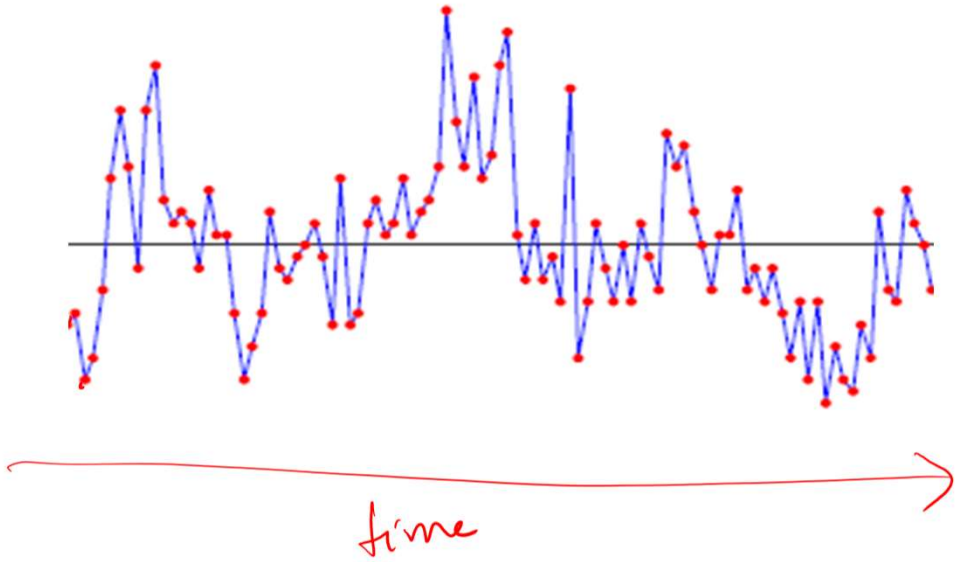
Time Series Analysis

Reading material

https://online.stat.psu.edu/stat510/lesson/1

What is a time-series

- Sequence of values recorded at regular time intervals
  - Time interval: E.g. Weekly, monthly, daily, hourly, annually, etc.
  - Values recorded:
    - Scalar: single value like sales
    - Vector of values

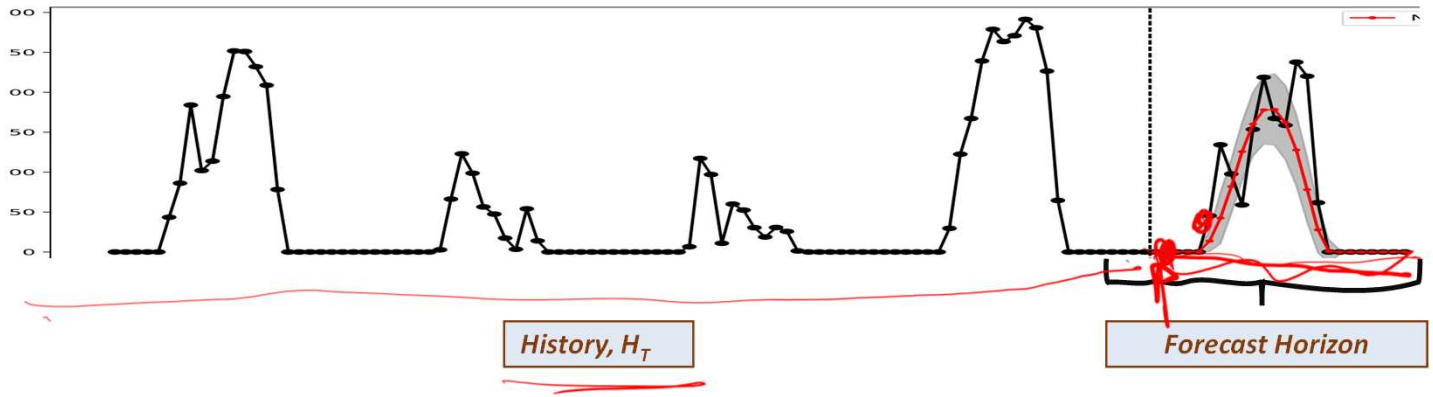


Motivation

- Daily traffic on individual webpages from different regions in Wikipedia
- Hourly load on various servers of different services in a Data center
- Monthly demand for products from different regions in Flipkart
- Stock price of various companies
- Rice production in Maharashtra each year
- Consumer price index of various food items.

Objective of time series analysis

- Identify characteristics of the time-series
  - Provide a model of the data (e.g. test scientific hypothesis)
- Forecasting
  - Predict future values as a function of past values.



- Finding outliers and filling in missing values
- Provide a compact description of the data (data compression)

Important characteristics of time-series

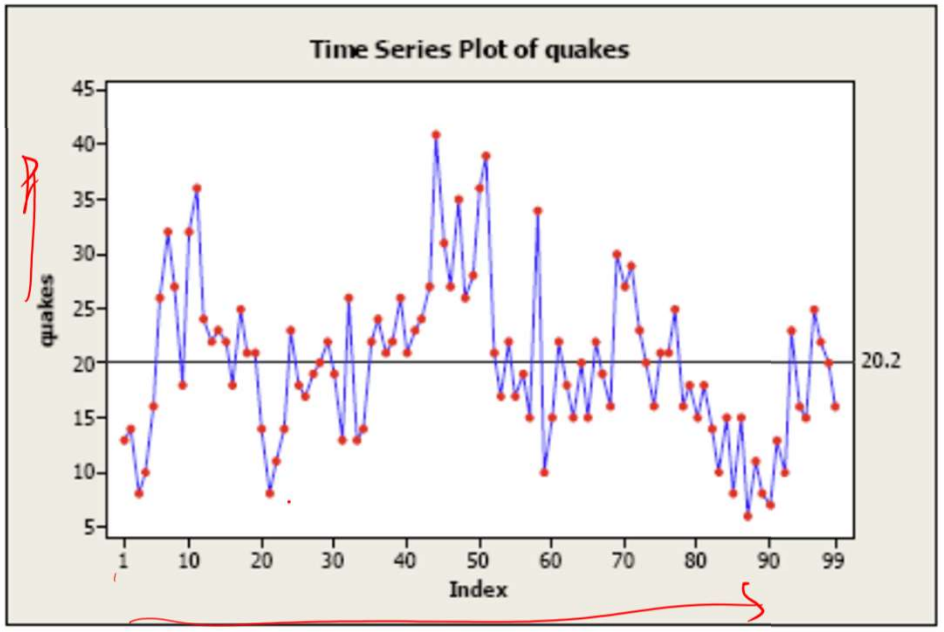
- Is there a **trend**, meaning that, on average, the measurements tend to increase (or decrease) over time?
- Is there **seasonality**, meaning that there is a regularly repeating pattern of highs and lows related to calendar time such as seasons, quarters, months, days of the week, and so on?
- Are there **outliers**? In regression, outliers are far away from your line. With time series data, your outliers are far away from your other data.
- Is there a **cycle**: data rises and falls but without a fixed frequency.
- Is there **constant variance** over time, or is the variance non-constant?
- Are there any **abrupt changes** to either the level of the series or the variance?

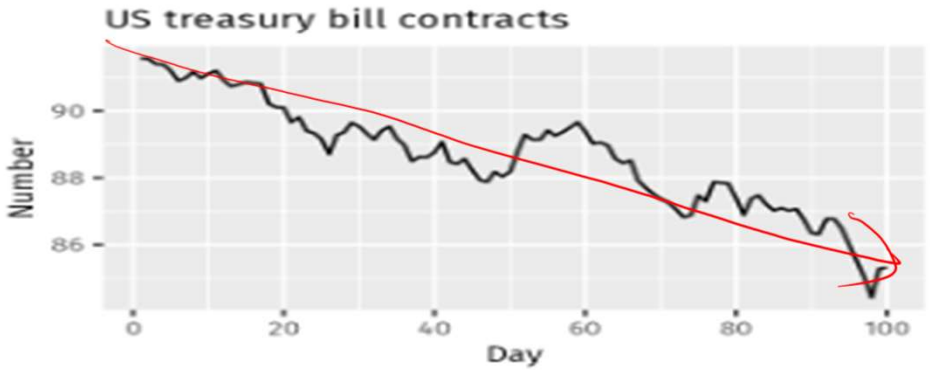
Example 1-1 (from book)

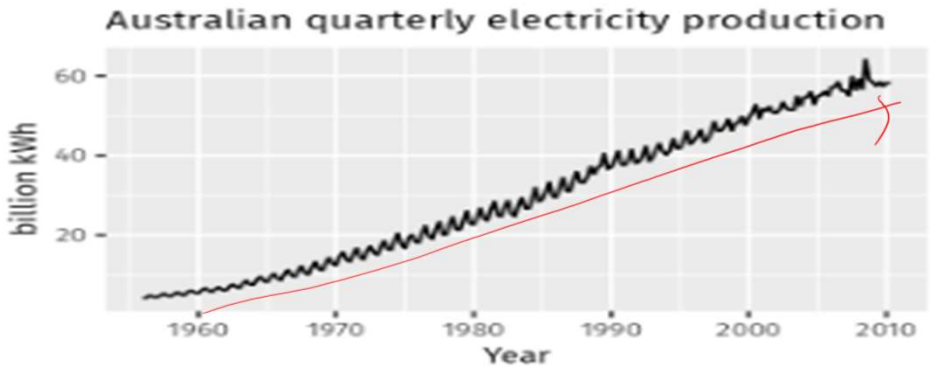
Annual number of earthquakes in the world with seismic magnitude over 7.0, for 99 consecutive years

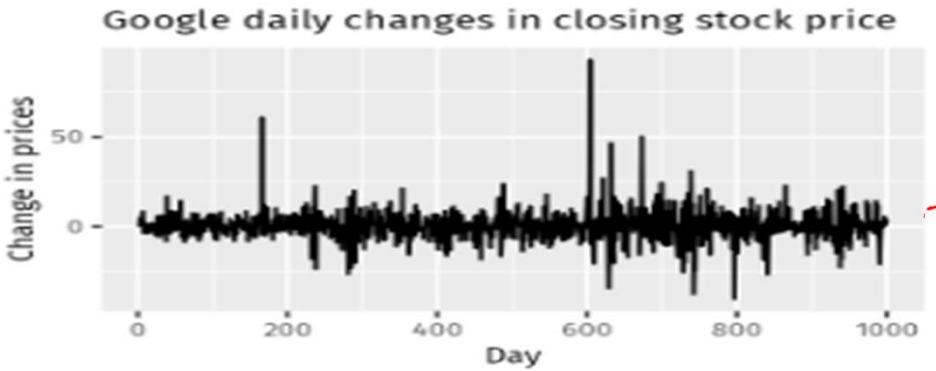
Characteristics:

- No consistent trend. Values on both sides of mean along time
- No seasonality
- No outliers



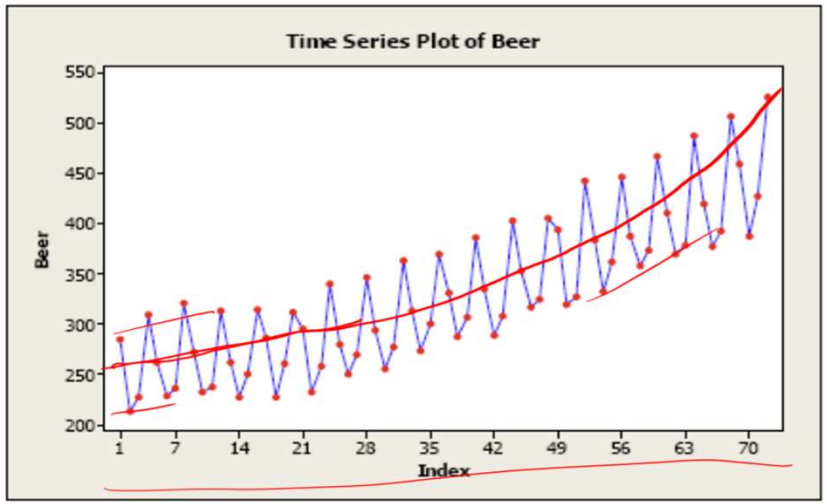






1.The monthly housing sales (top left) show strong seasonality within each year, as well as some strong cyclic behaviour with a period of about 6–10 years. There is no apparent trend in the data over this period.
2.The US treasury bill contracts (top right) show results from the Chicago market for 100 consecutive trading days in 1981. Here there is no seasonality, but an obvious downward trend. Possibly, if we had a much longer series, we would see that this downward trend is actually part of a long cycle, but when viewed over only 100 days it appears to be a trend.
3.The Australian quarterly electricity production (bottom left) shows a strong increasing trend, with strong seasonality. There is no evidence of any cyclic behaviour here.
4.The daily change in the Google closing stock price (bottom right) has no trend, seasonality or cyclic behaviour. There are random fluctuations which do not appear to be very predictable, and no strong patterns that would help with developing a forecasting model.

Example 1-2



Quarterly production of beer in Australia

For the past 18 years.

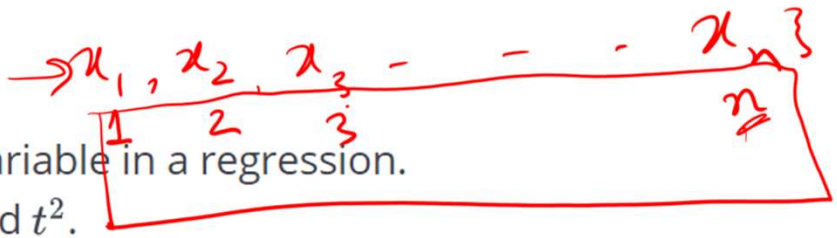
- Length = 4*18=72

Characteristics

- There is an upward trend, possibly a curved one.
- There is seasonality – a regularly repeating pattern of highs and lows related to quarters of the year.
- There are no obvious outliers.
- There might be increasing variation as we move across time, although that's uncertain.

Fitting simple regression models for quantifying pattern in time-series

Suppose that the observed series is x_t, for $t = 1, 2, \ldots, n$.



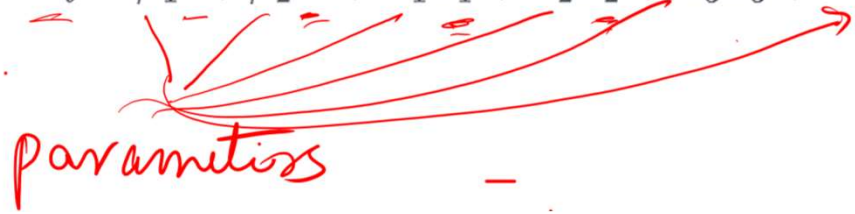
- For a linear trend, use t (the time index) as a predictor variable in a regression.
- For a quadratic trend, we might consider using both t and t^2.
- For quarterly data, with possible seasonal (quarterly) effects, we can define indicator variables such as $S_j = 1$ if the observation is in quarter j of a year and 0 otherwise. There are 4 such indicators.

Let $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. A model with additive components for linear trend and seasonal (quarterly) effects might be written

$$x_t = \beta_1 t + \alpha_1 S_1 + \alpha_2 S_2 + \alpha_3 S_3 + \alpha_4 S_4 + \epsilon_t$$

To add a quadratic trend, which may be the case in our example, the model is

$$x_t = \beta_1 t + \beta_2 t^2 + \alpha_1 S_1 + \alpha_2 S_2 + \alpha_3 S_3 + \alpha_4 S_4 + \epsilon_t$$



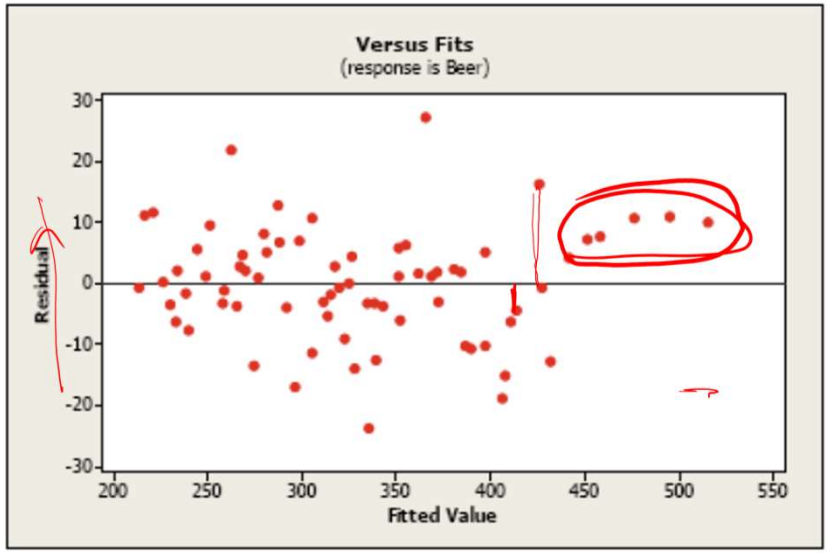
Fitting the dataset of 72 values using least square regression

Predictor	Coef	SE Coef
Noconstant		
Time	0.5881	0.2193
tsqrd	0.031214	0.002911
quarter_1	261.930	3.937
quarter_2	212.165	3.968
quarter_3	228.415	3.994
quarter_4	310.880	4.018

Residual analysis

- Ideal residuals: mean 0, normally distributed.

- For time-series: correlation between residues separated by a fixed time-span should be zero.



Limitation of regression models

- Does not account for strong temporal correlation among values.
- Predicts each value independent of its neighbors.
- Need a model that can directly exploits correlations within a series.

Sample Autocorrelation Function (ACF)

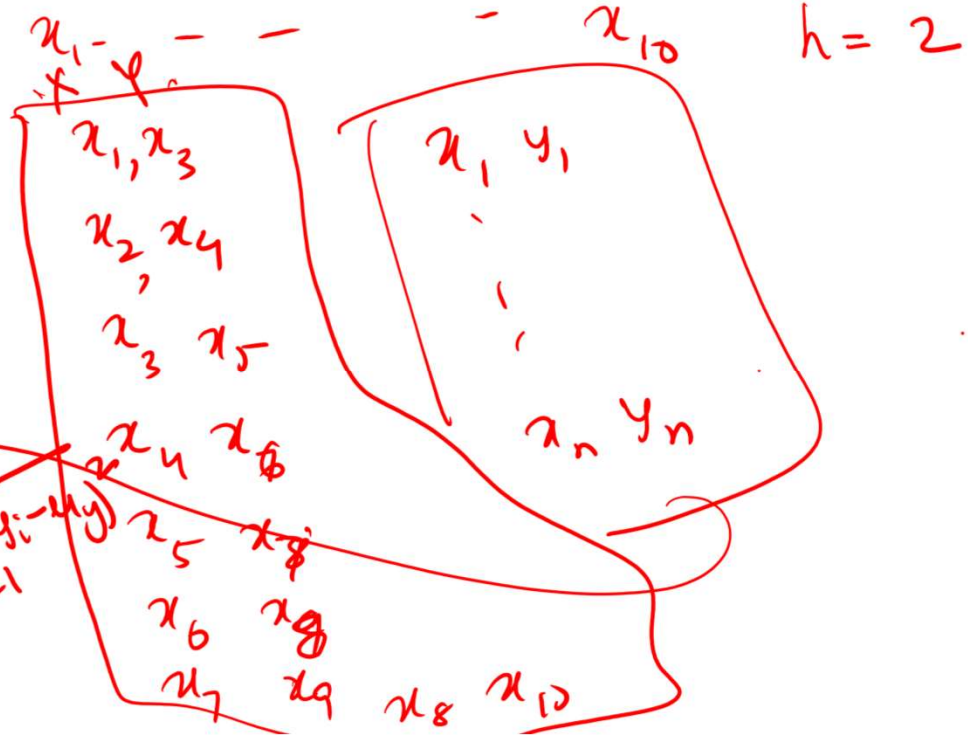
- Correlation between a series value x_t and lagged values for different values of lags h.
- Lag h auto-correlation function: correlation between x_t, x_{t-h} for all t.

$$\frac{\text{Covariance}(x_t, x_{t-h})}{\text{Std.Dev.}(x_t)\text{Std.Dev.}(x_{t-h})}$$

Correlation (X, Y)

$$= \frac{\text{Covariance}(X, Y)}{\text{Std Dev}(X)\ \text{Std Dev}(Y)}$$

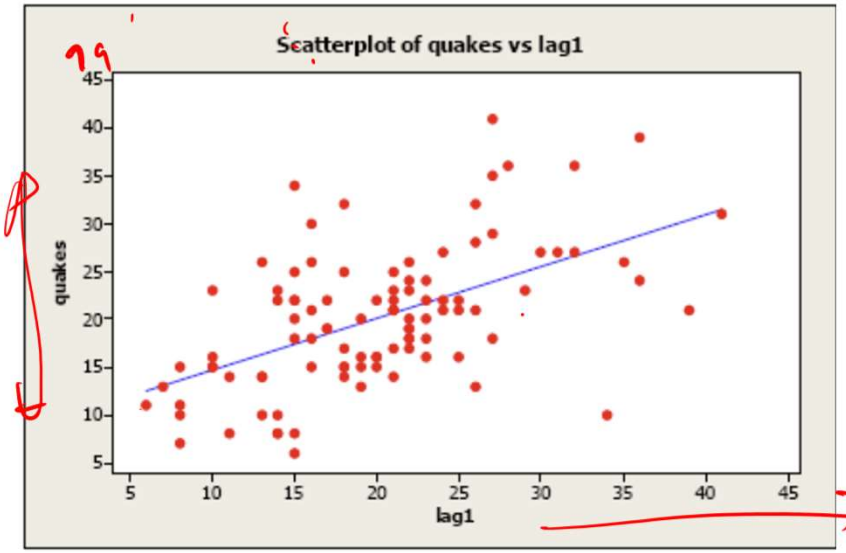
$$= \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i - \mu_x)^2}\sqrt{\sum_{i=1}^{n}(y_i - \mu_y)^2}}$$

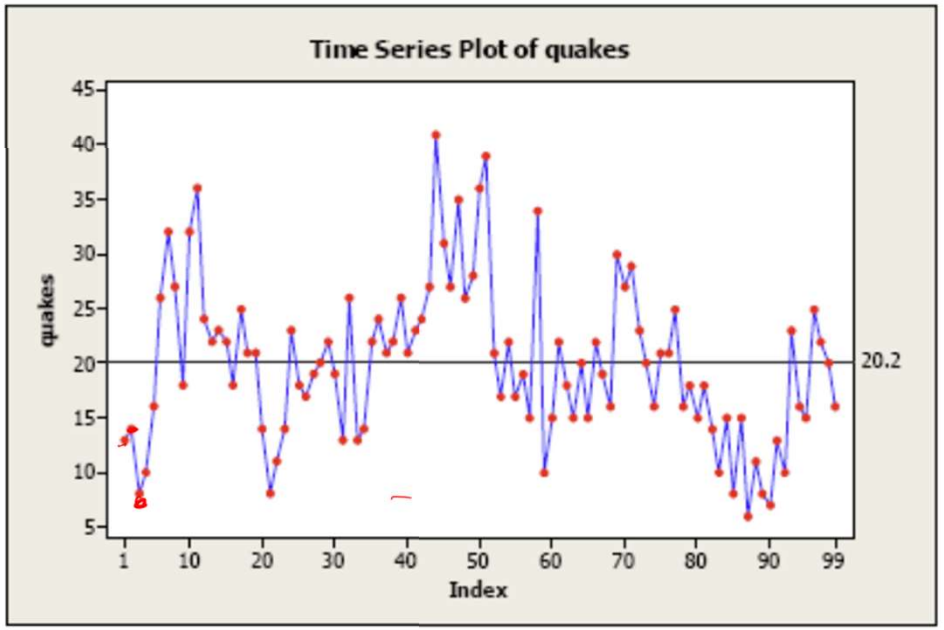


Example 1-1 (from book)

- Earthquake data with lag h=1

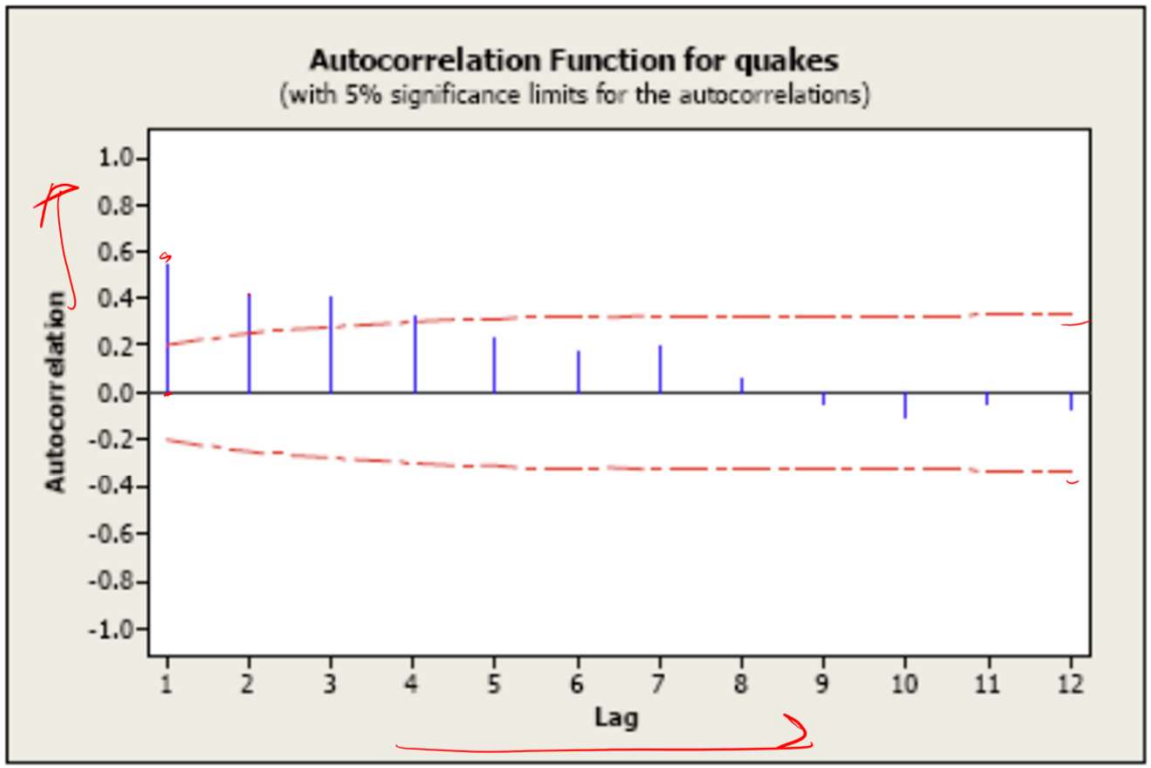
t	x_t	x_{t-1} (lag 1 value)
1	13	*
2	14	13
3	8	14
4	10	8
5	16	10





Example ACR for lag > 1

Lag.	ACF
1.	0.541733
2.	0.418884
3.	0.397955
4.	0.324047
5.	0.237164
6.	0.171794
7.	0.190228
8.	0.061202
9.	-0.048505



Stationary series

A series xt is said to be **(weakly) stationary** if it satisfies the following properties:

- The mean E(xt) is the same for all t.
- The variance of xt is the same for all t.
- The covariance (and also correlation) between x_t and x_{t-h} is the same for all t at each lag h = 1, 2, 3, etc.

Autocorrelation function for stationary series.

$$ACF(h) = \frac{\text{Covariance}(x_t, x_{t-h})}{\text{Std.Dev.}(x_t)\text{Std.Dev.}(x_{t-h})} = \frac{\text{Covariance}(x_t, x_{t-h})}{\text{Variance}(x_t)}$$

Which of the following series are stationary?

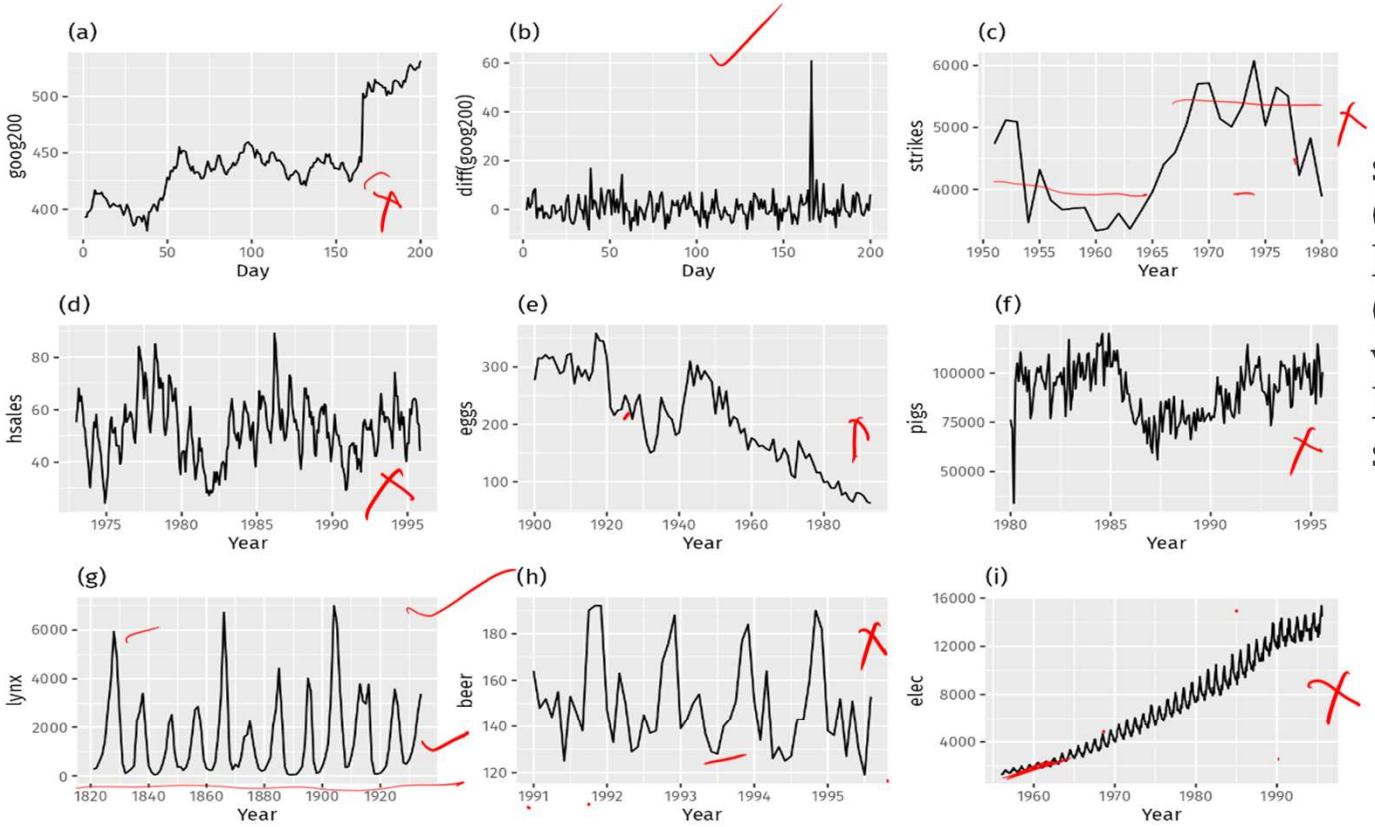


Figure 8.1: Which of these series are stationary? (a) Google stock price for 200 consecutive days; (b) Daily change in the Google stock price for 200 consecutive days; (c) Annual number of strikes in the US; (d) Monthly sales of new one-family houses sold in the US; (e) Annual price of a dozen eggs in the US (constant dollars); (f) Monthly total of pigs slaughtered in Victoria, Australia; (g) Annual total of lynx trapped in the McKenzie River district of north-west Canada; (h) Monthly Australian beer production; (i) Monthly Australian electricity production.

seasonality rules out series (d), (h) and (i). Trends and changing levels rules out series (a), (c), (e), (f) and (i). Increasing variance also rules out (i). That leaves only (b) and (g) as stationary series.

https://otexts.com/fpp2/stationarity.html

Models for time-series

- Auto-regressive models AR(p):
- Moving average models MA
- ARIMA models (Combines the above two)
- SARIMA: Generalization of ARIMA to handle seasonality.

Auto-regressive models

- AR(p) model: The value of x at time t is a linear function of the value of x at time t−1, t-2,…t-p.

$$AR(p): \quad x_t = \delta + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

- AR(1) $x_t = \delta + \phi_1 x_{t-1} + w_t$

(δ: parameter, w_t: error.)

$$x_t = x_{t-1} + w_t, \quad \phi_1 = 1$$

- $w_t \overset{iid}{\sim} N(0, \sigma_w^2)$, meaning that the errors are independently distributed with a normal distribution that has mean 0 and constant variance.
- Properties of the errors w_t are independent of x_t.
- The series $x_1, x_2, \ldots$ is (weakly) stationary. A requirement for a stationary AR(1) is that $|\phi_1| < 1$. We'll see why below.