Queen Mary School Hainan
Queen Mary University of London

QHP5701 Exploratory Data Analysis

Time Series Analysis

Nikesh Bajaj, PhD
Lecturer in Data Science,
Queen Mary University of London
nikesh.bajaj@qmul.ac.uk
https://nikeshbajaj.in

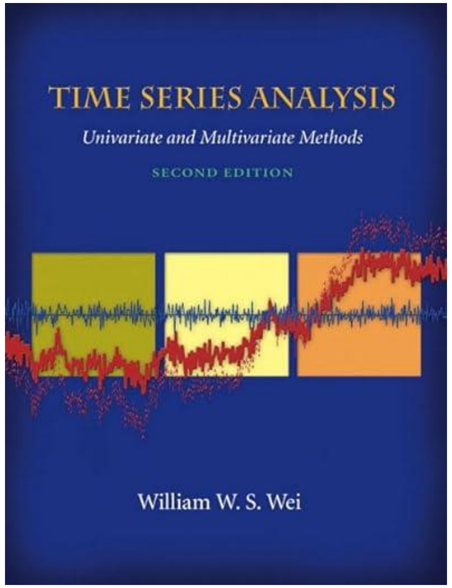
#Ref: Chapter 1 to 4, +
Time Series Analysis by William WS Wei

Time Series Analysis

Time Series Analysis
Sources & Processing

- Stochastics processes
- *Correlation*
- ACF, AVF, PACF, Noise, Visualisation
- MA, AR, ARMA, ARIMA
- Stationary & Non-stationary
- Seasonality

- Time Series Analysis



Queen Mary School Hainan
Queen Mary University of London

QHP5701 Exploratory Data Analysis

Stochastic Process

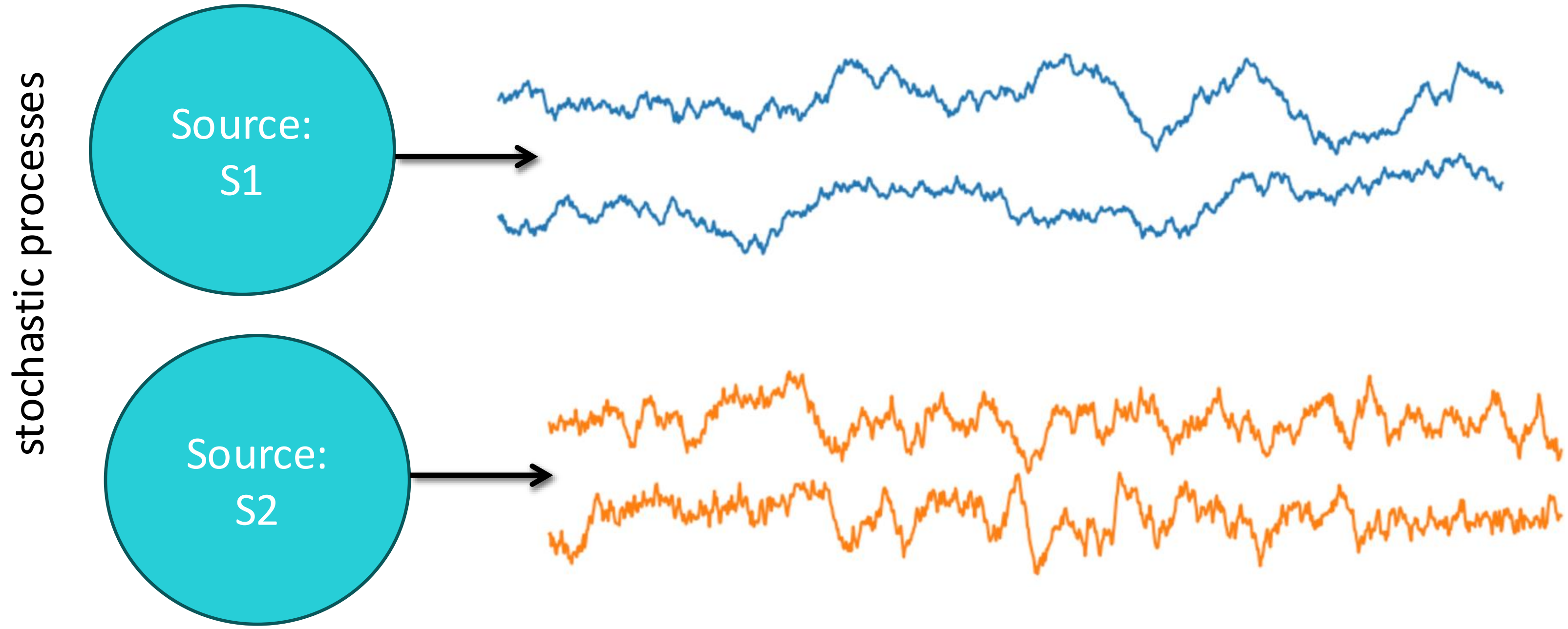
Nikesh Bajaj, PhD
Lecturer in Data Science,
Queen Mary University of London
nikesh.bajaj@qmul.ac.uk
https://nikeshbajaj.in

#Ref: Chapter 2, TSA, William WS Wei

Time Series Analysis: Source

- Time Series - temporal discreet data - time-dependent - ***signal x(n)***
- All time-series are considered to be generated by **a source**
- Objective of TSA is to analyse the source, from given a time-series

Time series is a realization of a stochastic process



Time Series Analysis: Source

- We analyse time-series, in order to know **more about the source**

- We will model these sources using some mathematical models (stochastic processes)
  - AR (q), MA(p), ARMA, ARIMA

- We will analyse properties of time-series, that help to understand, which model they come from
  - ACF, AVF, PACF

Time Analysis

Time series are typically referred to temporal data collected in specifics fields such as:

- Marketing time series
- Economics and Financial time series
- Demographic time series
- Population time series
- Physical time series

…

Queen Mary School Hainan
Queen Mary University of London

QHP5701 Exploratory Data Analysis

Stochastic Process: Fundamentals

Nikesh Bajaj, PhD
Lecturer in Data Science,
Queen Mary University of London
nikesh.bajaj@qmul.ac.uk
https://nikeshbajaj.in

#Ref: Chapter 2, TSA, William WS Wei

Random Variable $X: S \rightarrow R$

Random variable

- Not quite a variable x from algebra
- Ways to map random processes to numbers
- $X: S \rightarrow R$

- Continues/ Discreet random Variable

Examples

- $X = \begin{cases} 1, & for\ head \\ 0, & for\ tail \end{cases}$

$P(X = 1)$

- $\mathrm{Y} = sum\ of\ two\ dice$

$P(Y < 3)$

Expected value

Expected value $E[X]$ is a mean of random variable X

$$\mu_X = E[X] = \sum xp(X = x)$$

$$\mu_X = E[X] \equiv \text{mean of source X}$$

Expected value: Example

Expected value: Example

Estimation of Expected value: mean

Expected value $E[X]$ is a mean of random variable X

$$\mu_X = E[X] = \sum xp(X = x)$$

- Which is estimated from realization x, as sample mean

$$\hat{\mu}_X = \frac{1}{N}\sum x_i$$

Example

Example

Variance & Standard Deviation

Variance: $V[X] = E[(X - \mu_X)^2] = \sigma_X^2$

$Standard\ Deviation = \sigma_X$

Estimation of Variance & Standard Deviation

Variance: $V[X] = E[(X - \mu_X)^2] = \sigma_X^2$

$Standard\ Deviation = \sigma_X$

- This is estimated from realization x,

$$\hat{\sigma}_X^2 = \frac{1}{N-1}\sum(x_i - \hat{\mu}_X)^2$$

Covariance

Variance of x

$$\sigma_X^2 = E[(X-\mu_X)^2] = E[(X-\mu_X)(X-\mu_X)] = \sum (x-\mu_X)^2 P(X=x)$$

Covariance of x and y (cross-covariance)

$$\sigma_{XY}^2 = E[(X-\mu_X)(Y-\mu_Y)] = \sigma_{YX}^2$$

$$E[(X-\mu_X)(Y-\mu_Y)] = \sum (x-\mu_X)(y-\mu_Y)P(X=x, Y=y)$$

Example

Estimation of Covariance

Variance of x

$$\sigma_X^2 \ = \ E[(X-\mu_X)^2] = E[(X-\mu_X)(X-\mu_X)] = \sum (x-\mu_X)^2 P(X=x)$$

Covariance of x and y (cross-covariance)

$$\sigma_{XY}^2 = E[(X-\mu_X)(Y-\mu_Y)] = \sigma_{YX}^2$$

$$E[(X-\mu_X)(Y-\mu_Y)] = \sum (x-\mu_X)(y-\mu_Y)P(X=x, Y=y)$$

Covariance of x and y can be estimated as

$$\hat{\sigma}_{XY}^2 = \ Cov(x,y) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\hat{\mu}_X)(y_i-\hat{\mu}_Y)$$

Example

Example

Correlation

Correlation is closely related to covariance, so it can be computed as:

$$\rho_{XY} = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y}$$

$$-1 \leq \rho_{XY} \leq 1$$

$$Corr(x, y) = \frac{Cov(x, y)}{\hat{\sigma}_x \hat{\sigma}_y}$$

Example

Stochastic Processes: Time series

Collection of Random variable

- X_1 X_2 X_3 …
- 20 32 22
- $X_n \sim distribution(\mu_n, \sigma_n^2)$,

- Time series

Types of characteristics

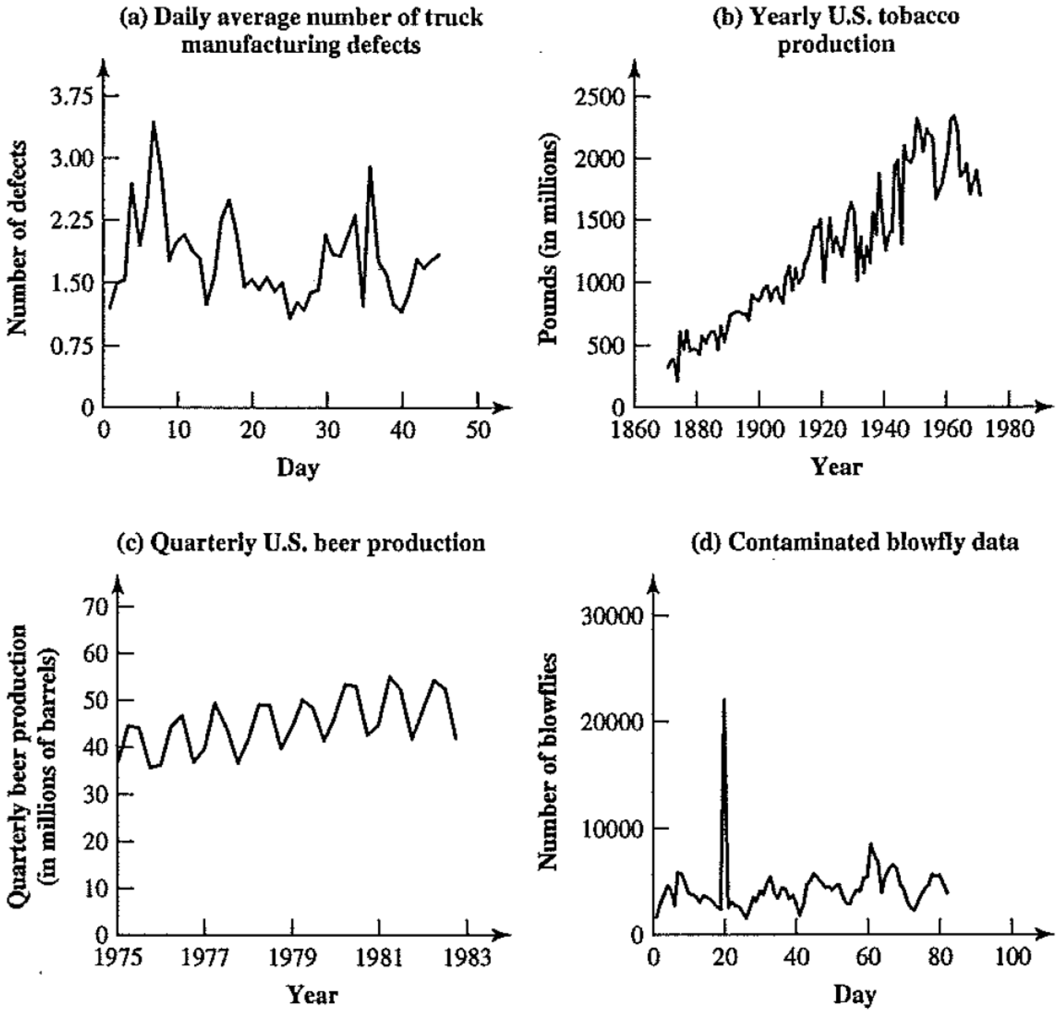
Time-series = signal

(a) (weak) Stationary

(b) Non-stationary

(c) Season variation: seasonality

(d) Non-stationary due to external disturbance

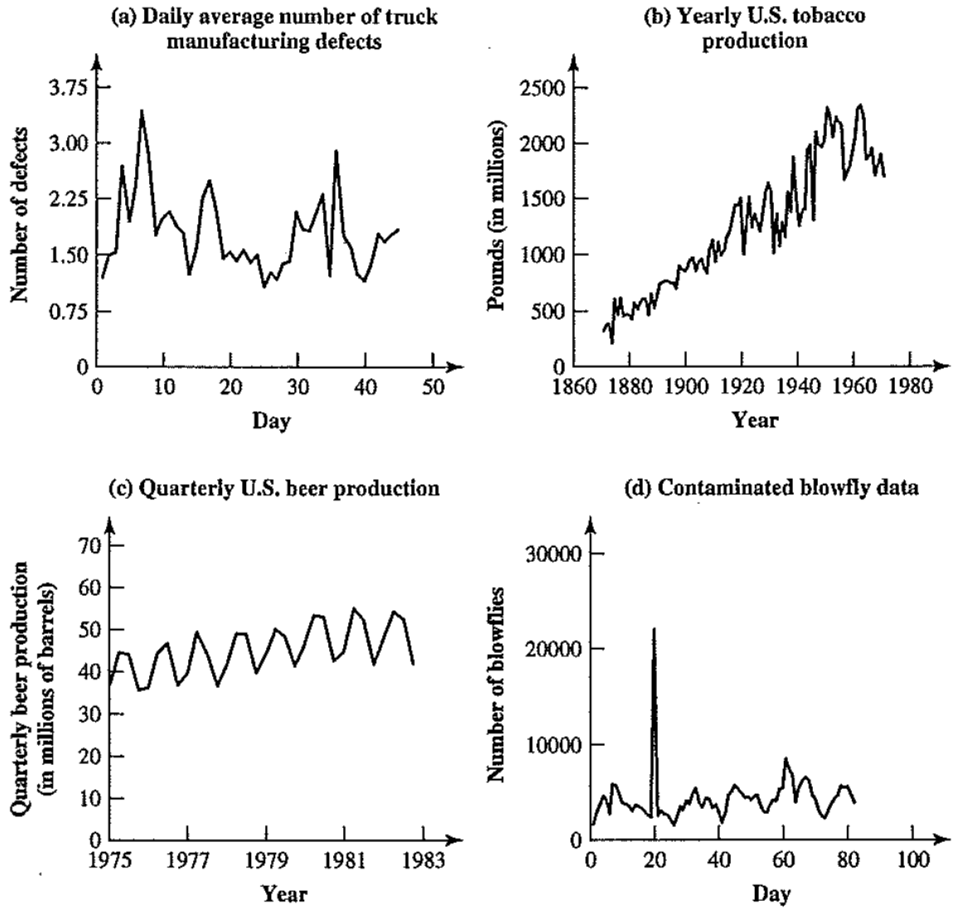


Stationarity

Stationary (a):

- No systematic change in time-series,
- No trends,
- No change in mean
- No period fluctuation

Properties of one part (section) of signal, is same as another part of signal



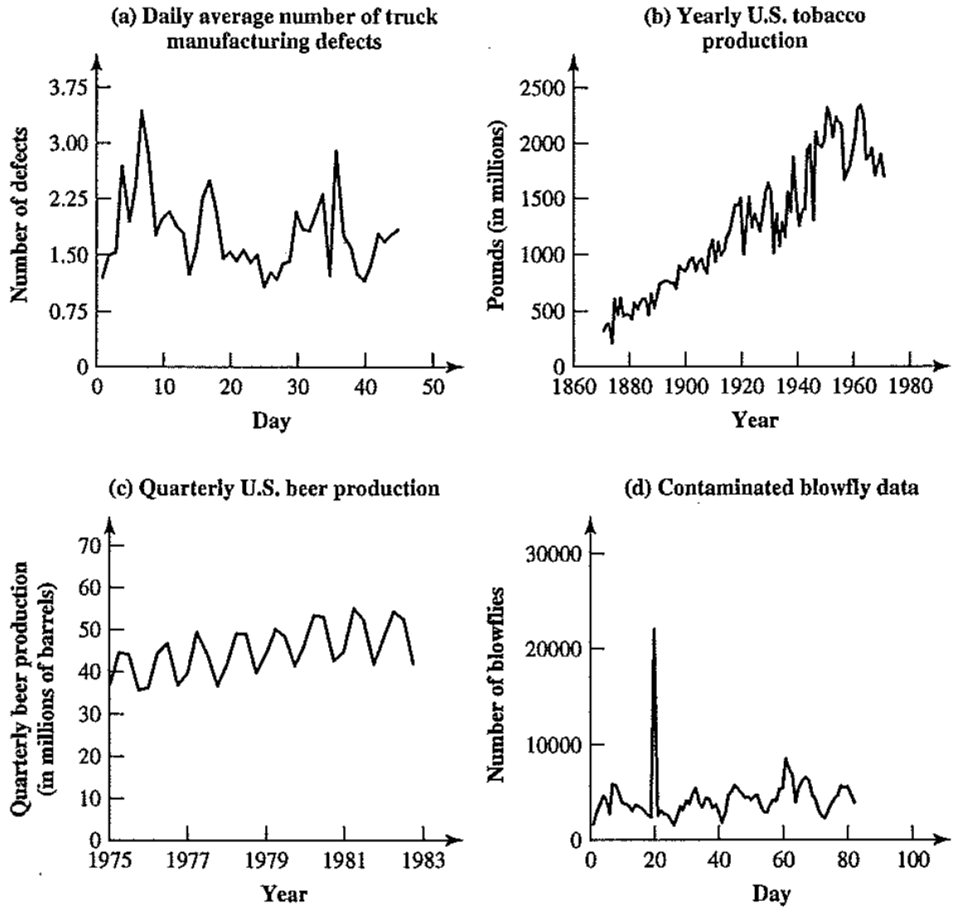
Non-Stationarity

Non-Stationary (b) & (c):

- trends,

*Properties of one part (section) of signal, is **NOT** same as another part of signal*

Transform into Stationary time series



Stationarity

Autocovariance Coefficient

Autocovariance Coefficient

$$C_k = \frac{1}{N-k}\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x}).$$

Where, $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

Autocovariance Coefficient

Autocovariance Coefficient

$$C_k = \frac{1}{N-k}\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x}).$$

Where, $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

Queen Mary School Hainan
Queen Mary University of London

QHP5701 Exploratory Data Analysis

Models

Nikesh Bajaj, PhD
Lecturer in Data Science,
Queen Mary University of London
nikesh.bajaj@qmul.ac.uk
https://nikeshbajaj.in

#Ref: Chapter 3, TSA, William WS Wei

Noise

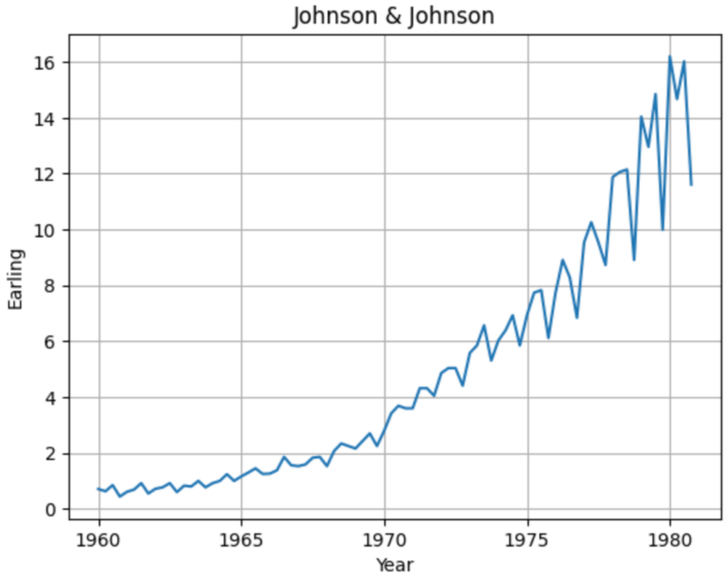
Random Walk

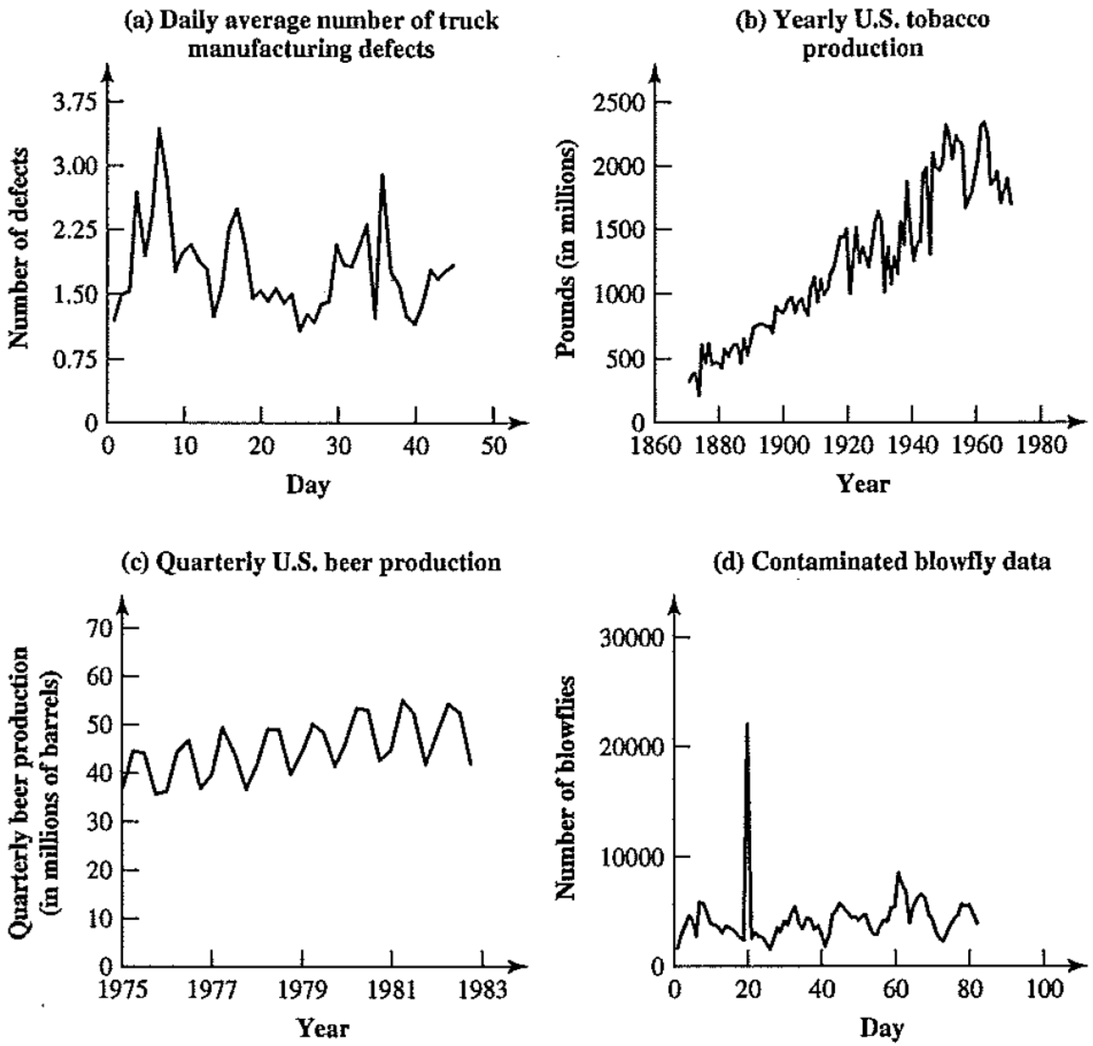
MA

AR

ARMA

Examples


Johnson & Johnson
Earling
Year
0
2
4
6
8
10
12
14
16
1960
1965
1970
1975
1980


(a) Daily average number of truck manufacturing defects
Number of defects
Day
(b) Yearly U.S. tobacco production
Pounds (in millions)
Year
(c) Quarterly U.S. beer production
Quarterly beer production (in millions of barrels)
Year
(d) Contaminated blowfly data
Number of blowflies
Day

Queen Mary
University of London