Lecture 1
Introduction

- **Instructor:**

Prof. Linlin Wang

Reading List

Compulsory Readings

Title	
1	Shumway, R. H., Stoffer D.S.. (2017). Time Series Analysis and Its Application: with R examples. Springer, 2017.
2	Goodfellow, I., Yoshua B., and Aaron C. (2016). Deep learning. MIT press.

Additional Readings

Title	
1	Brockwell, P.J., & Davis, R.A. (2016), Introduction to time series and forecasting. springer.
2	Chollet F., & Allaire J.J. (2018). Deep Learning with R. Manning Publications.

Lecture 1
Introduction

- **Keyword Syllabus**
  - Autoregressive(AR), Moving average(MA), Autoregressive moving average (ARMA) models
  - Parameter estimation
  - Model selection criteria
  - Properties of forecasts
  - Modelling volatility using ARCH and GARCH
  - Artifical neural networks
  - Recurrent neural networks
  - Long short-term memory

Lecture 1
Introduction

- **Prerequisite:**

  Linear algebra

  Mathematical statistics

- **Assessment:**

  25% test; → **Mid-term**

  25% assignments; → **~Small Project**

  50% final exam

- **Computer:**

  R software will be used for examples and exercises

- **Lecture Topics:**

  Introduction, perspective, examples of typical time series problems;

  Characteristics of time series;

  Time series regression;

  ARIMA models;

  Parameter estimation and forecasting;

  Unit root testing;

  ARCH and GARCH models;

  Multivariate ARMAX models;

  Feedforward Neural networks;

  Recurrent neural networks;

  Model training;

Background

- Due to the rapid development of data science technologies, researchers start to analyze useful patterns from time series data.
- This analysis refers to the behavior of summarizing valuable information from some of the data points that are ordered by time.
- Early in 1969, machine learning based approach has been leveraged to analyze time series tasks.
- More applications appear after 1980.

What is time series?

- ***A time series*** **is a series of data points indexed in time order.**
- **Application fields:**

  Economics: daily stock price, GDP, monthly unemployment rates

  Social sciences: population, birth rates, school enrollments

  Epidemiology: number of influenza cases, mortality rates

  Medicine: blood pressure tracking, fMRI

  Physical sciences: global temperatures, monthly sunspot observations

Examples

- Quarterly earnings per share

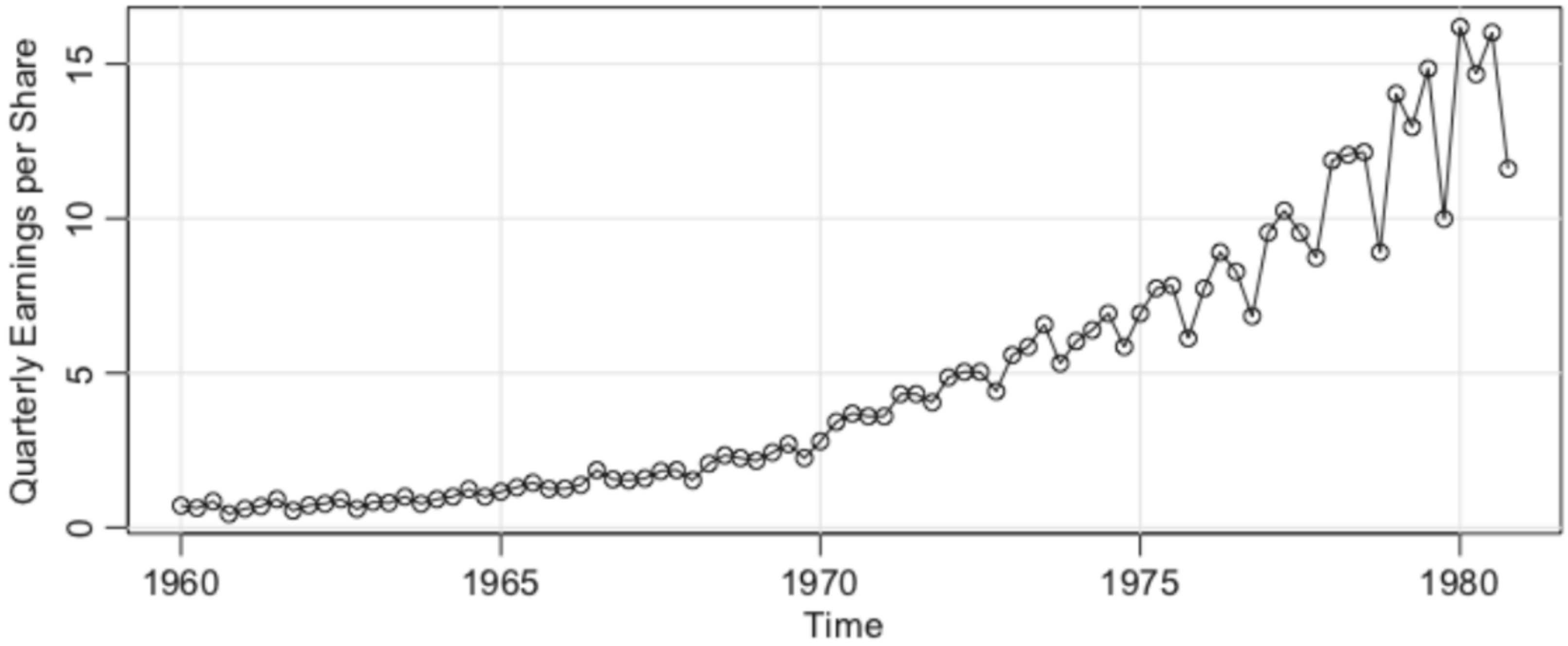


Fig.1. 1 in "Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer. 4th Edition.

Examples

- Global warming

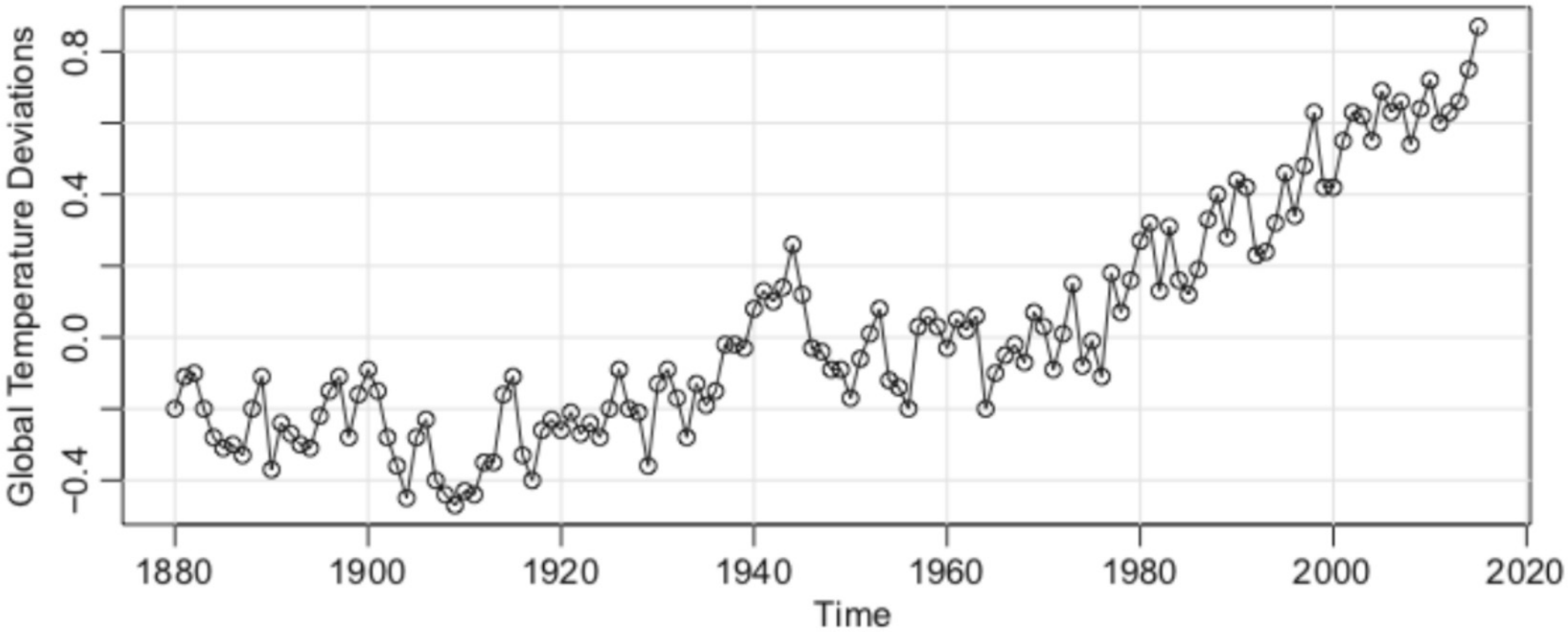
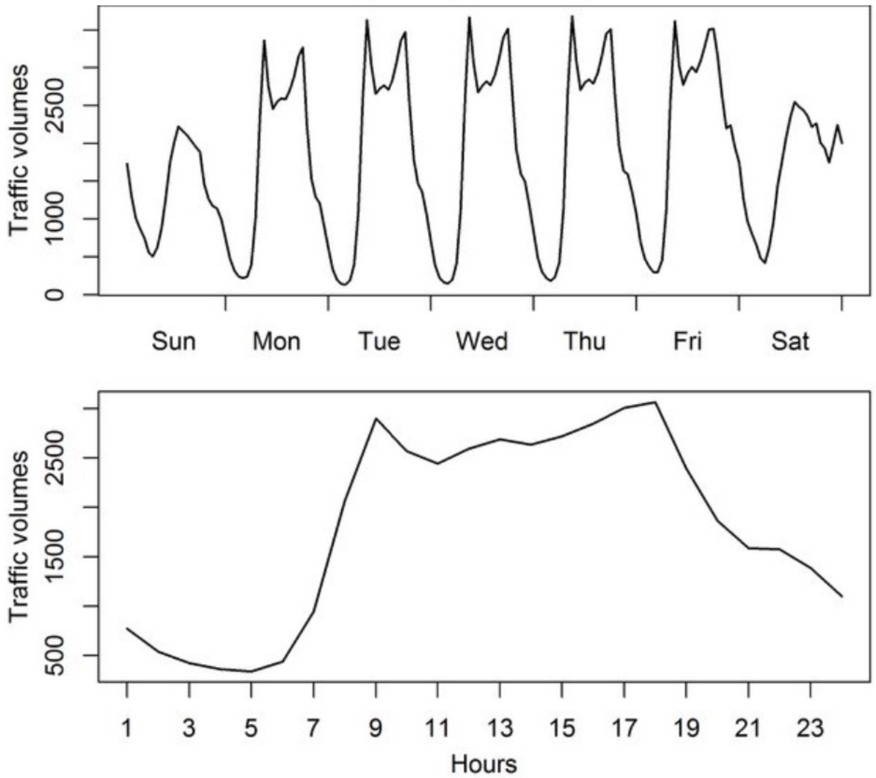


Fig. 1.2 in "Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer. 4th Edition.

Examples

- Traffic volume



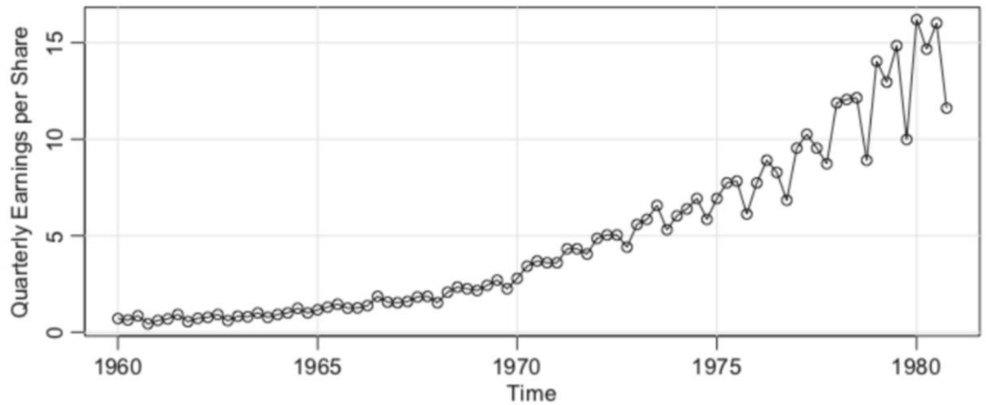
httos://wwwresearchaate.netloublication/336816787 Short-Term Traffic Flow Forecastina A Comoonent Wise Gradient Boostina Aooroach With Hierarchical Reconci/iation/fiaures?lo= 1

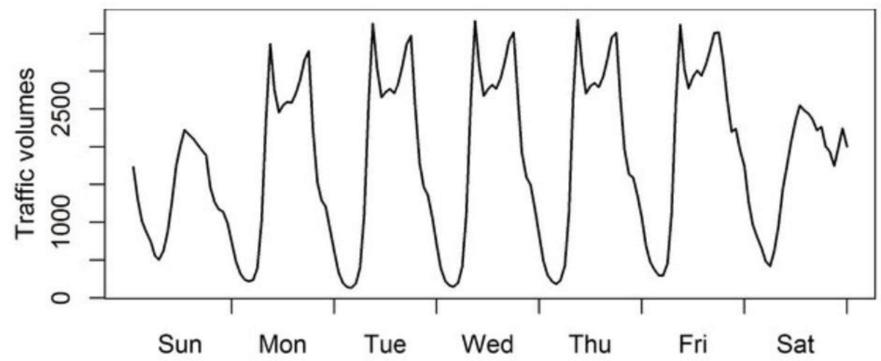
Question:

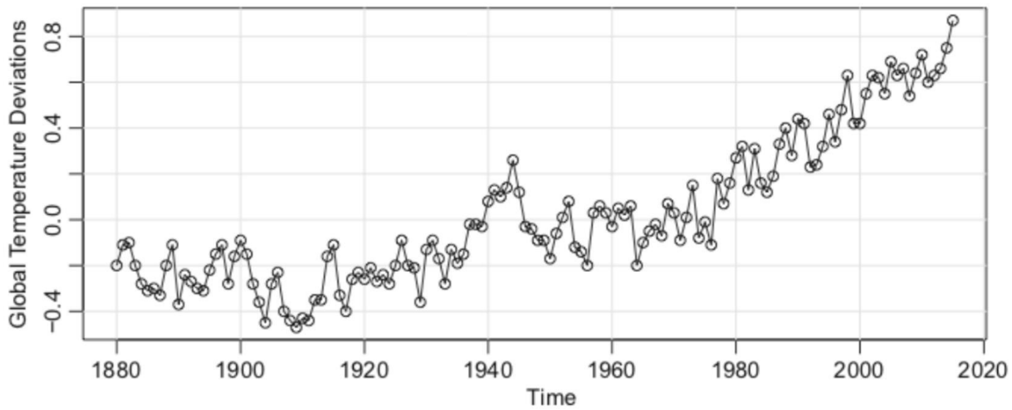
- Have you noticed any differences in term of the “time”?

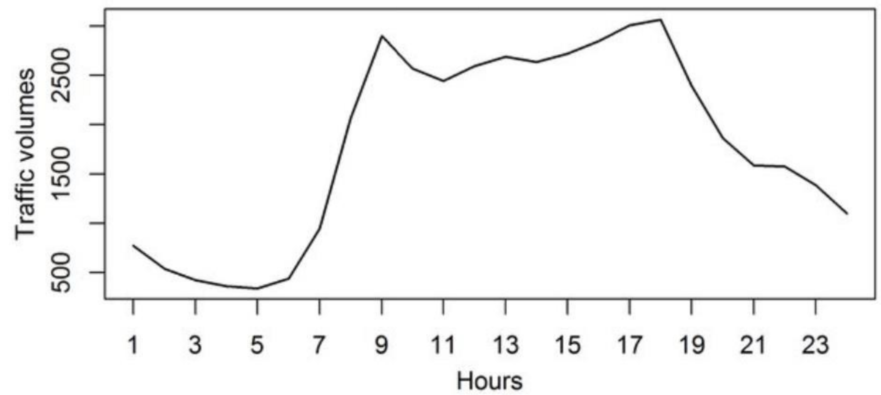
Question:

- Quarterly earnings per share









Time series analysis is very useful:

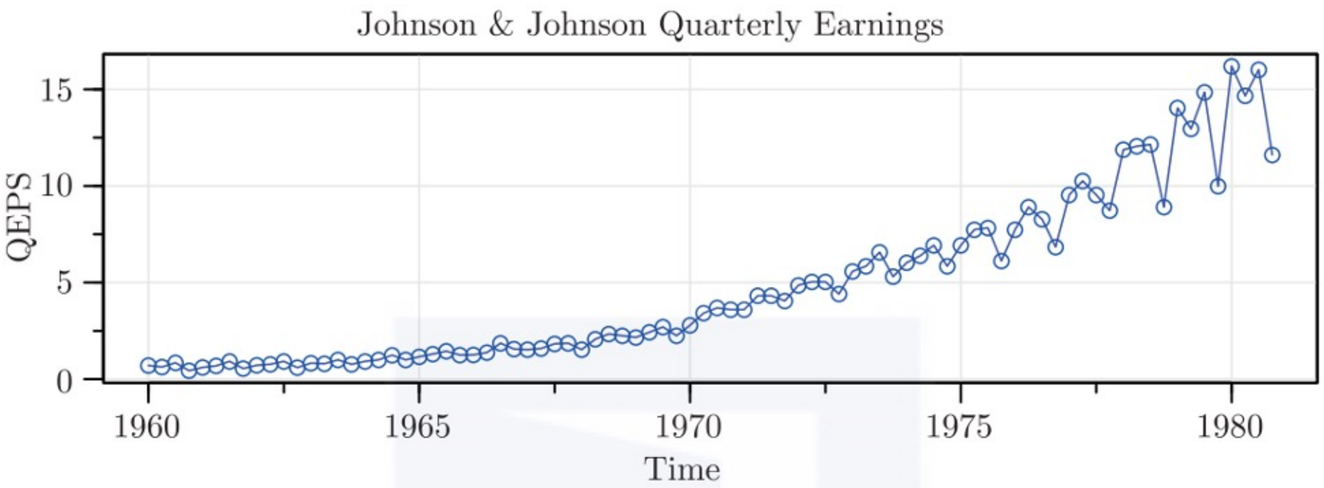
We can conduct “Time-Series Decomposition” to analyze:

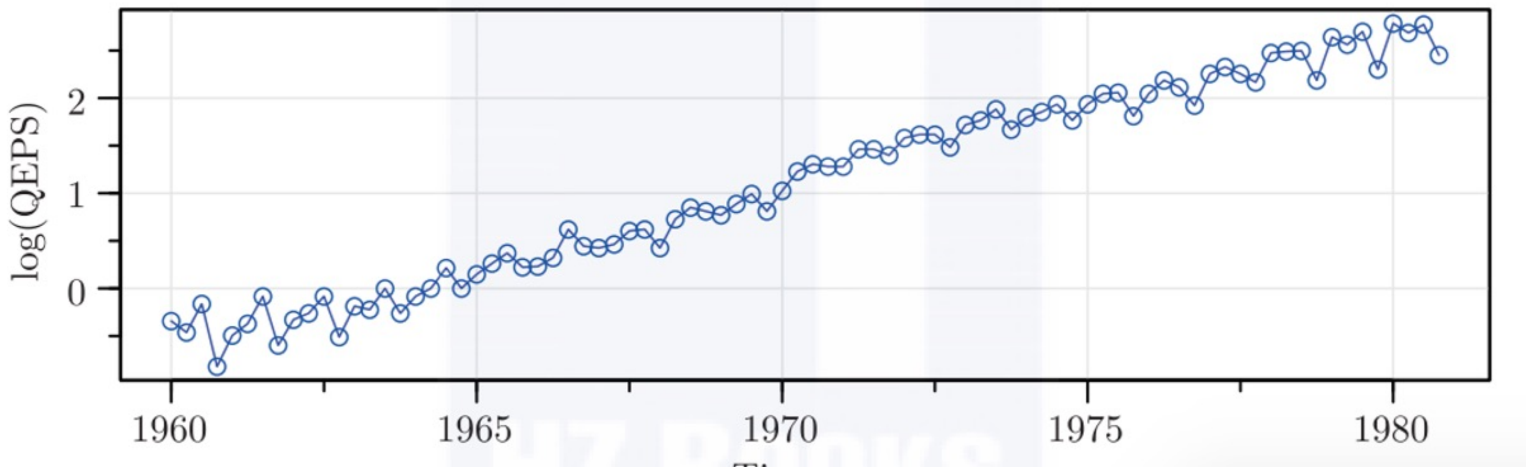
1. Long-term Tendency

2. Seasonal Changing rules

3. Repeated time-series patterns

4. Irregular Changes


Johnson & Johnson Quarterly Earnings
QEPS
0
5
10
15
1960
1965
1970
1975
1980
Time


log(QEPS)
0
1
2
1960
1965
1970
1975
1980

Objectives of time series analysis

- Description and explanation

  Deeper understanding of the mechanism that generated the time series (example: trend and seasonal components)

- Forecasting

  Predicting the future (example: predict unemployment)

- Control

  (example: Impact of monetary policy on unemployment)

- Hypothesis testing

  (example: global warming)

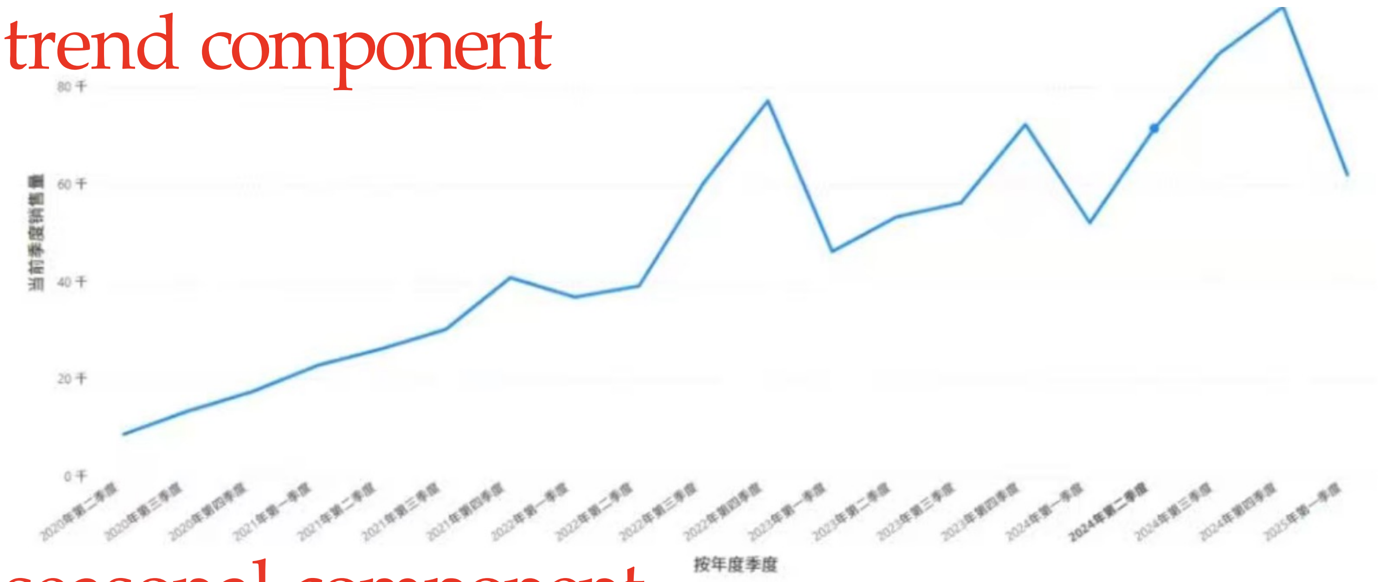
Time series modeling

- **Typically, a time series model can be described as**

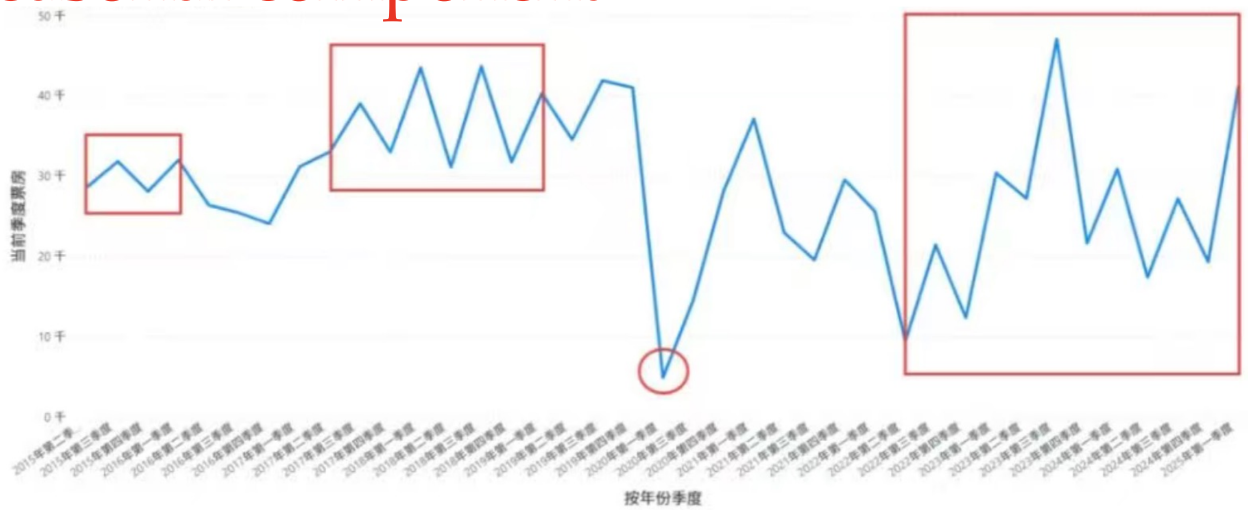
$$x_t = m_t + s_t + e_t$$

trend component   seasonal component   residual

trend component



seasonal component



Time series models

- **A time series can be defined as a collection of random variables indexed according to the order they are obtained in time.**

- **Example:**

  consider a time series as a sequence of random variables, $x1$, $x2$, $x3$, …

  $x1$: the value taken by the series at the first time point

  $x2$: the value taken by the series at the second time point

  $x3$: the value taken by the series at the third time point

  ⋮

Stochastic process

- **A collection of random variables, $\{x_t\}$, indexed by tis referred to as a *stochastic process*.**

- **The stochastic process is a model for the analysis of time series.**

- **The observed values of a stochastic process are referred to as a *realization* of the stochastic process.**

- **An observed time series is considered to be one realization of a stochastic process.**

Explanation

Imagine that you record the temperature at 3 PM every day for a year, and you obtain a curve of temperature changing over time—this is a time series. It is a sequence of chronologically ordered, observed factual data.

A stochastic process, then, is the "rule of nature" or the "God model" that generates this temperature curve.

Explanation

A time series is the specific set of data you actually observe. It is a concrete realization.

A stochastic process is a set of probabilistic rules that theoretically exists and can generate countless possible temperature curves.

White noise

$$x_t = \underbrace{m_t}_{\text{trend component}} + \underbrace{s_t}_{\text{seasonal component}} + \underbrace{e_t}_{\text{residual}}$$

- White noise is a stochastic process that satisfies three very strict conditions:

a. **Zero mean:** At any point in time, the expected value (average) is zero.

b. **Constant variance:** The magnitude of fluctuations is stable and does not change over time.

c. **No autocorrelation**: The value at any given time is entirely unrelated to the value at any other time. Today's value does not influence tomorrow's, and the value one minute from now will not affect the value an hour later.

White noise

- White noise is a stochastic process that satisfies three very strict conditions:

- This means that white noise contains no exploitable information, patterns, or structure. It is "purely random" and "unpredictable."

Example

- **White noise**

  A collection of uncorrelated random variables, w_t, with mean 0 and finite variance $\sigma_\omega{}^2$

  Used as a model for noise in many engineering applications

  Denoted as $w_t \sim \mathrm{wn}(0, \sigma_w^2)$

- **i.i.d. noise**

  A collection of uncorrelated random variables, Wt, with mean O and finite variance

  Denoted as $w_t \sim \mathrm{iid}(0, \sigma_w^2)$

Example

the most important and commonly used special case of white noise.

- **Gaussian white noise**

$$P\{w_t < c_t\} = \Phi(c_t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c_t} e^{-w^2/2} dw$$

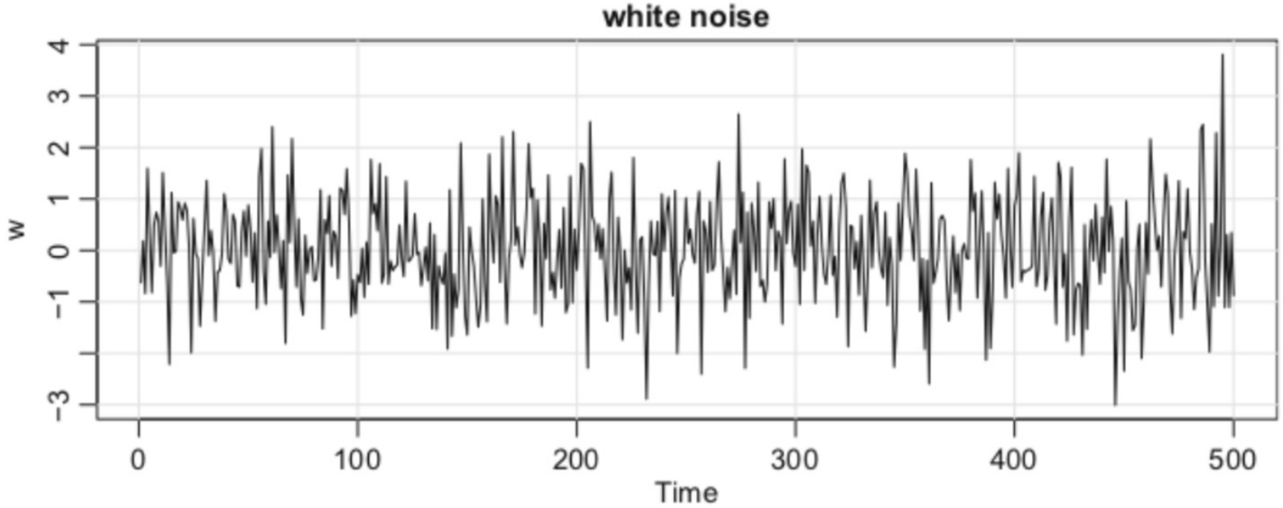


Fig.1.8 (Top) in "Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer. 4th Edition.

- **White Noise:**

Constant Mean: The mean of the sequence does not change over time and is typically zero.

Constant Variance: The degree of fluctuation (variance) in the sequence remains constant over time.

No Autocorrelation: The values at any two distinct time points (t and t+k, where $k \neq 0$) are completely uncorrelated. This means past values have no influence on or predictive power for future values.

- **Gaussian:**

Normal Distribution: Each data point in the sequence follows a normal distribution (also known as Gaussian distribution). This implies that most data points are concentrated near the mean, with extreme values occurring at low and symmetric probabilities.

Stationary Series VS Non-Stationary Series

- **Stationary Time Series:**

  Its statistical properties (such as mean and variance) do not change over time. The behavior of the series does not depend on time.

- **Non-Stationary Time Series:**

  Its statistical properties change over time. The behavior of the series is strongly dependent on time.

A series fluctuating around a fixed value (stationary) vs. a stock price series with a clear upward trend (non-stationary).

Stationary Series VS Non-Stationary Series

- **The problem with non-stationary series:**

Their statistical properties (mean, variance) are constantly changing. This means that patterns from the "past" may not apply at all in the future. For a series whose mean is 100 today and 110 tomorrow, you cannot use a fixed parameter to describe and predict it.

- **The advantage of stationary series:**

Their mean and variance are fixed. This means the behavioral patterns of the series are consistent over time. Patterns we learn from historical data (e.g., if the previous value is high, the next value tends to fall back) remain applicable in the future, making reliable prediction possible.

Example: selling ice cream in the summer.

- Suppose you own an ice cream shop and have recorded the number of ice creams sold daily over the past three years.

- **Plot the data:**

  The graph will clearly show **seasonality**: sales peak significantly every summer (July and August), forming sharp spikes, while sales drop to a trough every winter (December and January).

  This "repeating pattern of peaks and valleys" is also a **non-stationary** feature! Because its values are strongly correlated with time.

- **How to make it stationary?**

- **How to make it stationary? — "Seasonal Differencing"**
  Instead of looking at "how many were sold today," we calculate **"how many more (or fewer) were sold today compared to the same day last year."**

  This is **seasonal differencing** (with a period of 365 days):

  Difference = Sales on day 200 of this year - Sales on day 200 of last year
- Now, if we examine this "difference" series, it will likely become stationary because it removes the fixed seasonal effect that repeats every year.

Example

- **Moving averages and filtering**

$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$$

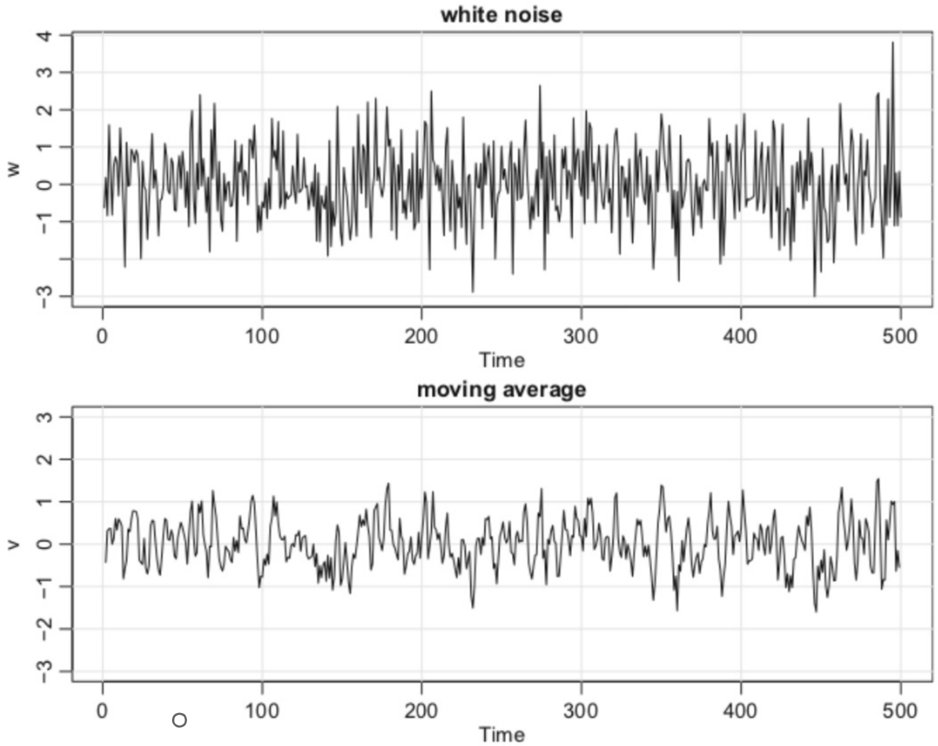


Fig.1.8 in "Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer. 4th Edition.

Moving averages

Main purposes:
Smoothing data: Removing tiny, random "noise."

- Suppose you have stock prices for five consecutive days: [100, 102, 101, 105, 103].
  To calculate a 3-day moving average:
- Day 3 average = (100 + 102 + 101) / 3 = 101
- Day 4 average = (102 + 101 + 105) / 3 ≈ 102.67
- Day 5 average = (101 + 105 + 103) / 3 = 103
- Thus, the original [100, 102, 101, 105, 103] becomes [-, -, 101, 102.67, 103] after applying the 3-day moving average.

A moving average is a computational method used to process data series by smoothing out short-term fluctuations and highlighting long-term trends or cycles.

Example

- **Random walk with drift**

$$x_t = \delta + x_{t-1} + \omega_t$$

1. Random Walk
Imagine a drunkard walking down a street. The direction of each step he takes (forward or backward) is completely random, as if decided by a coin toss. His next position **depends only on his current position**, plus a random step size.

Example

- **Random walk with drift**

$$x_t = \delta + x_{t-1} + \omega_t$$

1. Random Walk

Imagine a drunkard walking down a street. The direction of each step he takes (forward or backward) is completely random, as if decided by a coin toss. His next position **depends only on his current position**, plus a random step size.

- **2. Drift**
- Now, give this drunkard a slight preference or tendency. For example, he is on a slightly sloping ramp, or subconsciously wants to move in a certain direction. This way, not only does he take random steps, but **on average, he will also consistently move in a specific direction**.

$$Y_t = \delta + Y_{t-1} + \varepsilon_t$$

Not only does this drunkard take random steps, but he also has a loose rope tied to him, constantly and gently pulling him in a certain direction (e.g., north) by an unseen force. Thus, his overall movement path is:
(A continuous northward pull) + (The drunkard's own random, stumbling steps).

Example

- **Random walk with drift**

$$x_t = \delta + x_{t\text{-}1} + \omega_t$$

$$\begin{aligned} x_t &= \delta + (\delta + x_{t-2} + w_{t-1}) + w_t \\ &= 2\delta + x_{t-2} + w_{t-1} + w_t \\ &= 2\delta + (\delta + x_{t-3} + w_{t-2}) + w_{t-1} + w_t \\ &= 3\delta + x_{t-3} + w_{t-2} + w_{t-1} + w_t = \cdots \\ &= \delta t + \sum_{j=1}^{t} w_j \end{aligned}$$

trend component

Example

- **Random walk with drift**

$$x_t = \delta + x_{t-1} + w_t \quad \Rightarrow \quad x_t = \delta t + \sum_{j=1}^{t} w_j$$

drift, $\delta = 0$: random walk

trend component

$$\nabla x_t = x_t - x_{t-1} = \delta + w_t$$

Example

- **Random walk with drift**

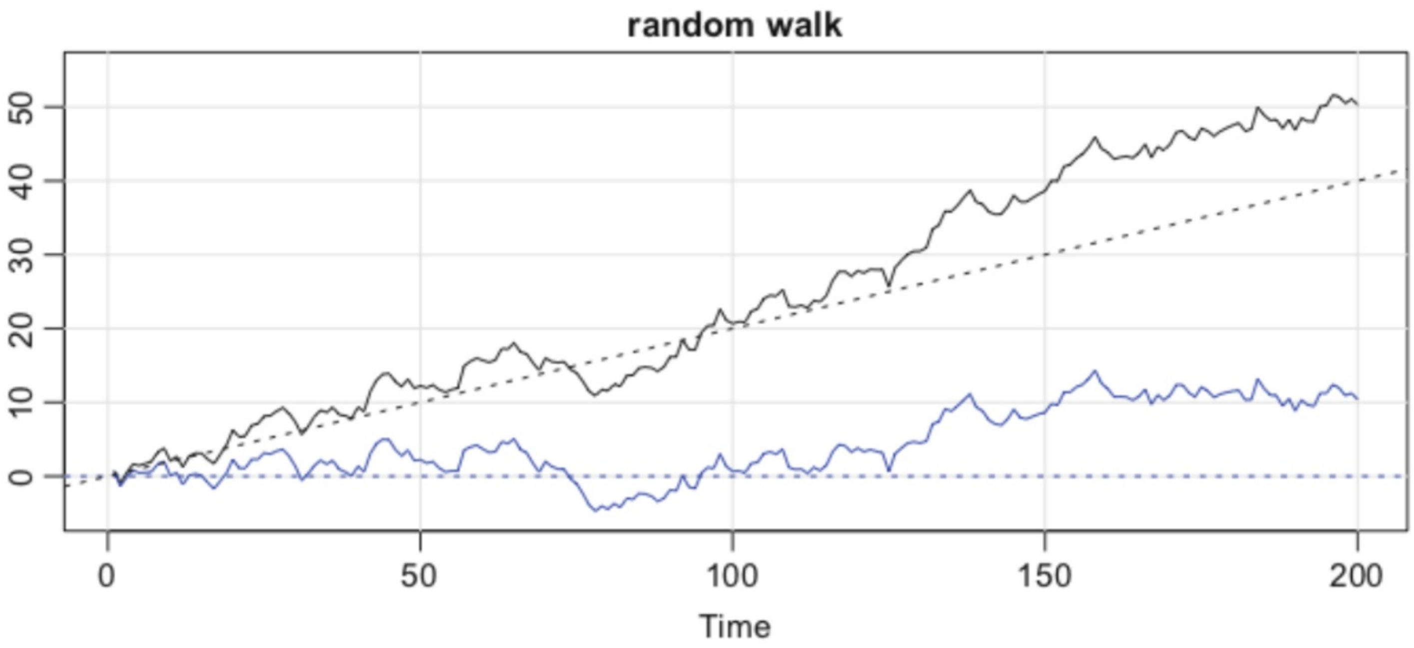


Fig. 1.10 in "Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer. 4th Edition.

Example

- **Signal in noise**

$x_t = A\,cos(2\pi\omega t + \Phi) + \omega_t$

underlying signal (seasonal component)

Example

- **Signal in noise**

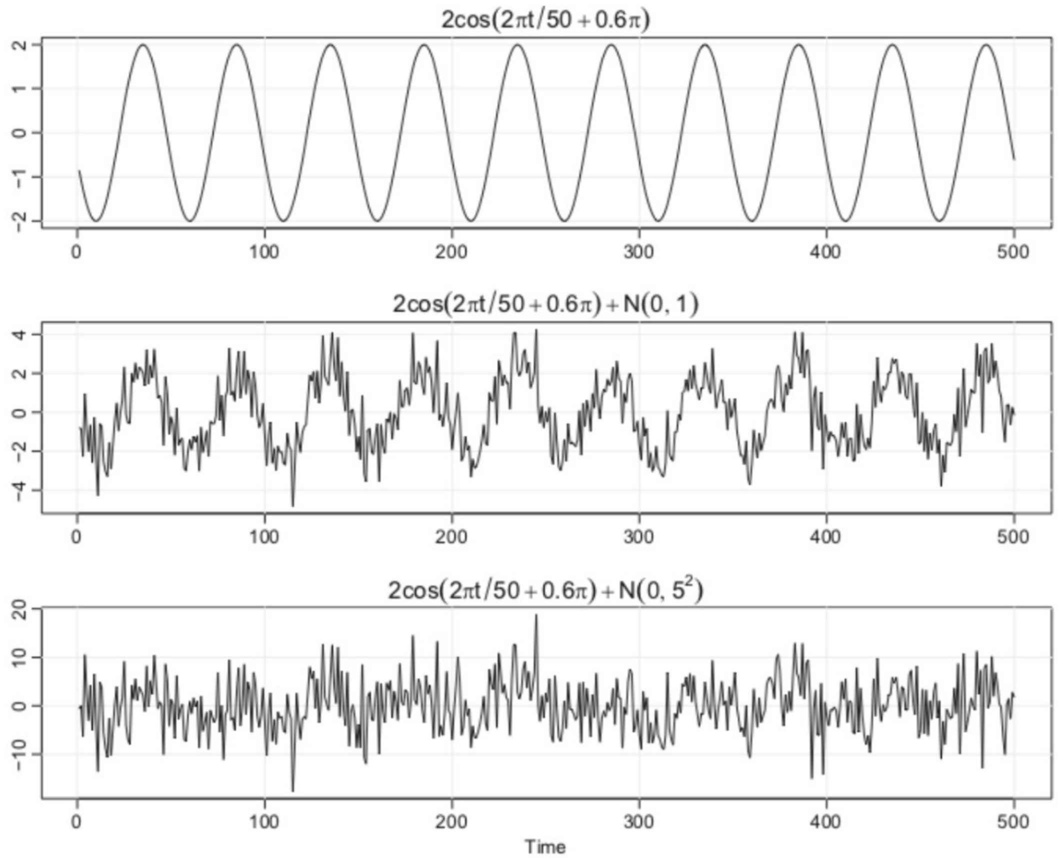


Fig. 1.10 in "Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer. 4th Edition.

Time series modeling

- **Typically, a time series model can be described as**

$$x_t = m_t + s_t + e_t$$

trend component   seasonal component   residual

Again!

Time series modeling

- **Plot the time series**

  Look for trends, seasonal components, step changes, etc.

- **Transform the data so that the residuals are *stationary***

  Estimate and subtract trend and seasonal components.

- **Fit model to residuals**

  The purpose of this process is to examine whether patterns that can be modeled still exist in the residuals.

Measure of dependence

- **Mean function of $\{x_t\}$**

$$\mu_t = \mathrm{E}(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx \text{ (provided it exist)}$$

Used to quantify the correlation between current values and past values in a time series, reflecting the dynamic dependency relationships within the sequence.

Core Question: When one variable changes, does another variable also change in a predictable way?
Purpose: To assess the existence and strength of this relationship.
Key Point: It describes the relationship between variables.

Mean function

- **Examples:**

$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$$

$$\mathrm{E}(v_t)?$$

Mean function

- **Examples:**

$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$$

$$\mathrm{E}(v_t)?$$

$$\begin{aligned}\mathrm{E}(v_t) &= \frac{1}{3}\mathrm{E}(w_{t-1} + w_t + w_{t+1}) \\ &= \frac{1}{3}\mathrm{E}(w_{t-1}) + \frac{1}{3}\mathrm{E}(w_t) + \frac{1}{3}\mathrm{E}(w_{t+1}) \\ &= 0\end{aligned}$$

Mean function

- **Examples:**

$$x_t = \delta t + \sum_{j=1}^{t} w_j$$

$\mathrm{E}(x_t)$?

Mean function

- **Examples:**

$$x_t = \delta t + \sum_{j=1}^{t} w_j$$

$\mathrm{E}(x_t)$?

$$\begin{aligned}
\mathrm{E}(x_t) &= \mathrm{E}\left(\delta t + \sum_{j=1}^{t} w_j\right) \\
&= \delta t + \sum_{j=1}^{t} \mathrm{E}(w_j) \\
&= \delta t
\end{aligned}$$

Mean function

- **Examples:**

$$x_t = A\cos(2\pi\omega t + \phi) + w_t$$

$$\mathrm{E}(x_t)?$$

Mean function

- **Examples:**

$$x_t = A\cos(2\pi\omega t + \phi) + w_t$$

$\mathrm{E}(x_t)$?

$$\begin{aligned}
\mathrm{E}(x_t) &= \mathrm{E}(A\cos(2\pi\omega t + \phi) + w_t) \\
&= A\cos(2\pi\omega t + \phi) + \mathrm{E}(w_t) \\
&= A\cos(2\pi\omega t + \phi)
\end{aligned}$$

Measure of dependence

- **Autocovariance function of $\{x_t\}$**

$$\gamma(s,t) = \mathrm{Cov}(x_s, x_t) = \mathrm{E}[(x_s - \mu_s)(x_t - \mu_t)]$$

Measures the linear dependence between two points on the same series observed at different times

Smooth series: stay large even when t and s are far apart

Choppy series: nearly zero for large separations

$\gamma(s,t) = 0$: x_t and x_s are not linearly related

$s = t$: reduces to variance

Autocovariance function

- **Examples:**

  white noise $\{w_t\}$

  $\gamma(s, t)$?

Autocovariance function

- **Examples:**

  white noise $\{w_t\}$

  $\gamma(s,t)$?

$$\gamma(s,t) = \text{cov}(w_s, w_t) = \begin{cases} \sigma_w^2 & s = t \\ 0 & s \neq t \end{cases}$$

Autocovariance function

- **Examples:**

$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$$

$$\gamma(s,t)?$$

Autocovariance function

- **Examples:**

$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$$

$$\gamma(s,t)?$$

$$\begin{aligned}\gamma(s,t) &= \mathrm{cov}(v_s, v_t) \\ &= \mathrm{cov}\left\{\frac{1}{3}(w_{s-1} + w_s + w_{s+1}), \frac{1}{3}(w_{t-1} + w_t + w_{t+1})\right\}\end{aligned}$$

Autocovariance function

- **Examples:**

$$
\begin{aligned}
\gamma(s,t) &= \mathrm{cov}(v_s, v_t) \\
&= \mathrm{cov}\left\{\frac{1}{3}(w_{s-1} + w_s + w_{s+1}), \frac{1}{3}(w_{t-1} + w_t + w_{t+1})\right\}
\end{aligned}
$$

$$s = t:$$

$$
\begin{aligned}
\gamma(t,t) &= \frac{1}{9}(\mathrm{cov}(w_{t-1}, w_{t-1}) + \mathrm{cov}(w_t, w_t) + \mathrm{cov}(w_{t+1}, w_{t+1})) \\
&= \frac{3}{9}\sigma_w^2
\end{aligned}
$$

Autocovariance function

- **Examples:**

$$\gamma(s,t) = \text{cov}(v_s, v_t)$$
$$= \text{cov}\left\{\frac{1}{3}(w_{s-1} + w_s + w_{s+1}), \frac{1}{3}(w_{t-1} + w_t + w_{t+1})\right\}$$

$s = t + 1$ (same logic when $s = t - 1$) :

$$\gamma(t+1,t) = \frac{1}{9}\text{cov}\{(w_t + w_{t+1} + w_{t+2}), (w_{t-1} + w_t + w_{t+1})\}$$
$$= \frac{1}{9}[\text{cov}(w_t, w_t) + \text{cov}(w_{t+1}, w_{t+1})]$$
$$= \frac{2}{9}\sigma_w^2$$

Autocovariance function

- **Examples:**

$$\begin{aligned}\gamma(s,t) &= \operatorname{cov}(v_s, v_t)\\ &= \operatorname{cov}\left\{\frac{1}{3}(w_{s-1}+w_s+w_{s+1}), \frac{1}{3}(w_{t-1}+w_t+w_{t+1})\right\}\end{aligned}$$

$$s = t+2 \text{ (same logic when } s = t-2\text{) :}$$

$$\begin{aligned}\gamma(t+2,t) &= \frac{1}{9}\operatorname{cov}\{(w_{t+1}+w_{t+2}+w_{t+3}), (w_{t-1}+w_t+w_{t+1})\}\\ &= \frac{1}{9}\operatorname{cov}(w_{t+1}, w_{t+1})\\ &= \frac{1}{9}\sigma_w^2\end{aligned}$$

Autocovariance function

- **Examples:**

$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$$

$$\gamma(s,t)?$$

$$\gamma(s,t) = \begin{cases} \frac{3}{9}\sigma_w^2 & s = t, \\ \frac{2}{9}\sigma_w^2 & |s-t| = 1, \\ \frac{1}{9}\sigma_w^2 & |s-t| = 2, \\ 0 & |s-t| > 2 \end{cases}$$

Autocovariance function

- **Examples:**

$$\text{random walk } x_t = \sum_{j=1}^{t} w_j$$

$$\gamma(s, t)?$$

Autocovariance function

- **Examples:**

$$\text{random walk } x_t = \sum_{j=1}^{t} w_j$$

$$\gamma(s,t)?$$

$$\gamma(s,t) = \mathrm{cov}(x_s, x_t) = \mathrm{cov}\left(\sum_{j=1}^{s} w_j, \sum_{k=1}^{t} w_k\right)$$

$$= \min\{s,t\}\sigma_w^2$$