



UNIVERSITÀ  
DEGLI STUDI  
DI TRIESTE

# Statistics

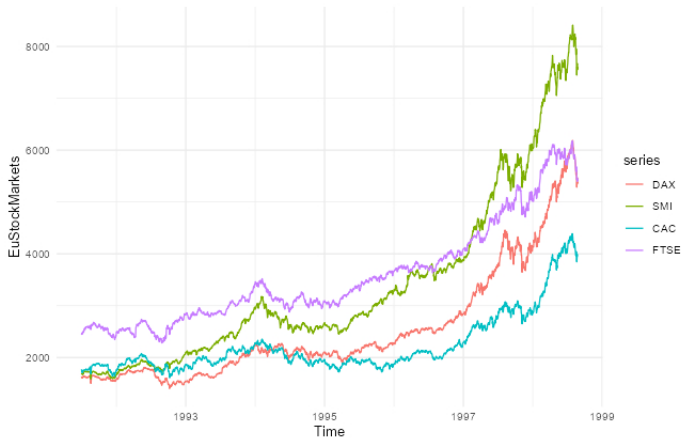
Time Series

Sara Geremia  
May 9th, 2024



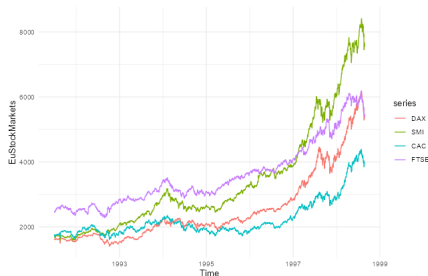
# Time Series

- ▶ A set of observations indexed by time  $t$ :  $Y_1, Y_2, \dots, Y_T$
- ▶ Discrete and continuous time series



# Time Series

- ▶ The interval between observations — that is, the period of time between observation  $t$  and observation  $t + 1$  — is some unit of time such as weeks, months, or years.
- ▶ The date  $t = 1$  corresponds to the first date in the data set, and  $t = T$  corresponds to the final date in the data set.



The data plotted relates to the **daily** closing prices of major European stock indices from 1991-1998.

# Time Series Notation and Terminology

- ▶ The change in the value of  $Y$  between period  $t - 1$  and period  $t$  is called the first difference in the variable  $Y_t$ .
- ▶  $\Delta$  is used to represent the first difference, so  $\Delta Y_t = Y_t - Y_{t-1}$ .
- ▶ The value of  $Y$  in the previous period (relative to the current period,  $t$ ) is called its first lagged value (or, first lag) and is denoted  $Y_{t-1}$ .
- ▶ Its  $j$ th lagged value (or,  $j$ th lag) is its value  $j$  periods ago, which is  $Y_{t-j}$ .
- ▶ Similarly,  $Y_{t+1}$  denotes the value of  $Y$  one period into the future.

# Missing Values

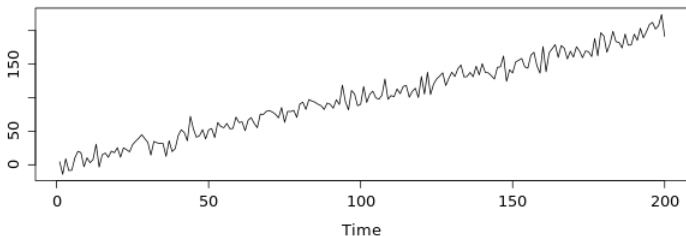
Sometimes there are missing values in time series data.

It is common to replace missing values with the mean of the observed values.

While mean imputation is simple and easy to implement, it often fails in accurately approximating what's really going on due to its inability to account for temporal patterns, variability, relationships between variables, and potential biases. More sophisticated imputation methods are generally required to capture the complexities of such datasets.

# Linear Trend

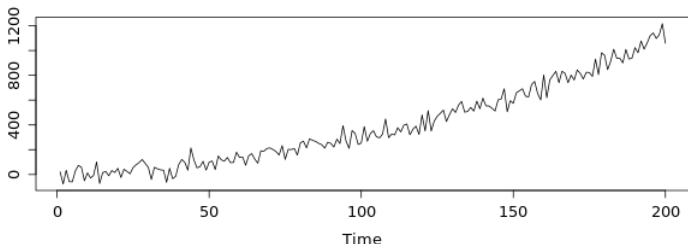
An example of a series with **Linear Trend** over time.



Data points show a consistent upward or downward movement at a constant rate. This type of trend can be represented by a straight line when plotted on a graph.

# Upward Trend

Upward trends may be increasing more quickly than linear.



Rapid decay is also a possibility, but it is not as common in most applications.

# Periodic Trend

Examples of series with **periodic trends** .

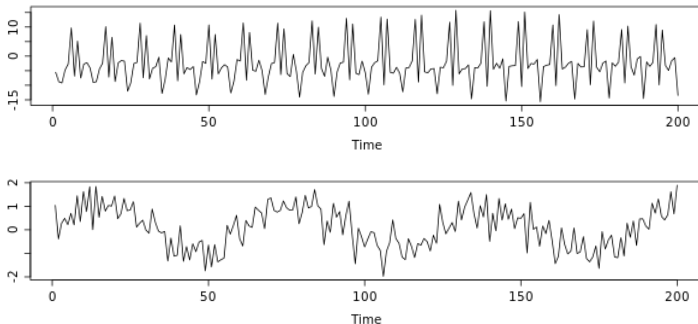
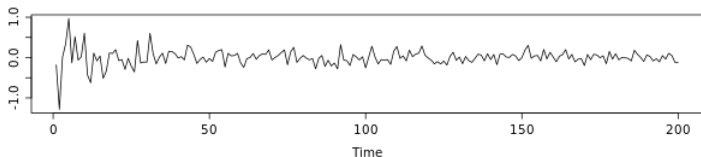


Exhibit patterns that repeat at regular intervals. The key characteristic of such a series is the regularity of peaks and troughs over specific periods.



# Trend in Variability

An example of a series with **trend in variability** .



It refers to changes in the fluctuation or spread of data points over time. This can manifest as either increasing or decreasing variability.

# Sample transformations: $\log()$

Data transformations are powerful tools in time series analysis, used to stabilize variance, linearize relationships, and make a dataset more amenable to modeling.

## Logarithm Transformations

### ► Purpose:

- Linearize exponential growth trends.
- Stabilize variance, particularly for series with increasing variance over time

### ► How It Works:

- Reduces the impact of large values, compressing the range of the data and making exponential growth trends appear linear over time.

# Sample transformations: `diff()`

## First Difference Transformations

### ► Purpose:

- Remove linear trends from a time series.

### ► How It Works:

- The first difference transformation involves subtracting each observation from the subsequent observation. This results in a new series of differences, which represents the increments or changes between consecutive observations in the original series.

Seasonal differencing can be applied to remove periodic trends; lag  $s$  difference transformation. For monthly or quarterly data, an appropriate value of  $s$  would be 12 or 4, respectively. The first difference transformation function has  $s = 1$ .

# Autocorrelation

- ▶ In time series data, the value of  $Y$  in one period typically is correlated with its value in the next period.
- ▶ This correlation of a series with its own lagged values is called autocorrelation or serial correlation.
- ▶ The first autocorrelation is the correlation between  $Y_t$  and  $Y_{t-1}$ —that is, the correlation between the value of  $Y$  at time  $t$  and its value in the previous period.

## Autocorrelation

The autocorrelation coefficient at lag  $k$  is given by:

$$\rho_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

# Autocorrelation

- ▶ Autocorrelation can reveal important patterns in time series data, such as trends, cycles, and seasonal effects.
- ▶ It is crucial for understanding the underlying structure of the data and for building accurate predictive models.
- ▶ High positive autocorrelation indicates that high values of  $Y$  tend to be followed by high values, and low values tend to be followed by low values.
- ▶ Negative autocorrelation indicates that high values of  $Y$  tend to be followed by low values, and vice versa.

# Autocovariance

- ▶ Autocovariance measures the degree to which a time series is linearly related to a lagged version of itself.
- ▶ The autocovariance at lag  $k$  quantifies the covariance between  $Y_t$  and  $Y_{t-k}$ .

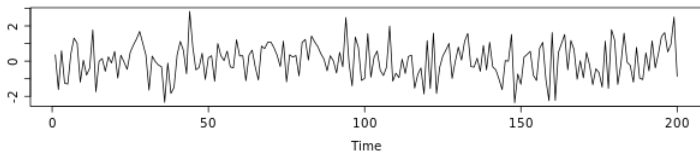
## Autocovariance

The autocovariance at lag  $k$  is given by:

$$\gamma_k = \frac{1}{n} \sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})$$

# Stationary Time Series

Some time series do not exhibit any clear trends over time.



A time series is considered **stationary** if its statistical properties do not change over time.

# Stationary Time Series

- ▶ Specifically, a time series is stationary if:
  - ▶ The mean is constant over time.
  - ▶ The variance is constant over time.
  - ▶ The autocovariance depends only on the lag between observations, not on the actual time at which the covariance is computed.

## Stationary Time Series

A time series  $\{Y_t\}$  is stationary if for all  $t$ :

$$E(Y_t) = \mu$$

$$\text{Var}(Y_t) = \sigma^2$$

$$\text{Cov}(Y_t, Y_{t+k}) = \gamma_k$$



# Stationary Time Series

- ▶ Stationarity is a key assumption in many time series models and methods, such as ARIMA.
- ▶ If a time series is not stationary, it can often be transformed to achieve stationarity through differencing, detrending, or other methods.
- ▶ Testing for stationarity can be done using statistical tests.

# White Noise (W-N) Time Series

- ▶ The White Noise (W-N) model is the simplest example of a stationary process.
- ▶ In a W-N model, the time series consists of a sequence of uncorrelated random variables.
- ▶ Each random variable has a mean of zero and a constant variance.

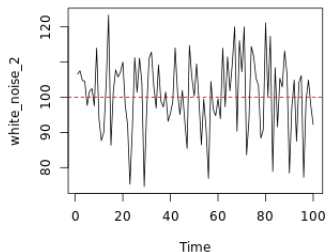
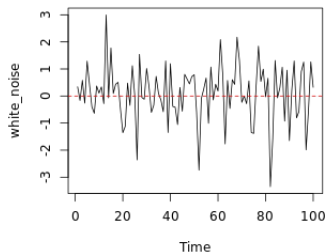
## White Noise

A time series  $\{Y_t\}$  is called white noise if:

$$Y_t \sim \text{i.i.d. } (0, \sigma^2)$$

where  $E(Y_t) = 0$ ;  $\text{Var}(Y_t) = \sigma^2$  and  $\text{Cov}(Y_t, Y_{t+k}) = 0$ .

# White Noise (W-N) Time Series



- ▶ **Left Figure:** The series shows no pattern or correlation in the data over time, indicating white noise with zero mean and constant variance.
- ▶ **Right Figure:** The series has shifted up, indicating a larger mean and exhibits more vertical variability, signifying a larger variance. However, there are still no patterns or trends over time.

# Random Walk (R-W) Time Series

A Random Walk (R-W) is a simple example of an unstable or non-stationary process.

## Random Walk

A random walk can be defined as:

$$X_t = X_{t-1} + \epsilon_t$$

where  $\epsilon_t$  is a white noise process:

$$\epsilon_t \sim \text{i.i.d. } (0, \sigma^2)$$

# Random Walk (R-W) Time Series

- ▶ **Characteristics:**

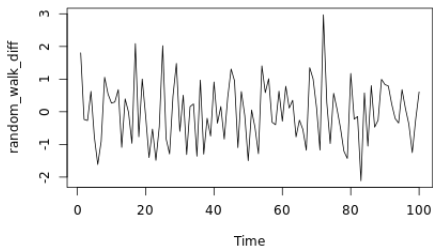
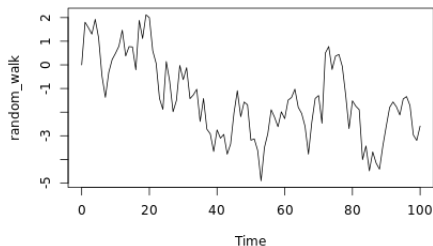
- ▶ **No Specified Mean or Variance:** Unlike stationary processes, a random walk does not have a defined mean or variance.
- ▶ **Strong Dependence Over Time:** Each observation in a random walk is closely related to its immediate predecessor, resulting in a very strong time dependency.

- ▶ **White Noise Increments:** The changes or increments  $\epsilon_t$  follow a white noise process, which is stable and stationary.

- ▶ **Implications:**

- ▶ The random walk's path can drift widely over time, lacking a central tendency or predictable long-term behavior.
- ▶ Despite the non-stationarity of the overall process, the increments  $\epsilon_t$  remain stationary and stable.

# Random Walk (R-W) Time Series



- ▶ **Left Figure:** In the series, each observation is strongly dependent on its immediate predecessor, resulting in a path that curves over time without a fixed mean or variance.
- ▶ **Right Figure:** The first difference of the random walk series is a white noise. This is because a random walk is simply recursive white noise data. By removing the long-term trend, you end up with simple white noise.