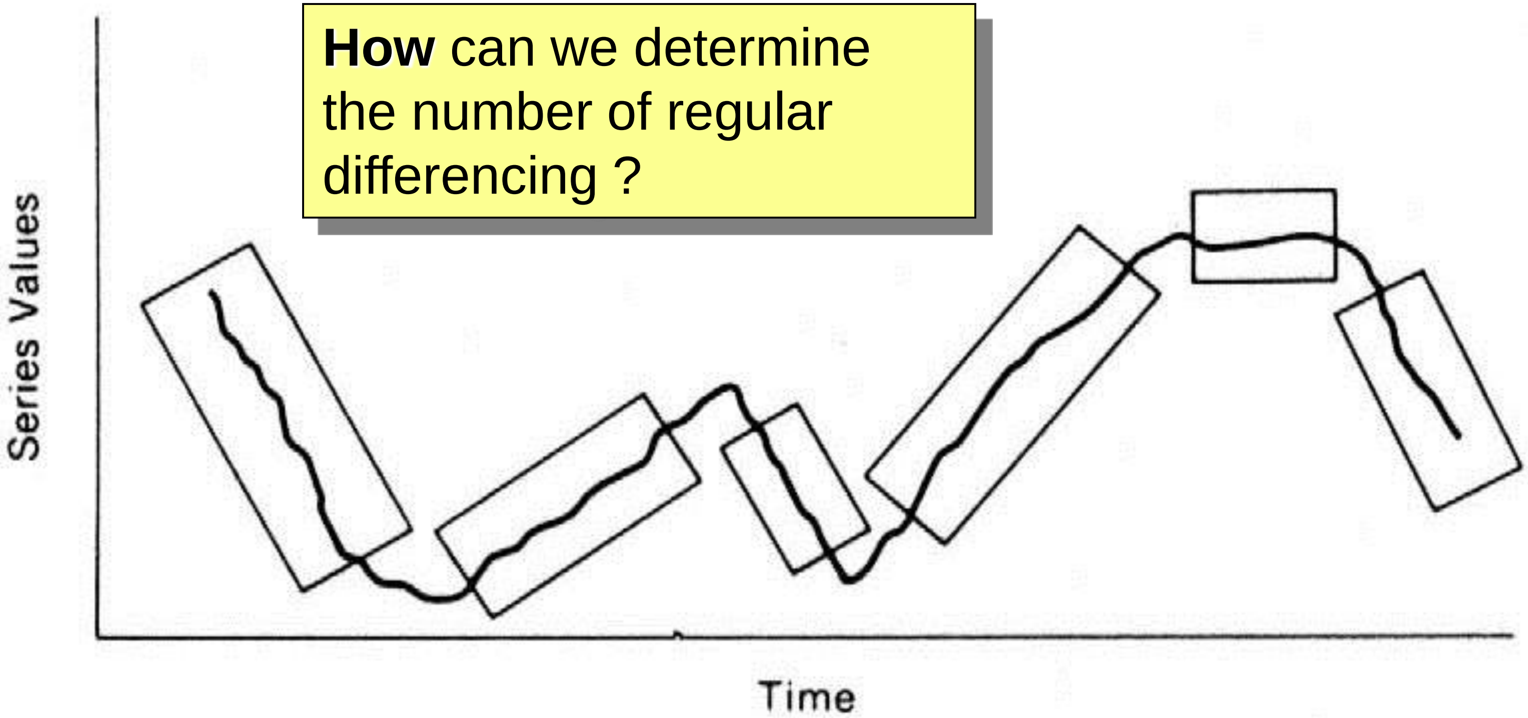
A Nonstationary Series: Overall Trend

Some nonstationary series (cont.)



A Nonstationary Series: Random Changes in Level

Some nonstationary series (cont.)



A Nonstationary Series: Random Changes in Both Level and Slope

Achieving stationarity - Differencing

- Regular differencing (RD)

$$(1^{st}\ \text{order}) \qquad \nabla y_t = y_t - y_{t-1}$$

$$(2^{nd}\ \text{order}) \qquad \nabla^2 y_t = y_t - 2y_{t-1} + y_{t-2}$$

- It is unlikely that more than two regular differencing would ever be needed
- Sometimes regular differencing by itself **is not** sufficient and **prior transformation** is also needed

Backshift Operator (B)

The backshift operator B is useful when working with time series model expression.

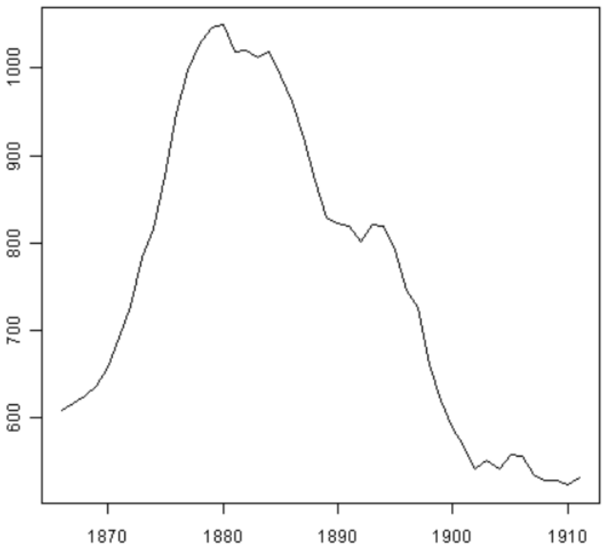
$$By_t = y_{t-1}$$
$$B(By_t) = B^2 y_t = y_{t-2}$$

Hence, 1st order differencing can be written as

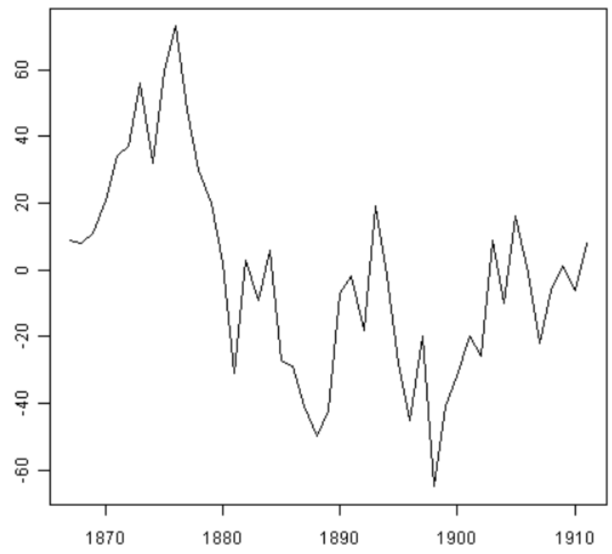
$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t$$

& 2nd order differencing : $(1 - B)^2 y_t$

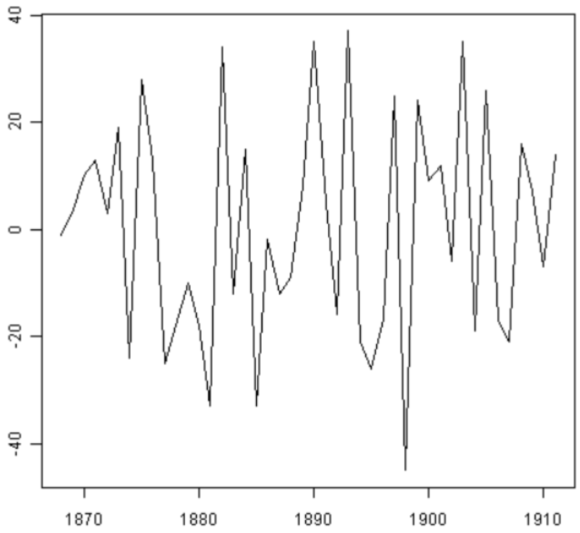
What happen to the series after differencing?



y_t



$y'_t = y_t - y_{t-1} = y_t - By_t = (1- B)y_t$

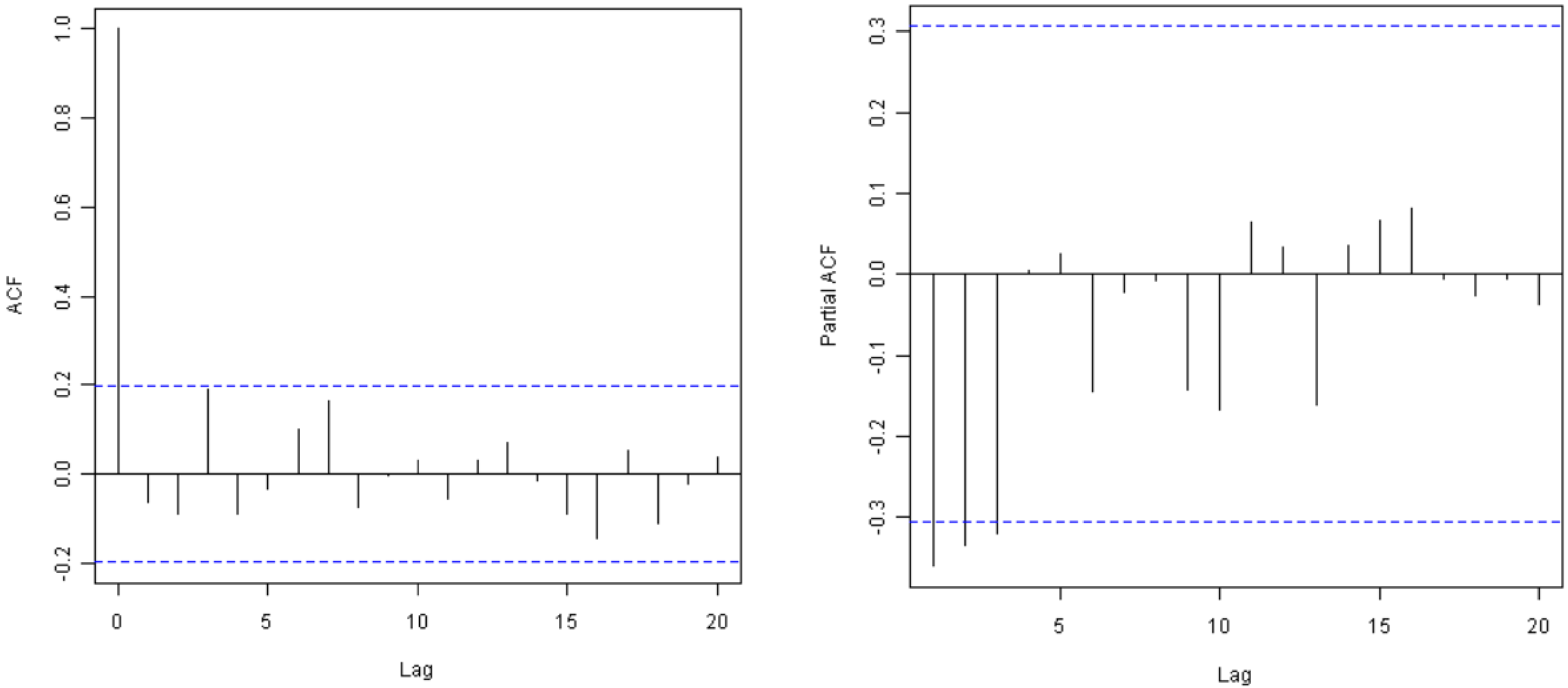


$y''_t = (1- B)^2 y_t$

Note: Assessing Stationarity

- A time series plot is often enough to convince a forecaster that the data are stationary or non-stationary. The **ACF and PACF** plots can readily expose non-stationarity in the mean.
  - The autocorrelations of stationary data drop to zero relatively quickly.
  - For a typical pattern of non-stationary series
    - ρ_1 is very large and positive (ACF plot).
    - ρ_k's are relatively large and positive, until k gets big enough (ACF plot).
    - PACF plot displays a large spike close to 1 at lag 1.

ACF & PACF



ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
20
Lag
Partial ACF
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
5
10
15
20
Lag

Unit root tests

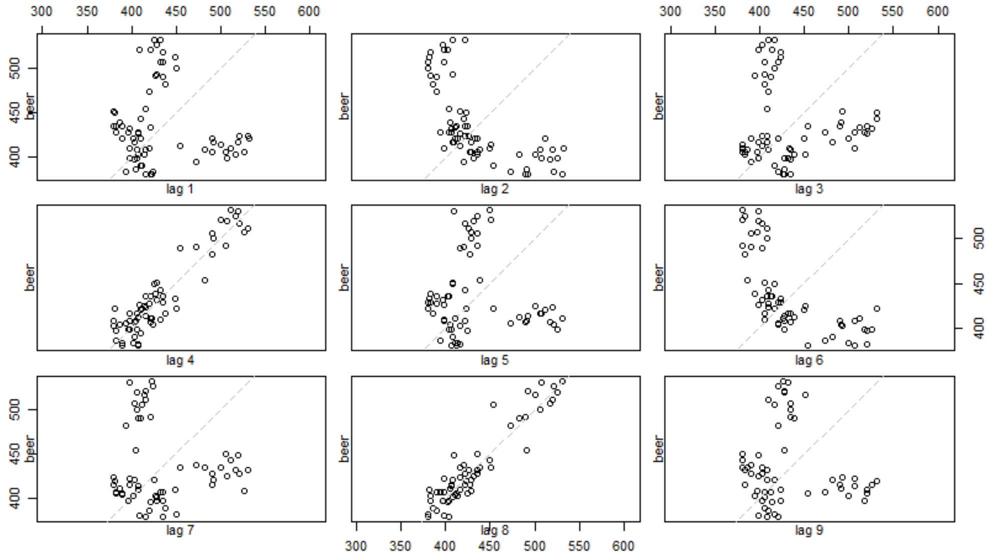
- More objective test of stationarity.
  - Augmented Dickey-Fuller (ADF) Test
    - Null hypothesis : data are non-stationary
  - Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test
    - Null hypothesis: data are stationary

Critical values for Dickey–Fuller *t*-distribution.

	Without trend		With trend	
Sample size	1%	5%	1%	5%
T = 25	−3.75	−3.00	−4.38	−3.60
T = 50	−3.58	−2.93	−4.15	−3.50
T = 100	−3.51	−2.89	−4.04	−3.45
T = 250	−3.46	−2.88	−3.99	−3.43
T = 500	−3.44	−2.87	−3.98	−3.42
T = ∞	−3.43	−2.86	−3.96	−3.41

Autocorrelations

- Covariances are often difficult to interpret because they depend on the units of measurement of the data. Correlations can be obtained through computing the autocorrelations of a time series.
- Autocorrelations are statistical measures that indicate how a time series is related to itself over time
- The autocorrelation at lag 1 is the correlation between the original series z_t and the same series moved forward one period (represented as z_{t-1})



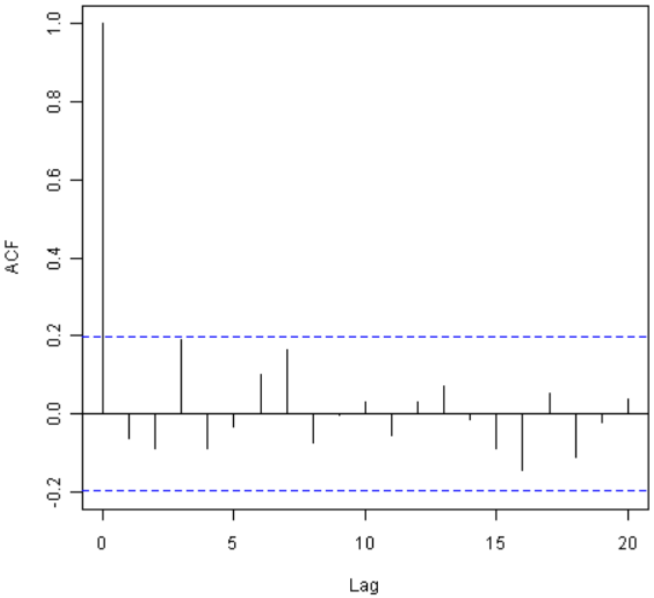
Autocorrelations (cont.)

- The theoretical autocorrelation function

$$\rho_k = \frac{E\left[(z_t - \mu)(z_{t+k} - \mu)\right]}{\sigma_z^2} = \frac{\operatorname{cov}\left[z_t z_{t+k}\right]}{\operatorname{var}(z_t)}$$

- The sample autocorrelation

$$r_k = \frac{\sum_{t=1}^{N-k}(z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^{N}(z_t - \bar{z})^2} \qquad k = 0,1,2,\ldots k$$



Autocorrelations (cont.)

- In practice, to obtain a useful estimate of the autocorrelation function, at least **50** observations are needed

- The estimated autocorrelations r_k would be calculated up to lag no larger than **N/4**

Partial Autocorrelations

- Another important function in the Box-Jenkins methodology is the partial autocorrelation function.

- It measures the strength of the relationship between observations in a series controlling for the effect of the intervening time periods.

Partial-autocorrelations (PACs)

- **Partial-autocorrelations** are another set of statistical measures used to identify time series models

- PAC is Similar to AC, **except** that when calculating it, the ACs with all the elements within the lag are **partialled out** (Box & Jenkins, 1976)

> *Partial autocorrelations are used to measure the degree of association between Y_t and Y_{t-k}, when the effects of other time lags (1, 2, 3, …, k – 1) are removed.*

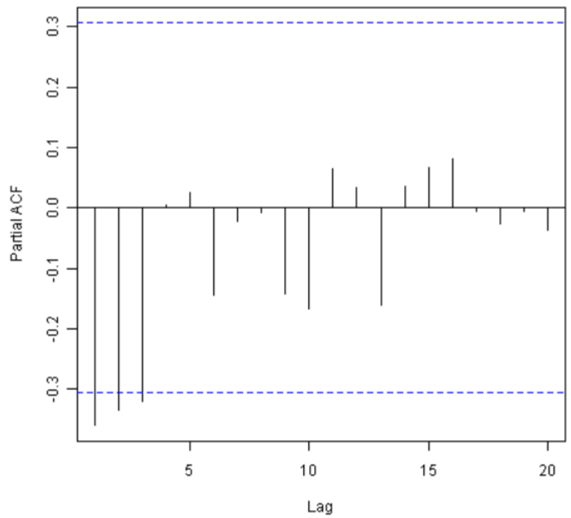
$$\frac{\text{Covariance}(y, x_3 | x_1, x_2)}{\sqrt{\text{Variance}(y | x_1, x_2)\text{Variance}(x_3 | x_1, x_2)}}$$

correlate the "parts" of *y* and x_3 that are not predicted by x_1 and x_2.

Partial-autocorrelations (cont.)

- PACs can be calculated from the values of the ACs where each PAC is obtained from a different set of linear equations that describe a pure autoregressive model of an order that is equal to the value of the lag of the partial-autocorrelation computed

- PAC at lag k is denoted by ϕ_{kk}
  - The double notation kk is to emphasize that ϕ_{kk} is the autoregressive parameter ϕ_k of the autoregressive model of order k



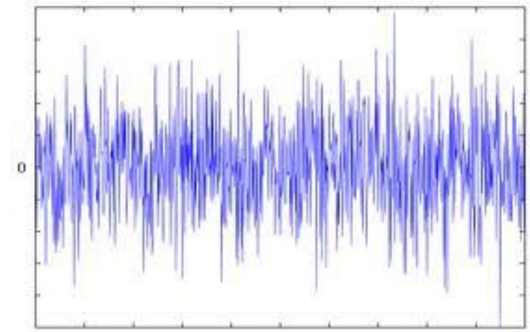
The white noise process

- The Box-Jenkins models are based on the idea that a time series can be usefully regarded as generated from (driven by) a series of **uncorrelated independent** "**shocks**" $\boxed{e_t}$

$$E[e_t] = 0 \quad \mathrm{var}[e_t] = \sigma_e^2$$

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

- Such a sequence e_t, e_{t-1}, e_{t-2},… is called a "white noise process"



AR Model Building

Autoregressive Models (AR)

- Autoregressive model of order p (AR(p))

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

i.e., y_t depends on its p previous values

- Using backshift notation

$$\phi_p(B) y_t = \phi_0 + \varepsilon_t$$

where

$$\phi_p(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$$

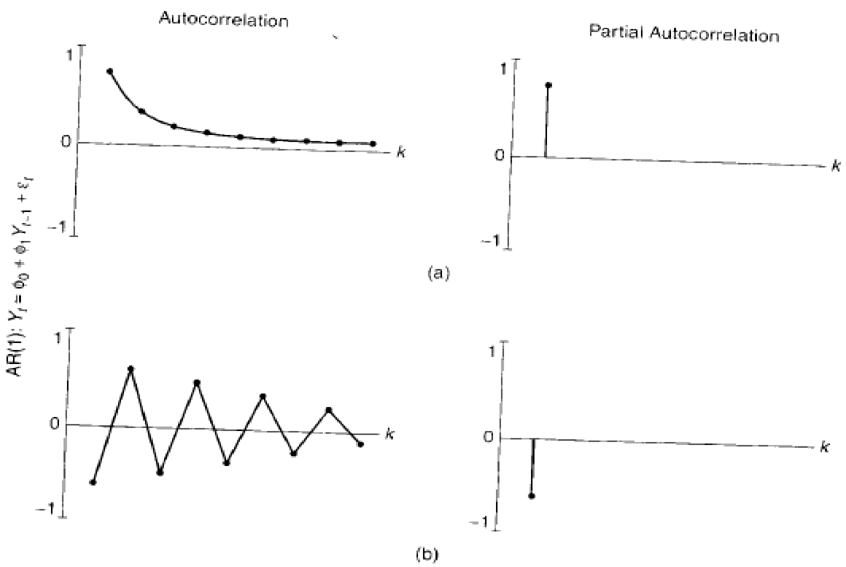
- Autoregression should be treated differently from ordinary regression models.
  - The basic assumption of independence of the error terms can easily be violated, since the explanatory variables usually have a built-in dependence relationship.
  - Determining the number of past values of y_t is not always straightforward.

An autoregressive model of order one AR(1)

- The basic form of an ARIMA (1, 0, 0) or AR(1) is:
  - Observation y_t depends on y_{t-1}.
  - The value of autoregressive coefficient ϕ_1 is between –1 and 1.

$$y_t = C + \phi_1 y_{t-1} + e_t$$

Theoretical ACF and PACF for AR(1)



- Characteristics:
  - ACF dies down
  - PACF cuts off after lag 1

AR(p)

- AR(2) process

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \qquad (\textit{where } \varepsilon_t \textit{ is white noise})$$

  where $|\phi_2| < 1$, $\phi_2 + \phi_1 < 1$, and $\phi_2 - \phi_1 < 1$, which are the stationarity requirement for an AR(2) process.

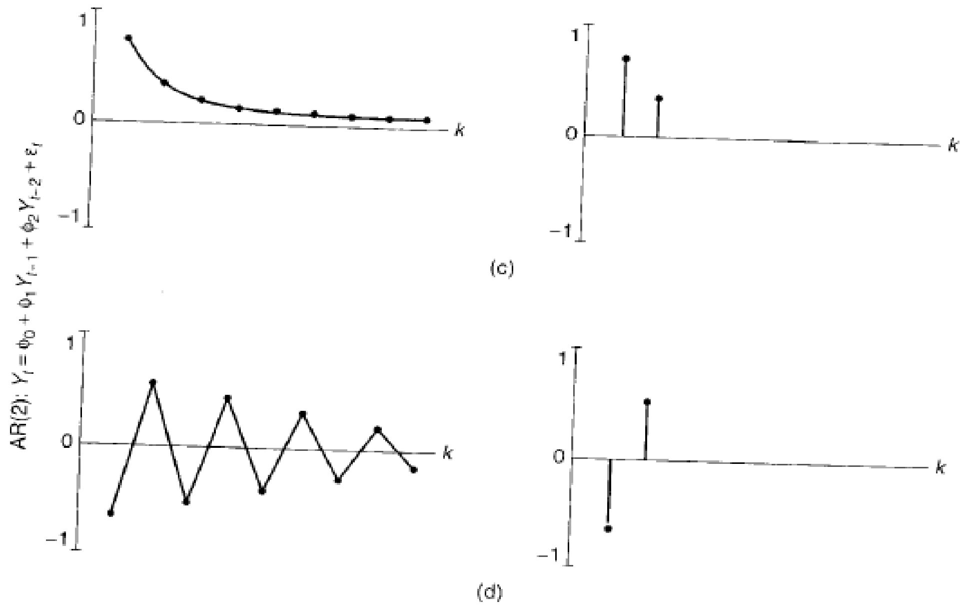
- AR(p) process

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

$$\textit{where } \varepsilon_t \textit{ is white noise}$$

  More complicated stationarity requirement of ϕ_i's holds for $p \geq 3$.

Theoretical ACF and PACF for AR(p)



- Characteristics:
  - ACF dies down.
  - PACF cuts off after lag p.

An autoregressive model of order one

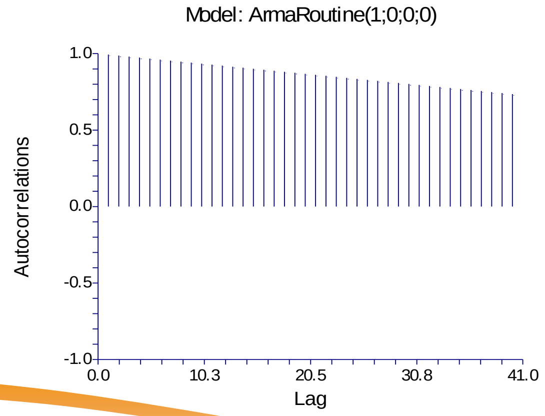
- The time plot of an AR(1) model varies with the parameter ϕ1..

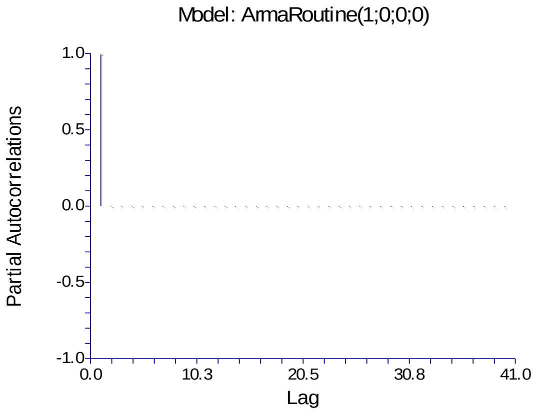
  When ϕ1= 0, yt is equivalent to a white noise series.

  When ϕ1= 1, yt is equivalent to a random walk series

  For negative values of ϕ1, the series tends to oscillate between positive and negative values.

- The following slides show the time series, ACF and PACF plot for an ARIMA(1, 0, 0) time series data.





Higher order auto regressive models

- A pth-order AR model is defined as

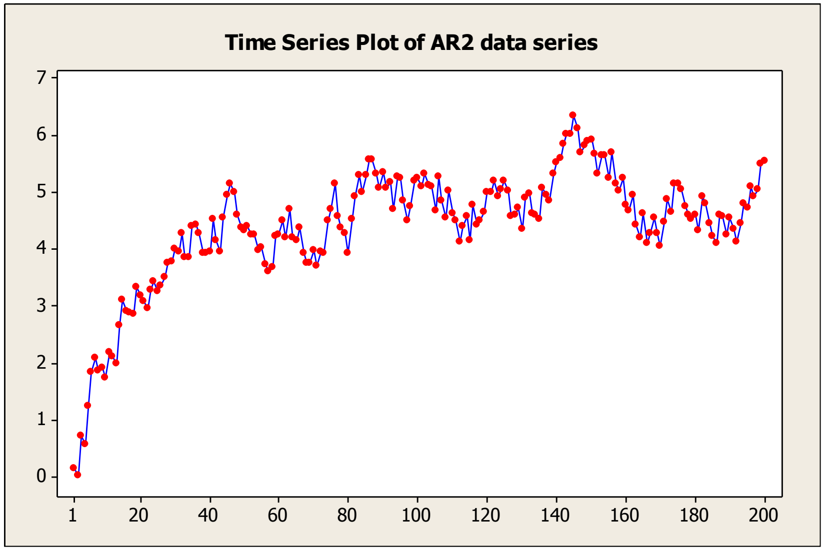
$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots \phi_p y_{t-p} + e_t$$

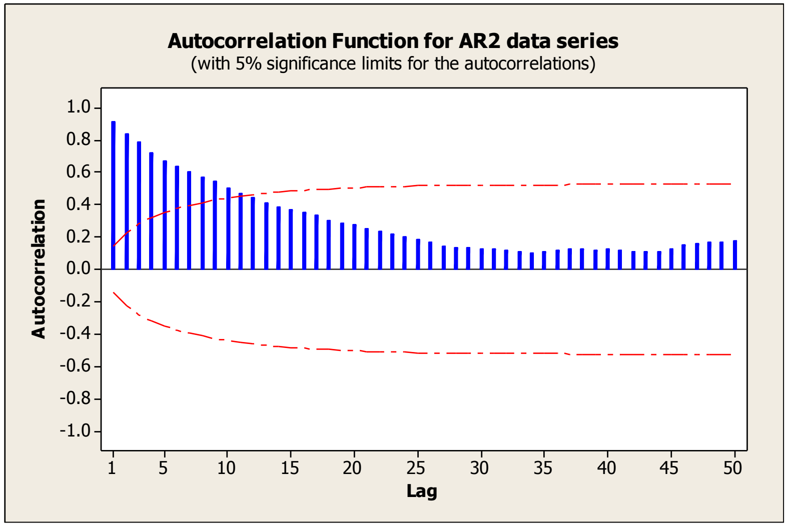
  - C is the constant term
  - ϕ_j is the jth auto regression parameter
  - e_t is the error term at time t.

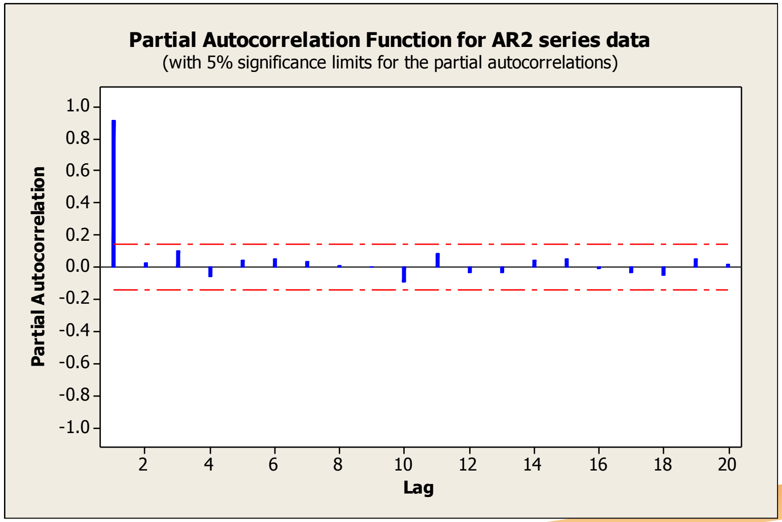
Higher order auto regressive models

- A great variety of time series are possible with autoregressive models.
- The following slides shows an AR(2) mode.
- Note that for AR(2) models the autocorrelations die out in a damped Sine-wave patterns.
- There are exactly two significant partial autocorrelations.

Higher order auto regressive models







Group discussion

- Is CO_2 series stationary?
- Build an AR model using CO_2 data.
- Hint: residuals check (is it white noise?)

MOVING AVERAGE MODELS (MA)

Moving Average Models (MA)

- Moving Average model of order *q* (MA(*q*))

  $$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

  i.e., y_t depends on q previous random error terms.
- Using backshift notation

  $$y_t = \mu + \theta_q(B)\varepsilon_t$$

  where $$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_q B^q$$
- The model is called a moving average because it is defined as a moving average of the error series, $\boxed{\varepsilon_t.}$
- Here we use *moving average* only in reference to a model of the above form.

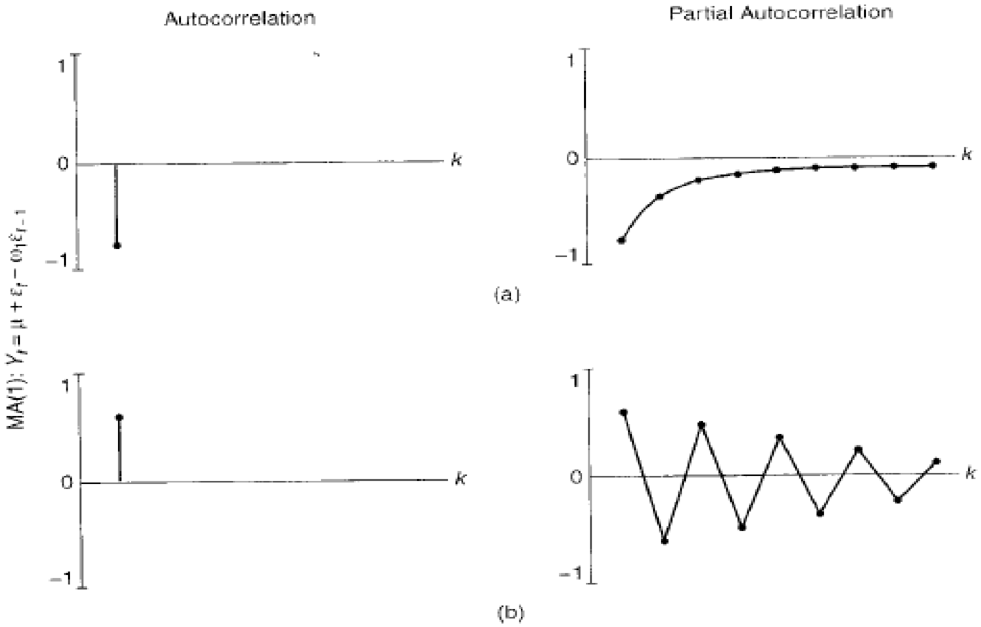
MA(1)

- MA(1) process:

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (\textit{where } \varepsilon_t \textit{ is white noise})$$

where $|\theta_1| < 1$, which is the stationarity requirement for an MA(1) process.

Theoretical ACF and PACF for MA(1)



- Characteristics:
  - ACF cuts off after lag 1
  - PACF dies down

A moving average of order one MA(1)

- Note that there is only one significant autocorrelation at time lag 1.
- The partial autocorrelations decay exponentially, but because of random error components, they do not die out to zero as do the theoretical autocorrelation.

MA(q)

- MA(2) process

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \quad (\textit{where } \varepsilon_t \textit{ is white noise})$$

  where $|\theta_2| < 1$, $\theta_2 + \theta_1 < 1$, and $\theta_2 - \theta_1 < 1$, which is the stationarity requirement for an MA(2) process.

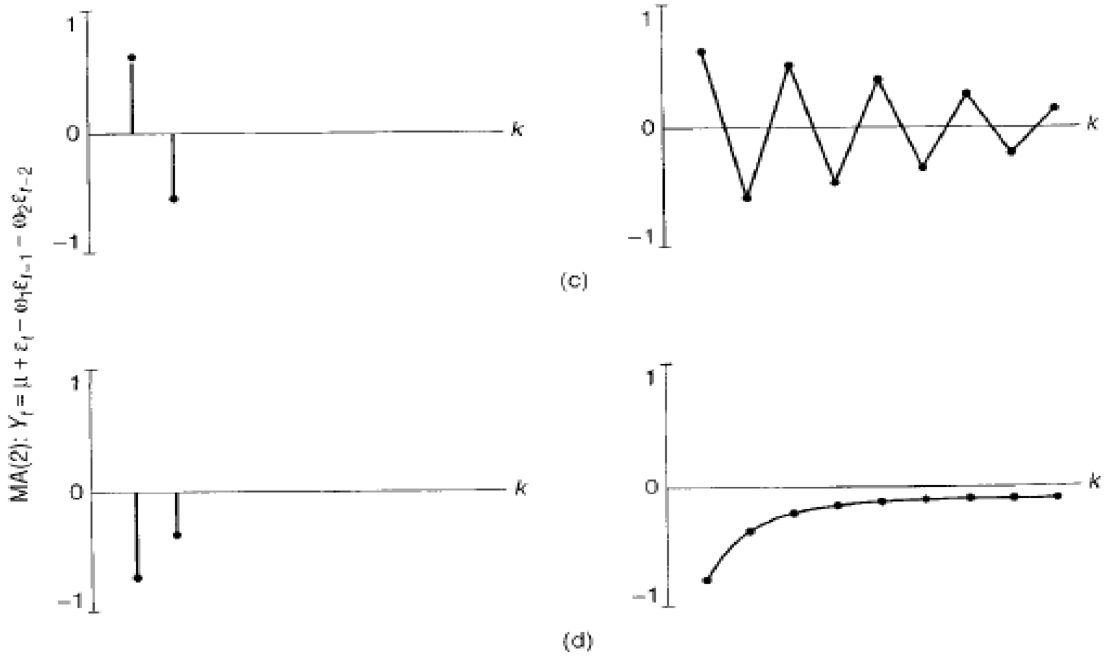
- MA(q) process

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

$$\textit{where } \varepsilon_t \textit{ is white noise}$$

  More complicated stationarity requirement of θ_i's holds for $q \geq 3$.

Theoretical ACF and PACF for MA(q)

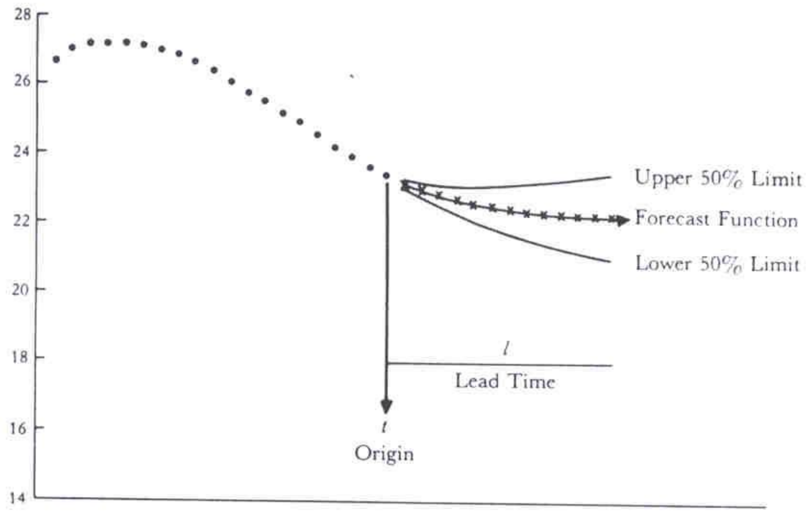


- Characteristics:
  - ACF cuts off after lag q.
  - PACF dies down.

ARMA, ARIMA, SARIMA

ARIMA models

- **A**utoregressive **I**ntegrated **M**oving-**a**verage
- Can represent a wide range of time series
- A "**stochastic**" modeling approach that can be used to calculate the probability of a future value lying between two specified limits



ARIMA models (Cont.)

- In the 1960's Box and Jenkins recognized the importance of these models in the area of economic forecasting
- "Time series analysis - forecasting and control"
  - George E. P. Box   Gwilym M. Jenkins
  - 1st edition was published in 1976

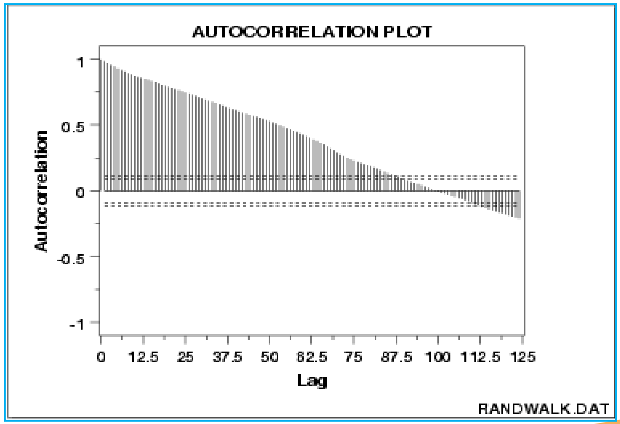
George E. P. Box

ARIMA models (Cont.)

- ARIMA models rely heavily on **autocorrelation** patterns in the data.

- ARIMA methodology of forecasting is different from most methods because it does not assume any particular pattern in the historical data of the series to be forecast.

- It uses an interactive approach of identifying a possible model from a general class of models. The chosen model is then checked against the historical data to see if it accurately describe the series.



Box-Jenkins Methodology

- Identification ~Determine, given a sample of time series observations, what is the model of the [stationary] data.
- Estimation ~Estimate the parameters of the chosen model

$$AR(p): X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t; \qquad \varepsilon_t \overset{i.i.d.}{\sim} N\left(\mu, \sigma^2\right)$$

$$MA(q): X_t = \alpha + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}; \qquad \varepsilon_t \overset{i.i.d.}{\sim} N\left(\mu, \sigma^2\right)$$

$$\begin{aligned} ARMA(p,q): X_t = {} & \alpha + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t \\ & - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}; \end{aligned} \qquad \varepsilon_t \overset{i.i.d.}{\sim} N\left(\mu, \sigma^2\right)$$

Autoregressive Moving Average Models (ARMA)

- Autoregressive-moving average model of order *p* and *q* (ARMA(*p*,*q*))

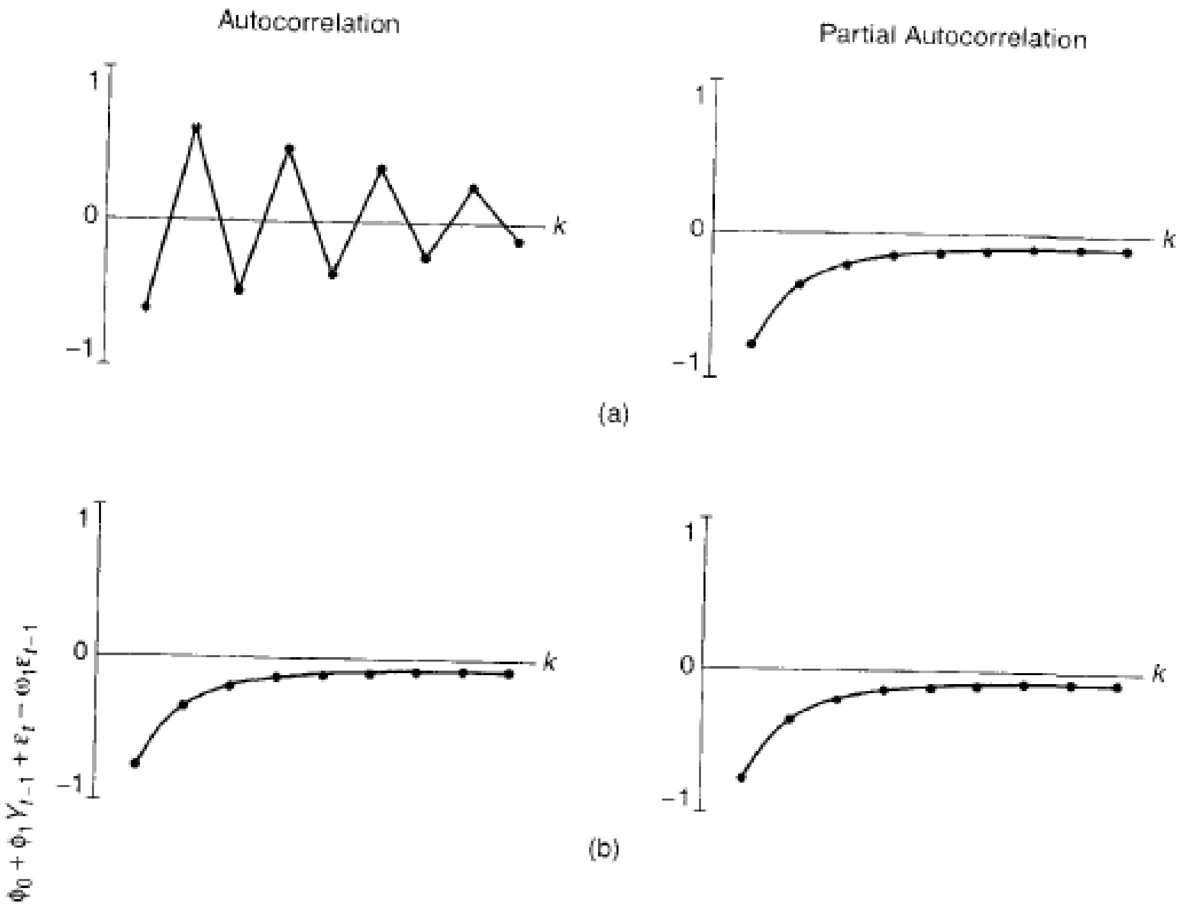
$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

  i.e., y_t depends on its *p* previous values and *q* previous random error terms
- Using backshift notation

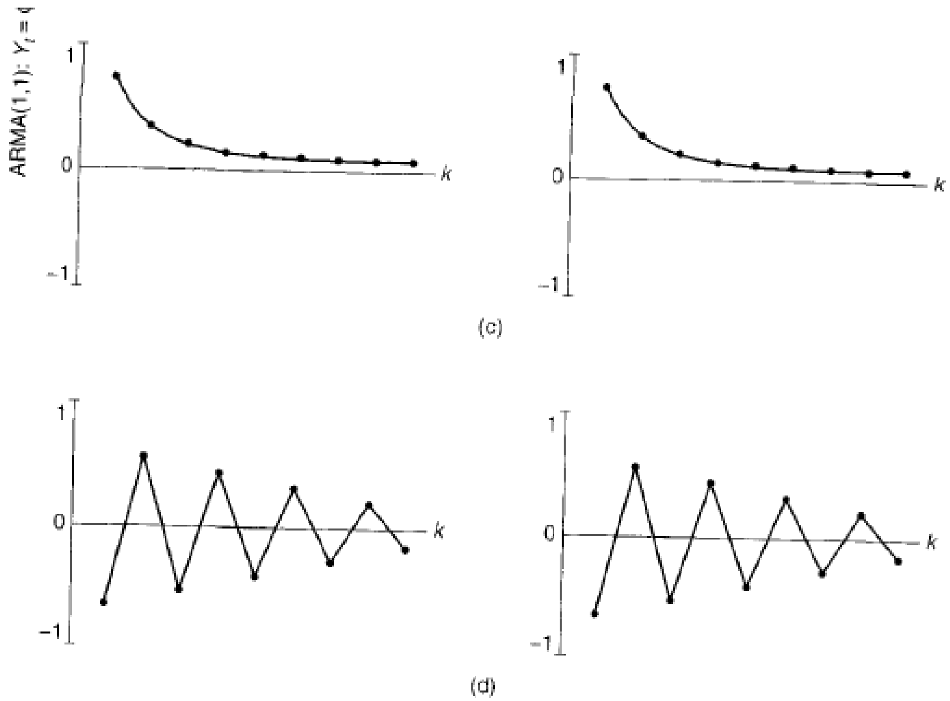
$$\phi_p(B) y_t = \phi_0 + \theta_q(B) \varepsilon_t$$

- When *q* = 0, the ARMA(*p*,0) model reduces to AR(*p*); when *p* = 0, the ARMA(0,*q*) model reduces to MA(*q*).

Theoretical ACF and PACF for ARMA(p,q)



Theoretical ACF and PACF for ARMA(*p*,*q*)



- Characteristics:
  - Both ACF and PACF die down.

Summary of the Behaviors of ACF and PACF

Behaviors of ACF and PACF for general non-seasonal models

Process	ACF	PACF
AR(p)	Dies down.	Cuts off after lag p.
MA(q)	Cuts off after lag q.	Dies down.
ARMA(p,q)	Dies down.	Dies down.

Non-seasonal Autoregressive Integrated Moving Average models (ARIMA)

- ARMA models can only be used for stationary data. This class of models can be extended to non-stationary series by allowing differencing the data series. $\Rightarrow$ ARIMA models
- Backshift notation

$$\underbrace{\phi_p(B)}_{AR}\underbrace{(1-B)^d}_{I} y_t = \phi_0 + \underbrace{\theta_q(B)\varepsilon_t}_{MA}$$

and $\delta = \mu\phi_p(B)\Phi_P(B^L)$

e.g. ARIMA(1,1,1) $\quad (1-\phi_1 B)(1-B)y_t = \phi_0 + (1-\theta_1 B)\varepsilon_t$

- The general non-seasonal model: ARIMA(p,d,q)
  - AR: p = order of the autoregressive part
  - I: d = order of integration
  - MA: q = order of the moving average part

Mixtures ARIMA models

- If non-stationarity is added to a mixed ARMA model, then the general ARIMA (p, d, q) is obtained.
- The equation for the simplest ARIMA (1, 1, 1) is given below.

$$y_t = C + (1 + \phi_1) y_{t-1} - \phi_1 y_{t-2} + e_t - \theta_1 e_{t-1}$$

Mixtures ARIMA models

- The general ARIMA (p, d, q) model gives a tremendous variety of patterns in the ACF and PACF, so it is not practical to state rules for identifying general ARIMA models.
- In practice, it is seldom necessary to deal with values p, d, or q that are larger than 0, 1, or 2.
- It is remarkable that such a small range of values for p, d, or q can cover such a large range of practical forecasting situations.

Seasonality and ARIMA models

- The ARIMA models can be extended to handle seasonal components of a data series.
- The general shorthand notation is

  $$\text{ARIMA } (p, d, q)(P, D, Q)_s$$

  – Where s is the number of periods per season.

Seasonality and ARIMA models

- The general ARIMA(1,1,1)(1,1,1)$_4$ can be written as

$$y_t = (1+\phi_1)y_{t-1} - \phi_1 y_{t-2} + (1+\Phi_1)y_{t-4} - (1+\phi_1+\Phi_1+\phi_1\Phi_1)y_{t-6}$$
$$- \Phi_1 y_{t-8} + (\Phi_1+\phi_1\Phi_1)y_{t-9} - \phi_1\Phi_1 y_{t-10} + e_t - \theta_1 e_{t-1} - \Theta_1 e_{t-4} + \theta_1\Theta_1 e_{t-5}$$

- Once the coefficients ϕ_1, Φ_1, θ_1, and Θ_1 have been estimated from the data, the above equation can be used for forecasting.

Seasonality and ARIMA models

- The seasonal lags of the ACF and PACF plots show the seasonal parts of an AR or MA model.
- Examples:
  - Seasonal MA model:
    - $\text{ARIMA}(0,0,0)(0,0,1)_{12}$
      - will show a spike at lag 12 in the ACF but no other significant spikes.
      - The PACF will show exponential decay in the seasonal lags i.e. at lags 12, 24, 36,…

Seasonality and ARIMA models

Seasonal AR model:

- ARIMA(0,0,0)(1,0,0)$_{12}$
  - will show exponential decay in seasonal lags of the ACF.
  - Single significant spike at lag 12 in the PACF.

Diagnostic Checking

Diagnostic Checking

- Often it is not straightforward to determine a single model that most adequately represents the data generating process. The suggested tests include

  (1) residual analysis,

  (2) model selection criteria.

Residual Analysis

- The residuals left over after fitting the model should be white noise.
  - ACF and PACF of the residuals show no significant autocorrelations or partial autocorrelations.
  - Residual autocorrelations as a group should be consistent with those produced by random errors.
    - Portmanteau test

White Noise Series

- White noise series $\{\varepsilon_t\}$
  - ε's are iid (independent and identically distributed) random variables with finite mean and variance.

$$(1)\ E(\varepsilon_t) = c \qquad for\ all\ t.$$

$$(2)\ Var(\varepsilon_t) = b \qquad for\ all\ t.$$

$$(3)\ Cov(\varepsilon_t, \varepsilon_{t+s}) = 0 \qquad for\ all\ t,\ \ s \neq t.$$

Portmanteau Test

- Rather than study the autocorrelation values one at a time, an alternative approach is to consider a whole set of autocorrelation values all at one time, and test to see whether the set is significantly different from a zero set.
- Ljung-Box test

$$Q = n(n+2)\sum_{k=1}^{m}\frac{r_k^2(e)}{n-k} \sim \chi^2_{m-r}$$

where $r_k(e)$ = the residual autocorrelation at lag k

n = number of residuals

k = time lag

m = number of time lags to be tested

r = number of parameters estimated in the model

Residual Analysis

- If the portmanteau test is <u>insignificant</u>, the model is adequate.

- If the portmanteau test is significant, the model is inadequate. Then we need to go back and consider other ARIMA models.

- The pattern of significant spikes in the ACF and PACF of the residuals may suggest how the model can be improved.
  - Significant spikes at low lags suggest the nonseasonal AR or MA components of the model.

Model Selection Criteria

- Akaike Information Criterion (AIC)

  $$\text{AIC} = -2\ln(L) + 2k$$

- Schwartz Bayesian Criterion (SBC)

  $$\text{SBC} = -2\ln(L) + k\ln(n)$$

  where L = likelihood function

  k = number of parameters to be estimated,

  n = number of observations.

- Ideally, the AIC and SBC will be as small as possible.
- Usually the model with the smallest AIC and SBC will have residuals which resemble white noise.

Interpreting the results: AR

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

$$\phi_p(B) y_t = \phi_0 + \varepsilon_t$$

where $\phi_p(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$

- AR(1) $\quad Y_{(t)} = \varphi_0 + \varphi_1 BY_{(t)} + \varepsilon_{(t)}$

- AR(1,1) $\quad (1\text{-}B)Y_{(t)} = \varphi_0 + \varphi_1 B(1\text{-}B)Y_{(t)} + \varepsilon_{(t)}$

$$Y_{(t)} = \varphi_0 + BY_{(t)} + \varphi_1 BY_{(t)} - \varphi_1 B^2 Y_{(t)} + \varepsilon_{(t)}$$

$$Y_{(t)} = \varphi_0 + (1+\varphi_1)Y_{(t-1)} - \varphi_1 Y_{(t-2)} + \varepsilon_{(t)}$$

Interpreting the results: MA

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

$$y_t = \mu + \theta_q(B)\varepsilon_t$$

where $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_q B^q$

- MA(1) $Y_{(t)} = \mu + \varepsilon_{(t)} - \theta_1 B \varepsilon_{(t)}$
- MA (1,1) $(1\text{-}B)\, Y_{(t)} = \mu + \varepsilon_{(t)} - \theta_1 B \varepsilon_{(t)}$

  $Y_{(t)} = \mu + B\, Y_{(t)} + \varepsilon_{(t)} - \theta_1 \varepsilon_{(t-1)}$

Interpreting the results: ARIMA

ARMA $\phi_p(B)y_t = \phi_0 + \theta_q(B)\varepsilon_t$

ARIMA $\phi_p(B)(1-B)^d y_t = \phi_0 + \theta_q(B)\varepsilon_t$

- ARIMA(1,1,1) $(1-\phi_1 B)(1-B)y_t = \phi_0 + (1-\theta_1 B)\varepsilon_t$

  $(1\text{-}B - \varphi_1 B + \varphi_1 B^2)Y_t = \varphi_0 + \varepsilon_t - \theta_1 B\varepsilon_t$

  $Y_t = \varphi_0 + (1+\varphi_1)\, Y_{t-1} - \varphi_1\, Y_{t-2} + \varepsilon_t - \theta_1\varepsilon_{t-1}$

Interpreting the results: Seasonal ARIMA

$$\varphi_p B(1-B)^d \psi_P B^4 (1-B^4)^D \ Yt= \ (\theta_q B)(\Theta_Q B^4)\ \varepsilon_t$$

- ARIMA(1,1,1)(1,1,1)$_4$

$$(1-\varphi_1 B)(1-B)(1-\psi_1 B^4)(1-B^4)^{\,1}\, Y_t = \ (1-\theta_1 B)(1-\odot_1 B^4)\ \varepsilon_t$$

….

$$Y_t = \quad (1+\varphi_1)Y_{t-1} - \varphi_1 Y_{t-2} + (1+\psi_1)Y_{t-4} - (1+\varphi_1+\psi_1+\varphi_1\psi_1)Y_{t-6}$$

$$- \psi_1 Y_{t-8} + (\psi_1 + \varphi_1\psi_1)Y_{t-9} - \varphi_1\psi_1 Y_{t-10}$$

$$+ \varepsilon_t - \theta_1\varepsilon_{t-1} - \odot_1\varepsilon_{t-4} + \theta_1 \odot_1\varepsilon_{t-1}$$

Special Cases of ARIMA

White noise	ARIMA(0,0,0)
Random Walk	ARIMA(0,1,0) with no constant
Random Walk with drift	ARIMA(0,1,0) with constant
Autoregression	ARIMA(p,0,0)
Moving Average	ARIMA(0,0,q)

Process Steps to Consider

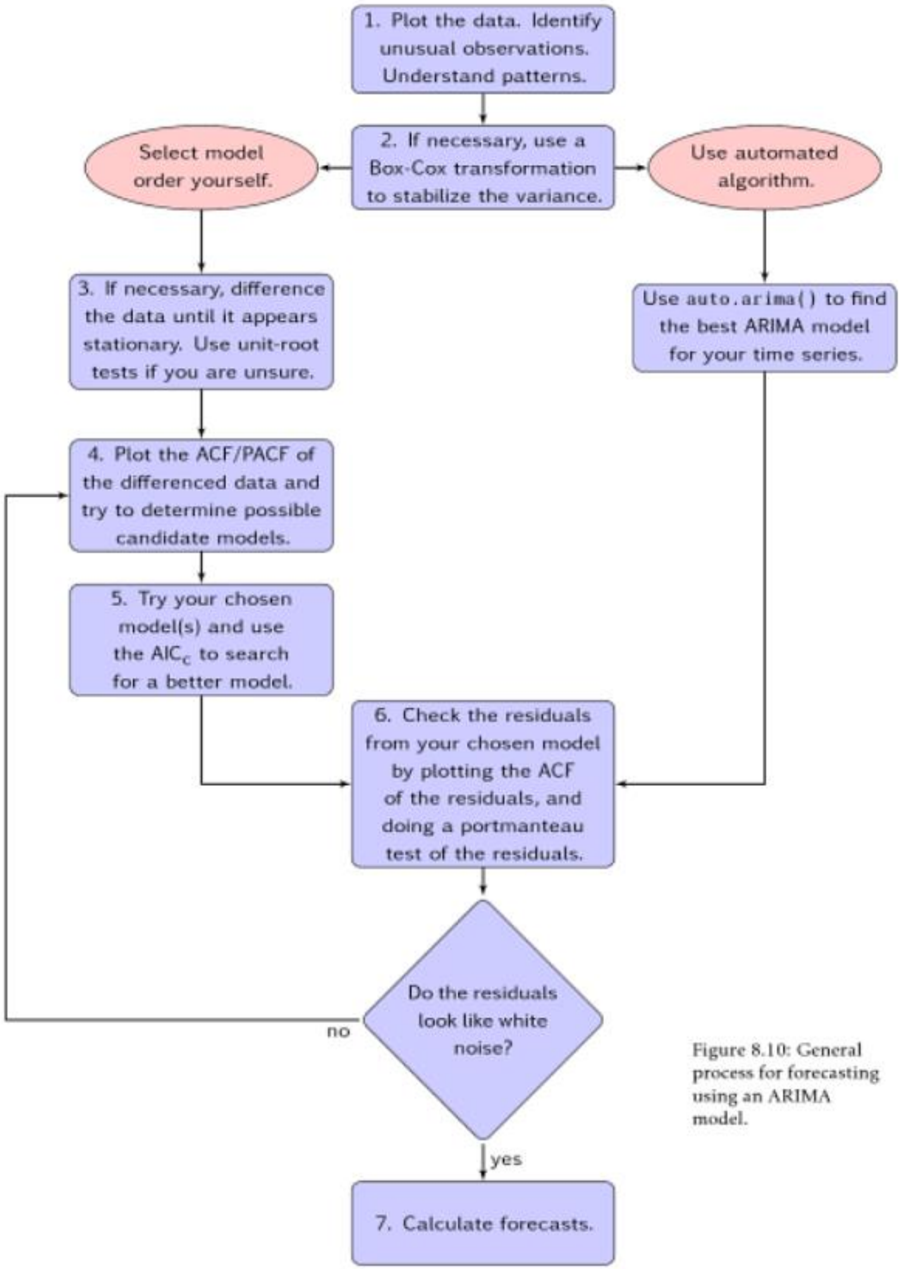


Figure 8.10: General process for forecasting using an ARIMA model.

Source : Forecasting Principles and practice – Rob J.H. & George A.

Group Exercise

- Objective Statement
- Source of data
- Observations & Analysis
- Model Building
- Recommendation

References

“Time Series Analysis and Forecasting by Example” – Soren Bisgaard and Murat Kulahci

Forecasting: principles and practice – Rob J Hyndman, George Athanasopoulos

Analysis of Financial Time Series - Ruey S. Tsay