Introduction to Applied Time Series

STAT 1321/2320

Kiran Nihlani

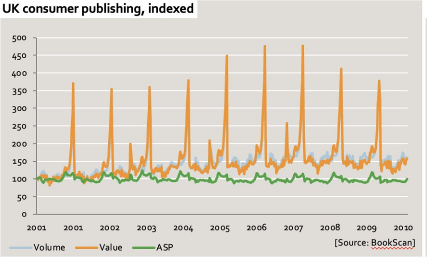
Department of Statistics
University of Pittsburgh

Outline

What is a Time Series?

A sequence of data points organized in time order.

- The sequence captures data at equally spaced points in time
- Data collected irregularly is not considered a time series



Credit: *Twitter: Benedict Evans* (@benedictevans)

UK publishing data

→ Descr. → consistent inc/dec.

→ seasonality / periodic fluctuations. every 2 years.

↳ how often ?

↳ reasoning.

Prevalence of Time Series

- Economics: stock prices, unemployment rate, inflation rate
- Social sciences: population series, such as birthrates or school enrollments
- Epidemiology/Medicine: the number of influenza/COVID-19 cases observed over some time period, blood pressure measurements traced over time for evaluating drugs
- E-Commerce: page views, new users, searches, targeted advertising
- Global warming?

Applications - Economic Forecasting

Great stagflation.
→ food, energy, govt wage price control program.

Macroeconomic Predictions:

- International Monetary Fund (IMF) collects data on trade exports, imports, and trade for all member and some non-member states for forecasting.
- Federal reserve uses time series based on Consumer Price Index to forecast inflation rates.

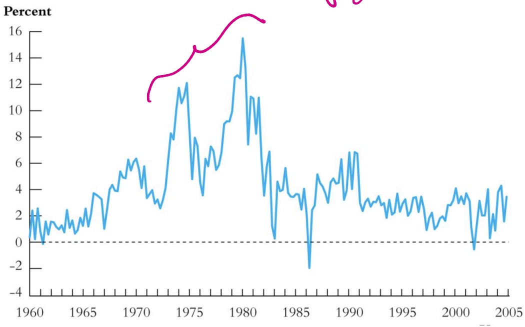


Figure: US rate of price inflation, as measured by the quarterly percentage change in the Consumer Price Index (CPI)

Applications - E-Commerce

demand in a long term setup.

more seasonal expectations.

Used to predict demand, both overall and at more granular levels

- Anticipatory package shipping where a product is shipped to a local distribution center before you buy them.
- Inventory forecasting (also known as demand planning) is the practice of using past data, trends and known upcoming events to predict needed inventory levels for a future period.

Figure: Amazon Fulfillment Center, Baltimore, MD by Maryland GovPics

Credit: *https://www.forbes.com/sites/onmarketing/2014/01/28/why-amazons-anticipatory-shipping-is-pure-genius/?sh=225ff41c4605*

Applications - Medical Research

The use ranges from tracking and forecasting infectious diseases to using individual data for personalized medicine.

- The CDC actively encourages researchers to work on forecasting the flu by sponsoring flu forecasting competitions.
- Labs use time series to understand the progression and clinical trajectories in a wide variety of diseases, including cancer, cystic fibrosis, Alzheimer's, cardiovascular disease, and COVID-19.

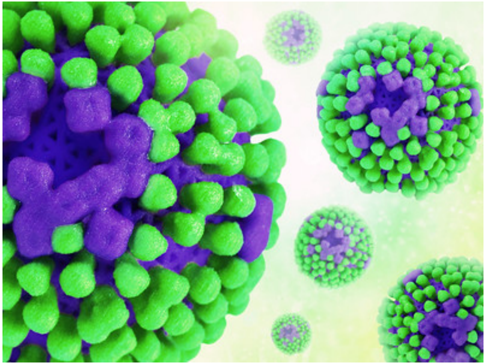


Figure: 3D print of influenza virus NIAID (NIH)

Why do we need separate analysis for time series data?

- Time series is a collection of observations made sequentially through time. Implications?
  - The observations are ordered. ✓
  - There is a possible dependence between successive observations.

  $\Longrightarrow$ past observations can be used to predict future values

- This gives us the first basis for classifying time series
  - Deterministic → if a time series can be predicted exactly.
    $y_t = 0.6t$, $y_t = 2\cos(2\pi t/5)$
  - Stochastic ↳ partly determined by the past values, exact pred^n is impossible.
    ↳ future values have a probab. dist. which is cond^n by past values.

challenging + fun

Objectives of Time Series Analysis

- Description
  - Plot the data and obtain simple descriptive measures of the main properties of the series.
- Explanation
  - Find a model to describe the time dependence in data.
- Prediction
  - Given a finite sample from the series (observations), forecast the next value or the next several values.
- Control
  - After forecasting, adjust various control/tune parameters.

Technical Definition

X_t : time series. $\{X_1, X_2, \dots X_n\}$.

x_t : realization.

- Time series is a sequence of random variables, $X_1, X_2, X_3, \dots,$ where random variable X_1 corresponds to the value taken by the series at the first time point, the variable X_2 corresponds to the value for the second time point, and so on.
  - In general, a collection of random variables $\{X_t\}$ indexed by t is referred to as a stochastic process.
  - The observed values $x_1, x_2, x_3, \dots,$ of a stochastic process are referred to as a realization of the stochastic process.
  - Most texts don't make any notational distinction between the two concepts and just refer to them as time series.

X_t, x_t

Classification of Time Series

daily, monthly expense, yearly snowfall data.

discretized.

- **Discrete**
  - Observations are made at fixed time intervals i.e. set T_0 of times at which observations are made is a discrete set.
  - However, the underlying data source may be in continuous time.

    temp.

- **Continuous**
  - Observations are recorded continuously over some time interval, e.g., when $T_0 = [0, 1]$.

    complex

Components of Time Series → Description + Explanation.

Components of time series are important because either we would like to know about the specific component or we would like to analyze the series after eliminating the effect of a particular component.

A time series is composed of 4 parts:

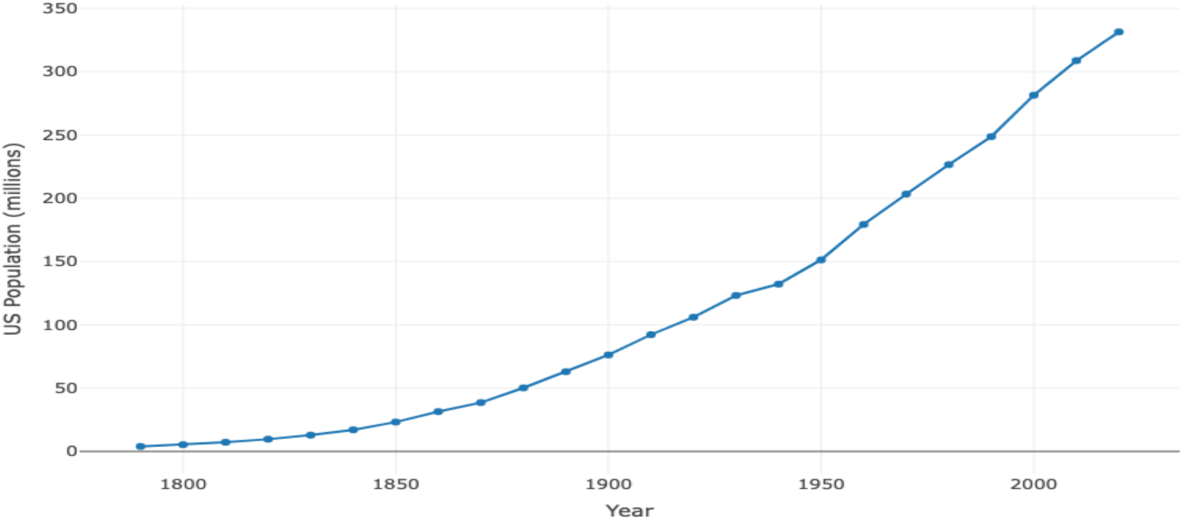
- Secular Trend(T) (or, just Trend)
- Seasonal Variation(S)
- Cyclical Variation(C)
- Irregular Fluctuation(I) (or, error/random component)

Trend

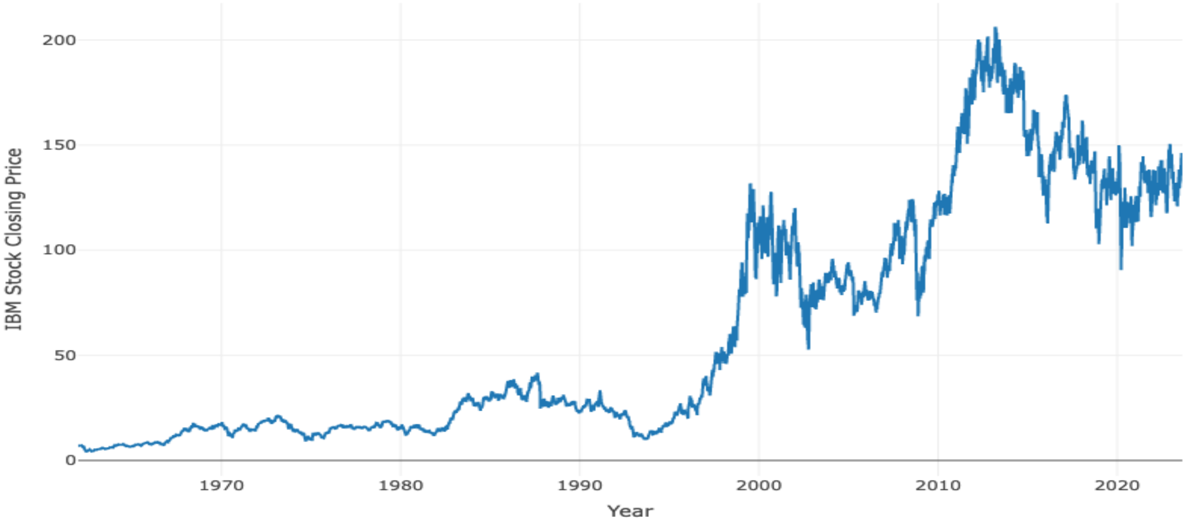
Trend captures the general direction of the time series.

- For example, increasing job growth year over year despite seasonal fluctuations.
- Trend can be increasing, decreasing, or constant.
- It can increase or decrease in different ways (linearly, exponentially, or in other ways).

US Population at Ten Year Intervals, 1970-2020



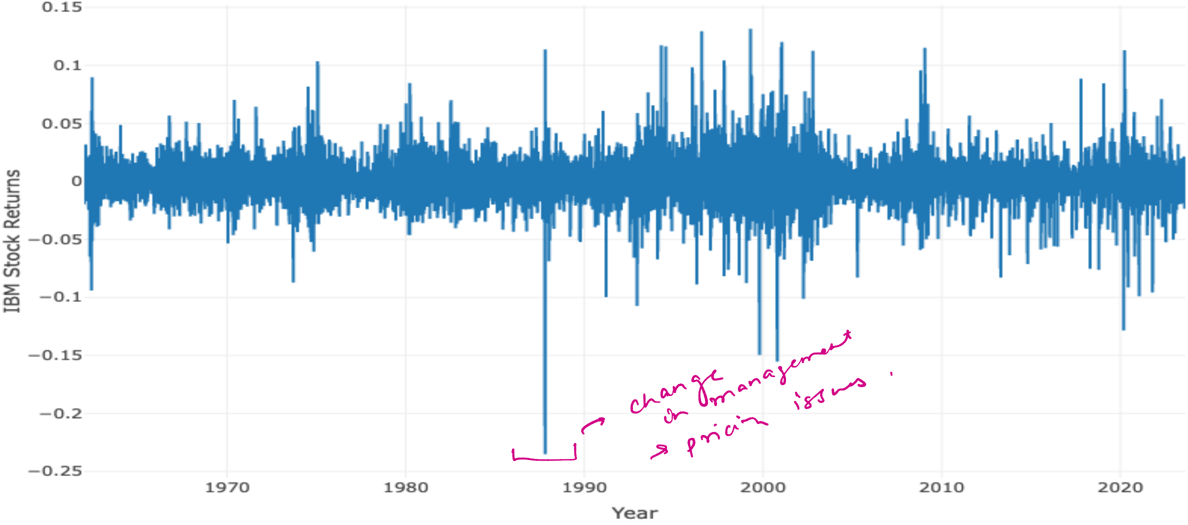
IBM Daily Stock Closing Price



Heteroskedasticity & Volatility Clustering

- Heteroskedasticiy
  - Heteroskedasticiy implies that the variance of the time series changes with time.
  - Smaller changes can be handled using transformations but we will need separate tools if the change is consistent over time.
- Volatility Clustering
  - This is a phenomenon characterized by significant changes in values of the time series that form groups or clusters.
  - This clustering occurs primarily due to the extreme volatility experienced by the market in response to local/global events.

IBM Stock Returns



Seasonality

Seasonality captures effects that occur with specific frequency.

- It can be driven by many factors - quarterly/annually, holidays or religious observances, weather fluctuations, etc.

  → time of year.

- Seasonality is always of a fixed and known period. Hence, seasonal time series are sometimes called periodic time series.
  - This fixed period is called seasonality period, and it is important for analysing time series.

    → interpretation.

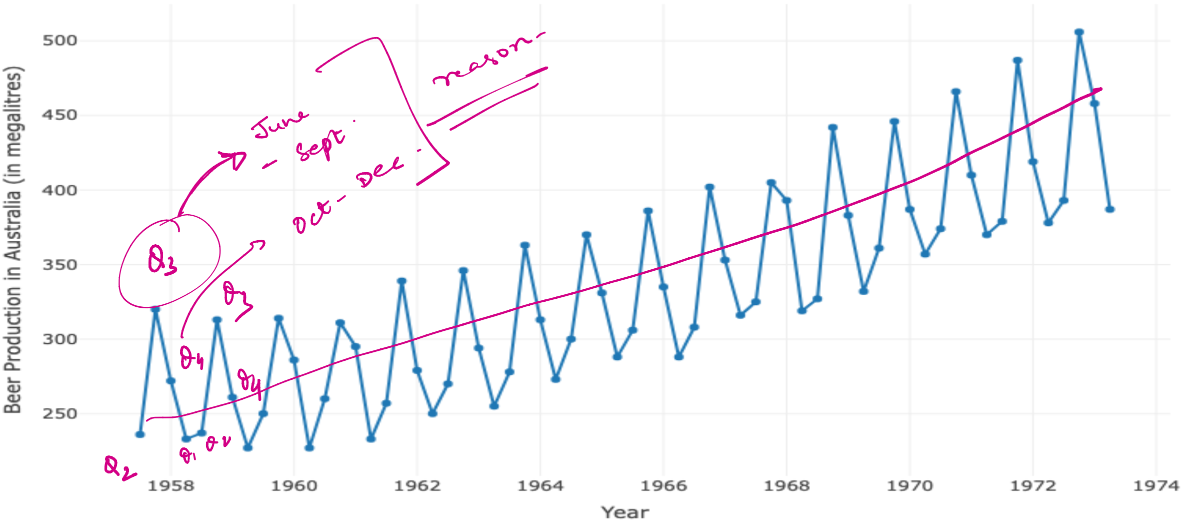
    → model fitting.

- Handling seasonality in data depends on the objective of the analyses. The objective may be to understand trend (by eliminating seasonality) or just explore the seasonal fluctuation.

Beer Production in Australia

Quarterly data.

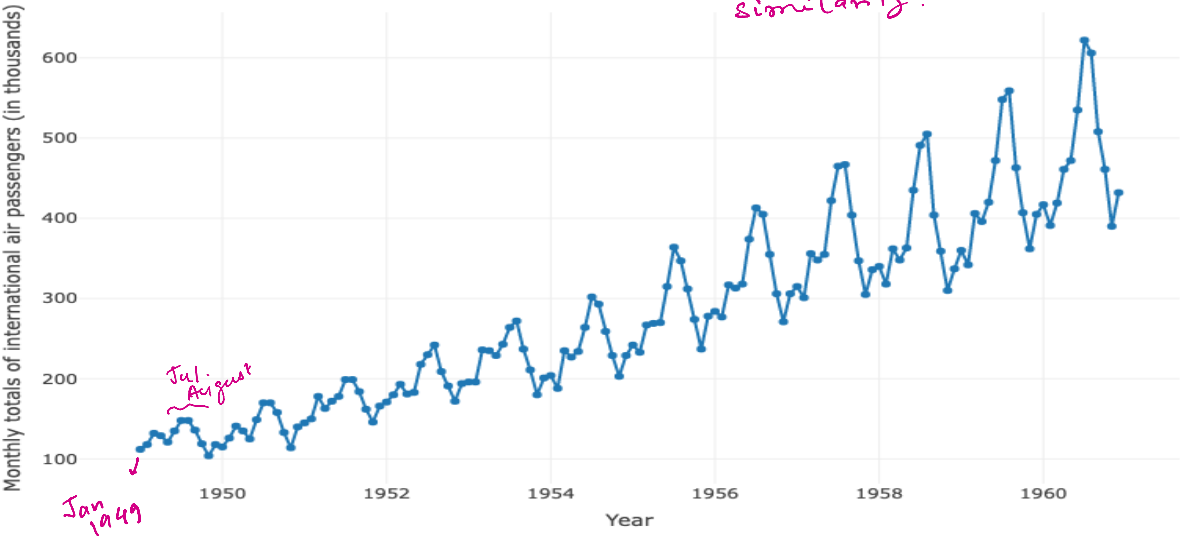
Trend + seasonality.



International Airline Passengers

Diff. b/w the 2 graphs.

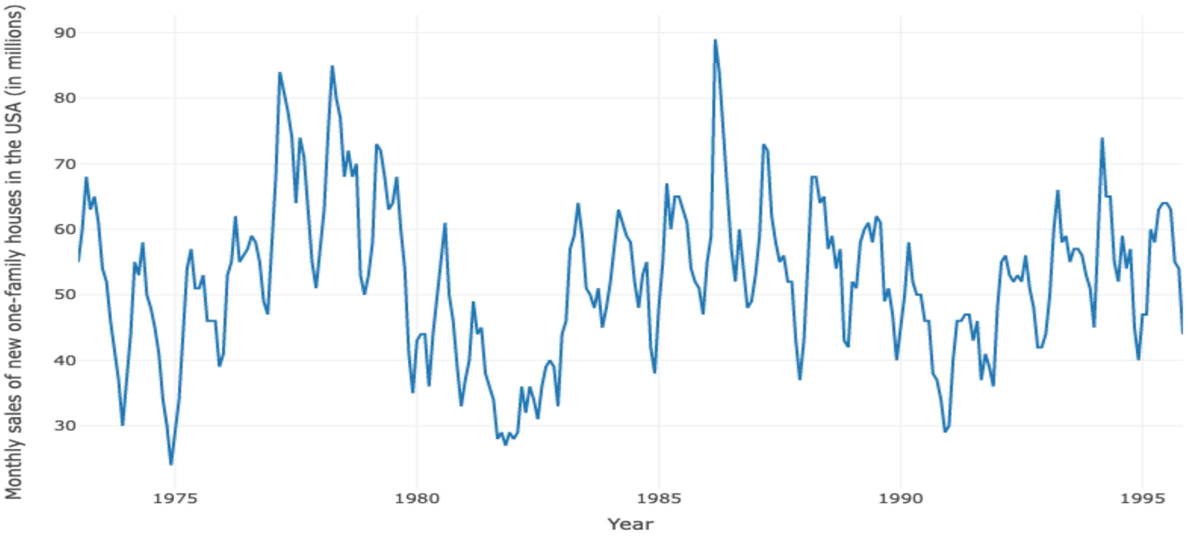
similarity.



Cyclical Fluctuations

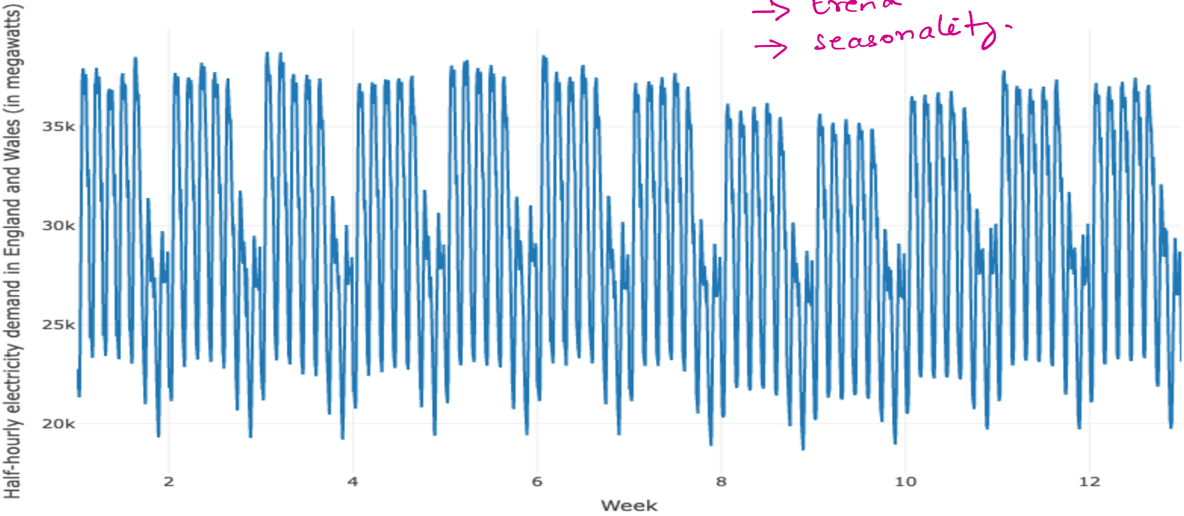
- A cyclic pattern exists when data exhibit rises and falls that are not of fixed period.
- One complete period is known as a cycle.
- The duration of these fluctuations is usually of at least 2 years.
- Cycles may last several years, but the length of the current cycle is unknown beforehand. Examples?

Sales of New 1-Family Houses in USA (01/73 - 11/95)



Electricity Demand in England and Wales from Mon 06/05/2000 - Sun 08/27/2000

→ trend.
→ seasonality.



$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. $\longrightarrow$ Additiv

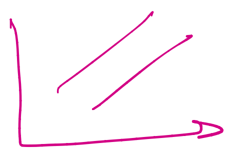
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \cdot x_2)$

$H_0: \beta_3 = 0$.

y: satisfaction.

x_1: food item. → ice cream. / → hot dog.

x_2: condiments → chocolate sauce. / → mustard.



X_t = Trend + Seasonality. + Cyclical + Random/irregular fluctuation.

$$X_t = T_t + S_t + C_t + I_t.$$

$$X_t = T_t * S_t * C_t * I_t$$

Decomposition Models

Time series components, very basically, can be decomposed with the following models:

- Additive decomposition
  - Additive models assume that the observed time series is the sum of its components

$$\text{Observation} = \text{Trend} + \text{Seasonality} + \text{Cyclic} + \text{Irregular Fluctuation}$$

- Multiplicative decomposition
  - The observed time series multiplicative models assume that the observed time series is the product of its components

$$\text{Observation} = \text{Trend} \times \text{Seasonality} \times \text{Cyclic} \times \text{Irregular Fluctuation}$$

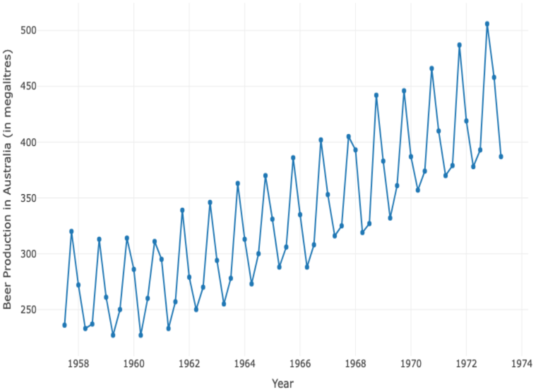
Additive vs. Multiplicative

- Additive models are used when the magnitudes of the seasonal and residual values are independent of trend.
- Multiplicative models are used when the magnitudes of the seasonal and residual values fluctuate with trend.
  - It is possible to transform a multiplicative model to an additive by applying a log transformation
- Just FYI - a lot of texts combine the Trend and Cyclical components together so the formula might look different.

$$X_t = T_t * S_t * C_t * I_t .$$

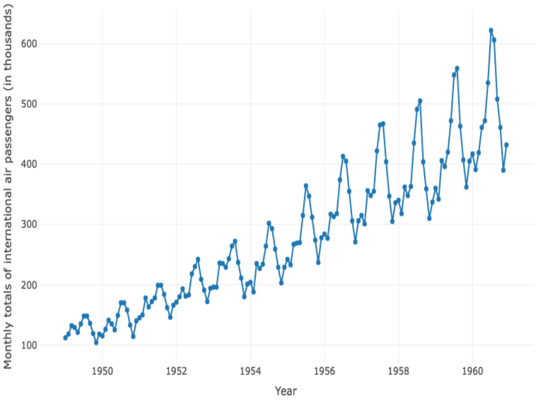
$$\log X_t = \log T_t + \log S_t + \log C_t + \log I_t .$$

Beer Production



- Magnitude of seasonality is constant over time.
- Preferred Model: Additive.

International Air Passengers



- Magnitude of seasonality changes (increases) over time.
- Preferred Model: multiplicative.

Irregular Fluctuation

- All series contain another factor that cannot be accounted for by the regular variations (trend, seasonality, cyclical).
- These fluctuations are random and unpredictable, and often, short term.

Time Series Model

$\{x_1, x_2, \ldots\ldots\}$

A time series model specifies the joint distribution of the sequence $\{X_t\}$ of random variables; e.g.,

$P(X \le x) = F(x)$.

$$F \qquad P(X_1 \le x_1, \ldots, X_t \le x_t) \text{ for all } t \text{ and } x_1, \ldots, x_t$$

where $\{X_1, X_2, \ldots\}$ is a stochastic process, and $\{x_1, x_2, \ldots\}$ is a single realization.

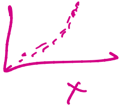
- Such a specification is rarely used in time series analysis, since in general it will contain far too many parameters to be estimated from the available data.
- We will only focus on second order properties of $\{X_t\}$ - first and second order moments of the joint distribution

$\mu_1' = E(X)$

$\sigma^2 = \mu_2 = E[(X-\mu)^2]$.

White Noise - Definition

$y = x^2$



- Collection of uncorrelated random variables with mean 0 and finite variance σ_w^2

$$W_t \sim wn\,(0, \sigma_w^2)$$

white noise.

- If the noise is independent and identically distributed,

mean.

$$W_t \sim \text{iid}\,(0, \sigma_w^2)$$

iid.

F.

- If the noise is iid and Gaussian,

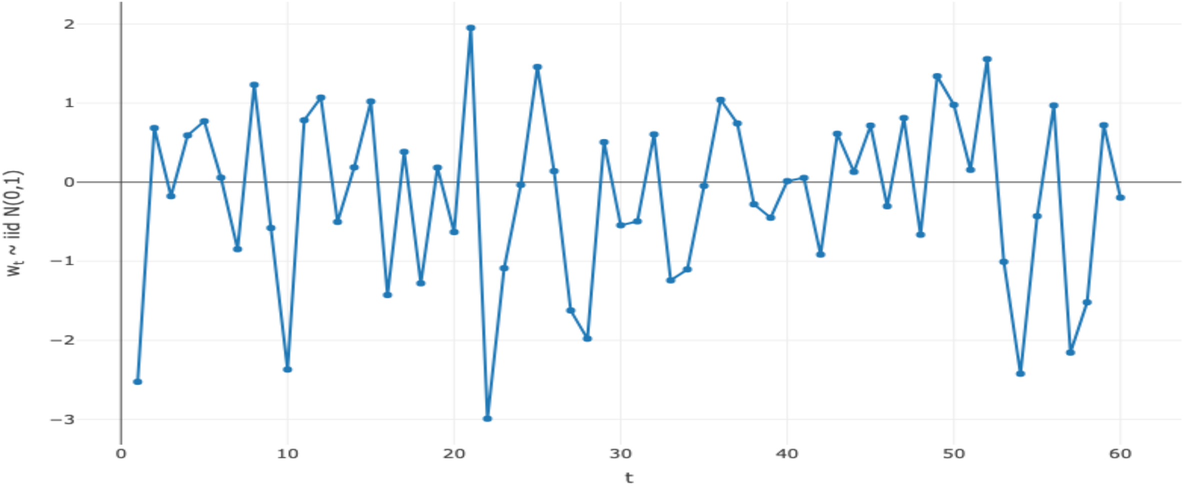
$$W_t \sim \text{iid}\, N(0, \sigma_w^2)$$

Any wn is iid wn.

Any iid wn is wn.

T

White Noise - Plot



White Noise - Importance

$x_t - T_t - S_t - C_t = I_t$

White noise is an important concept in time series analysis and forecasting, mainly, for two reasons.

- Predictability
  - If your time series is white noise, then, by definition, it is random. You cannot reasonably model it and make predictions.
- Model Diagnostics
  - The series of errors from a time series forecast model should ideally be white noise.
  - A sign that model predictions are not white noise is an indication that further improvements to the forecast model may be possible.

IID White Noise - Binary Process

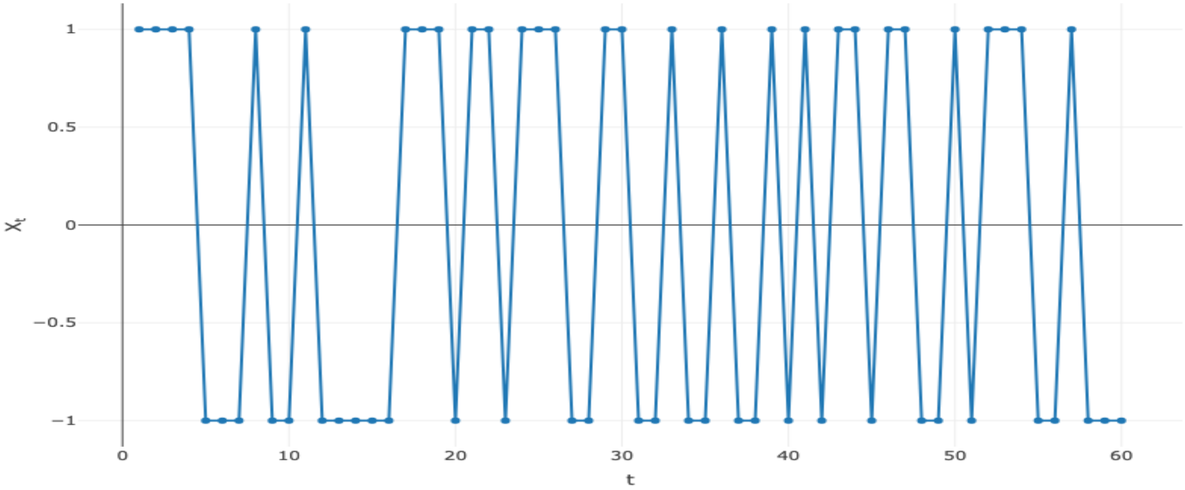
A binary process $\{X_t\}$ is a sequence of iid random variables with

$$P(X_t = 1) = 0.5, \quad P(X_t = -1) = 0.5$$

Some Applications:

- The time series obtained by tossing a coin repeatedly and scoring +1 for each head and -1 for each tail could be modeled as a realization of this process.
- A simple symmetric random walk

Binary Process - Plot



$$X_t = \begin{cases} 1 & p = 0.5 \\ -1 & p = 0.5 \end{cases}$$

$$E(X_t) = (1 * 0.5) + (-1 * 0.5) = 0$$

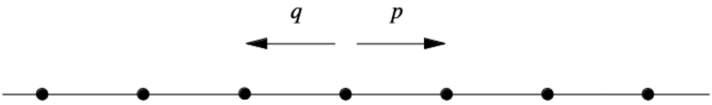
$$X_t = \begin{cases} 1 & p = 0.7 \\ -1 & p = 0.3 \end{cases}$$

Random Walk

- Consider an iid white noise $\{X_t\}$.
- Further consider another stochastic sequence $\{S_t\}$ such that $S_0 = 0$ and

$$S_t = X_1 + X_2 + \cdots + X_t, \quad \text{for } t = 1, 2, \ldots$$

- The process $\{S_t\}$ is known as a random walk.
- If $\{X_t\}$ is a binary process as discussed before, the random walk is a simple symmetric random walk.

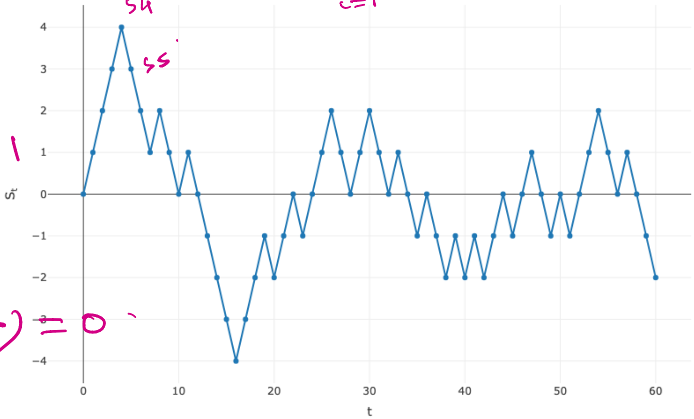


Simple Symmetric Random Walk - Plots

A random walk based on the symmetric binary process is plotted below. Read the following values from the graph:

$$S_t = \sum_{i=1}^{t} x_i$$

- S_2 $= 2$.
- X_2 $\quad S_1 = x_1, \quad S_2 = x_1 + x_2, \quad x_2 = S_2 - S_1 = 1$
- X_5 $= S_5 - S_4 = -1$.
- $E(S_t)$ $= E\left(\sum_{i=1}^{t} x_i\right) = \sum_{i=1}^{t} E(x_i) = 0$.
- $Var(S_t)$



$X_1 \quad X_2 \quad \ldots\ldots \quad X_t.$

$\{x_1 \quad x_2 \quad \ldots \quad x_t.\}$

$x_1 \quad x_2 \quad \ldots\ldots$

$$\mathrm{Var}(S_t) = \mathrm{Var}\left[\sum_{i=1}^{t} X_i\right]$$

indep.

$$= \sum_{i=1}^{t} \mathrm{Var}(X_i)$$

$$= t.$$

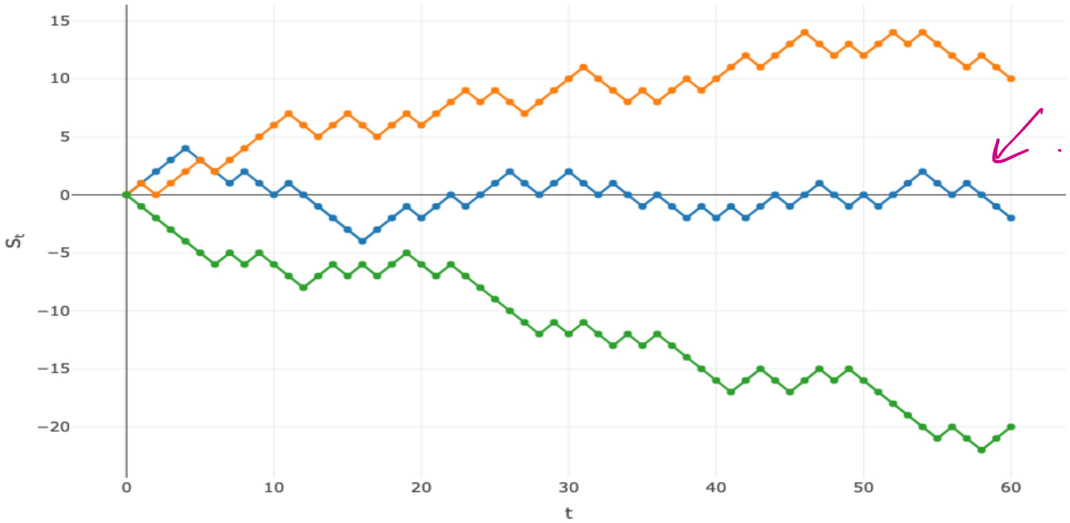
$$\mathrm{Var}(X_i) = E(X_i - \mu)^2$$

$$= E\left[X_i^2\right]$$

$$= \left[(-1)^2 * 0.5\right] + \left[1^2 * 0.5\right]$$

$$= 1.$$

Random Walk - Multiple Paths



Random Walk with Drift

random

$S_t = \sum_{i=1}^{t} X_t$

iid wn

Let

$$X_t = \delta + X_{t-1} + W_t \quad \text{for } t = 1, 2, \ldots$$

with $X_0 = 0$, and $W_t \sim WN(0, \sigma_w^2)$

$X_t = \delta + X_{t-1} + W_t$

$= \delta + (\delta + X_{t-2} + W_{t-1})$

$+ W_t$

$= 2\delta + X_{t-2}$

$+ (W_t + W_{t-1})$

- The constant δ is known as the drift.
- $\delta = 0$ gives us a simple random walk (without drift).
- Recursively, X_t can be rewritten as

$$X_t = \delta t + \sum_{j=1}^{t} W_j$$

Random Walk with Drift - Plot

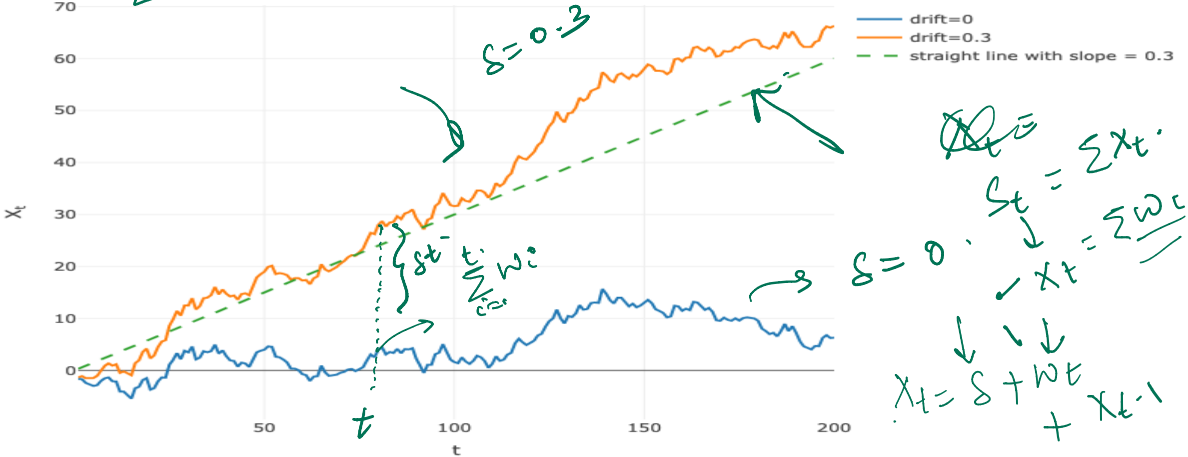
$w_t \sim wn(0, \sigma_w^2 = 1)$

$X_t = \delta + X_{t-1} + w_t$

$N(0, \sigma^2)$

Random walk with and without drift and $\sigma_w^2 = 1$

$E(X_t) = \delta t$



Moving Average Smoother

- Smoothing is the process of removing random variations that appear as coarseness in a plot of raw time series data.
- It reduces the noise to emphasize the signal that can contain trends and cycles.
- Analysts also refer to the smoothing process as filtering the data.
- Moving average is the simplest method of smoothing a series by averaging the neighbors.

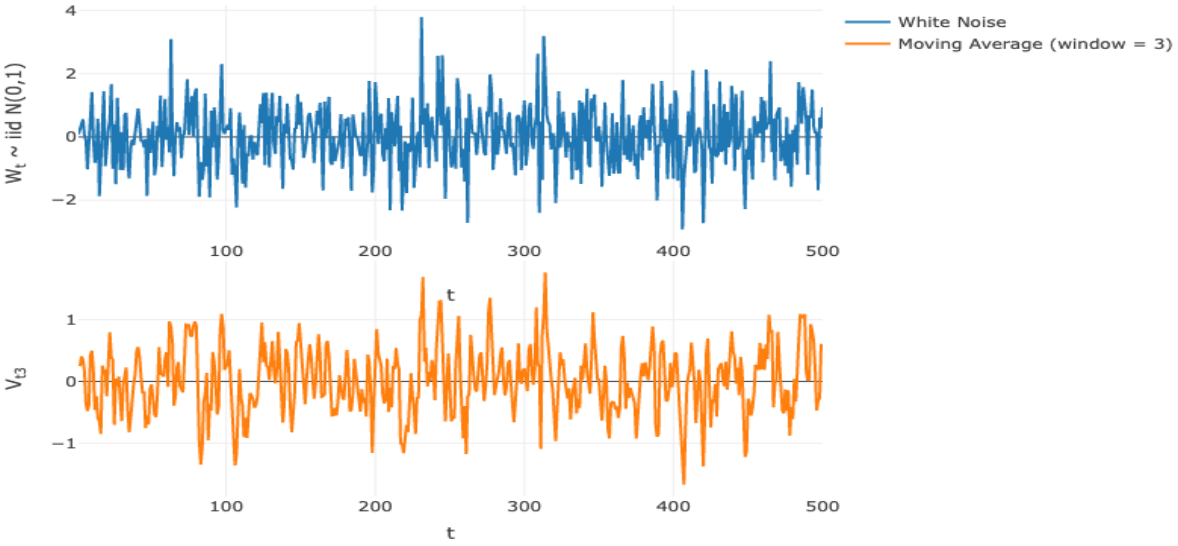
3 Point Moving Average

- Consider a white noise series $\{W_t\}$ with variance σ^2.
- The series $\{W_t\}$ can be replaced by another smooth series given as

$$V_t = \frac{1}{3}(W_{t-1} + W_t + W_{t+1})$$

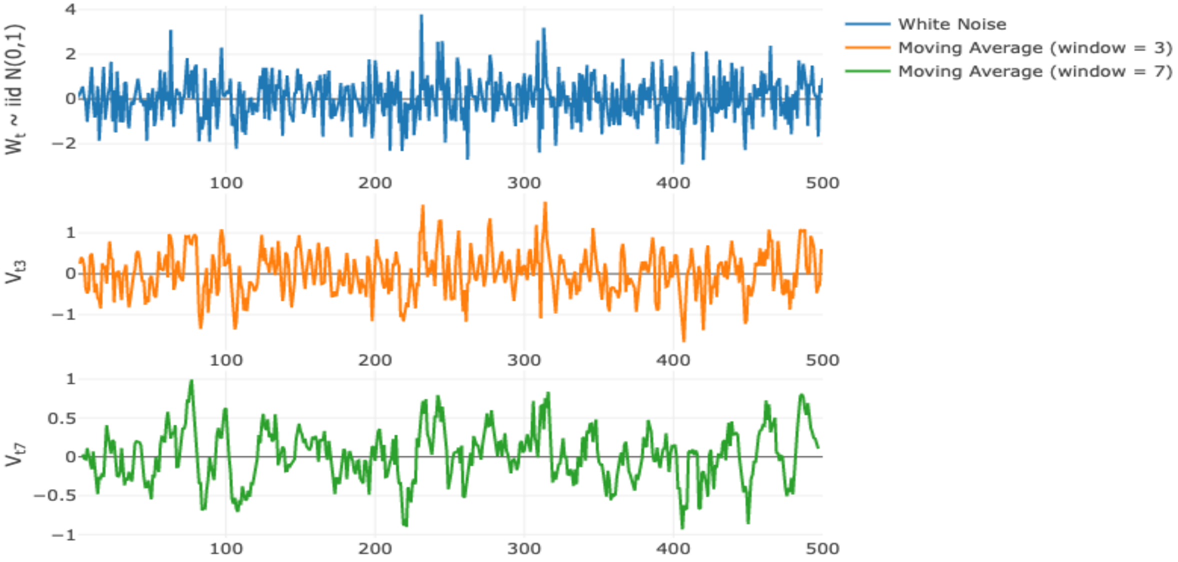
- The series $\{V_t\}$ is (centered/two-sided) three-point moving average of the white noise series $\{W_t\}$.
- Number of points used to calculate the average is known as the 'window' of a moving average.

3 Point Moving Average - Plot



3 Point Moving Average - Plot

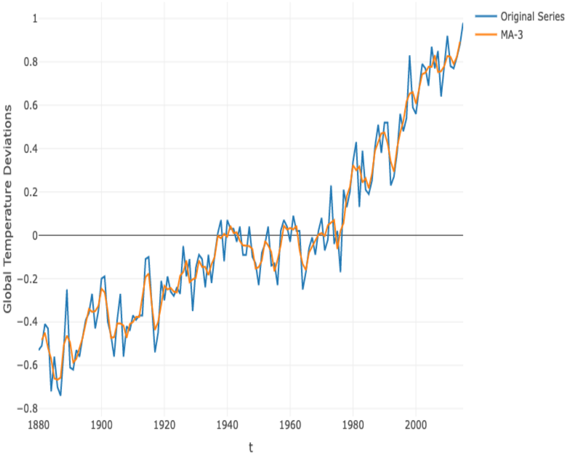
- What do you infer from the plots on the previous slide?
- How do you think the Moving Average plot would change if the window is changed?



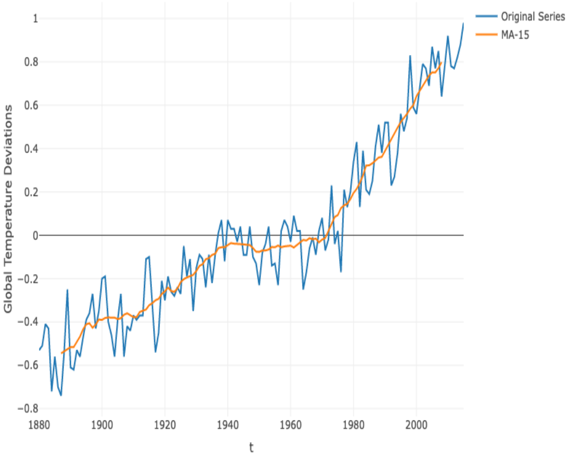
White Noise
Moving Average (window = 3)
Moving Average (window = 7)
$W_t \sim$ iid N(0,1)
V_{t3}
V_{t7}

Global Temperature Deviations - MA

MA - 3



MA - 15



Where do we go from here?

- Dig deeper into statistical characteristics of these time series.
    - Review basics of expectations - mean, variance, covariance, and moments.
- Learn how to use R to do basic time series operations - accessing data, plotting, smoothing, etc.