UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

Statistics

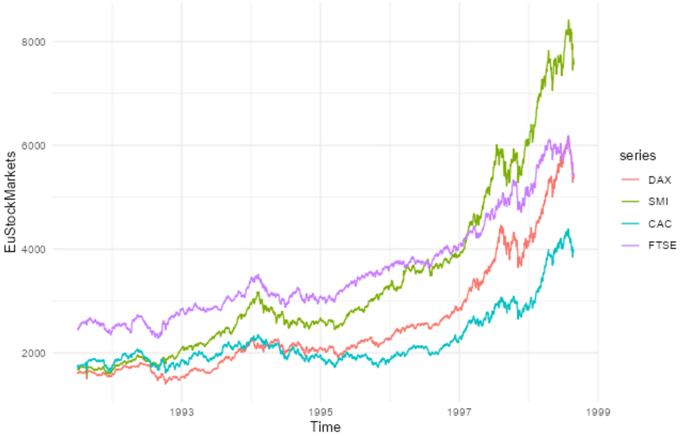
Time Series

Sara Geremia

May 9th, 2024

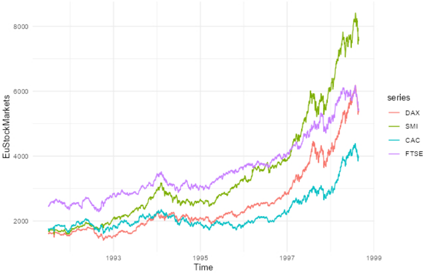
Time Series

- A set of observations indexed by time t: $Y_1, Y_2, ..., Y_T$
- 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.
- Δ 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 jth lagged value (or, jth 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.

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 (or autocorrelation coefficient) 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:

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

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:

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

Stationary Time Series

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

Mathematical Definition

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

$$E(Y_t) = \mu$$
$$Var(Y_t) = \sigma^2$$
$$Cov(Y_t, Y_{t+k}) = \gamma_k$$

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.