

Time Series Analysis

Nik Bear Brown

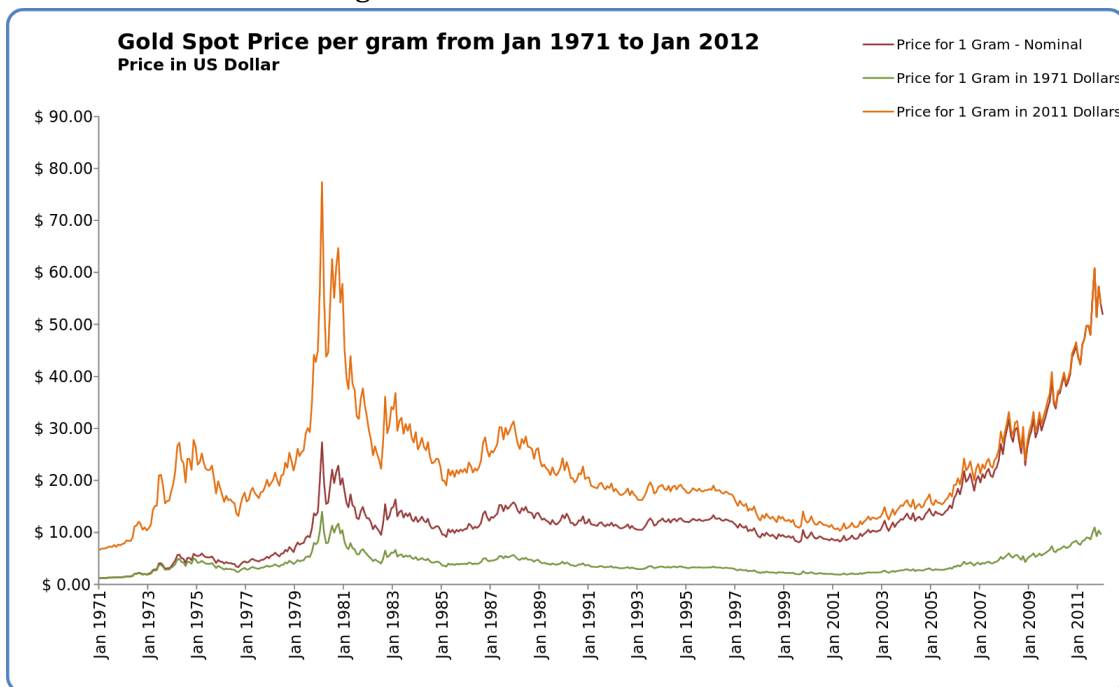
In this lesson we will discuss time series analysis theory. There are no data sets, or libraires to be installed.

Time Series

A **time series** is a sequence of data points, typically consisting of successive measurements made over a time interval. This is a very common type of data as we frequently measure how something varies over time. usually a time series is an ordered sequence of values of a variable at equally spaced time intervals; but methods exist to deal with irregular sampling.

Examples of time series include:

- Stock Market
- The change of weather
- An electrocardiogram (EKG or ECG), that is, the electrical activity of your heart.
- The popularity of a celebrity or politician
- Economic Forecasting

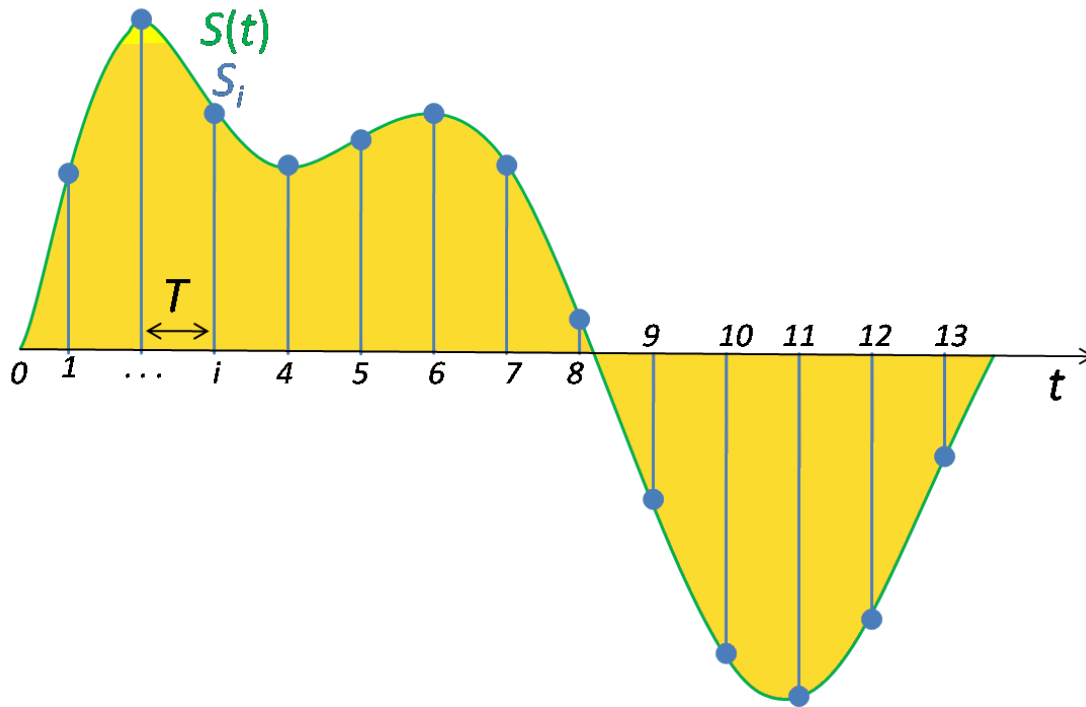


Gold Spot Price

- from https://commons.wikimedia.org/wiki/File:Gold_Spot_Price_per_Gram_from_Jan_1971_to_Jan_2012.svg

Univariate (bivariate, multivariate) time series: collection of observations of one (two, several) state variables, that are made in sequential moments in time.

The **sampling** frequency (or sample rate) is the number of samples per unit of time. (e.g once per minute).



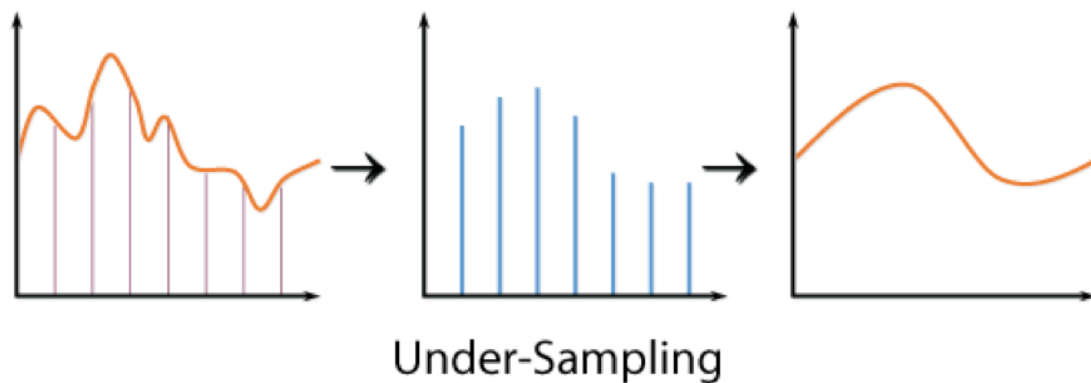
- from https://commons.wikimedia.org/wiki/File:Signal_Sampling.png

Nyquist criterion

Nyquist criterion, or **Nyquist frequency**, named after electronic engineer **Harry Nyquist**, is half of the sampling rate of a discrete signal processing system. In order to correctly determine the frequency spectrum of a signal, the signal must be measured at least twice per period.

Reconstruction

Reconstruction is the opposite of sampling, the process by which a sampled signal is converted into a continuous signal.



Under-Sampling

Components of Time Series

The pattern or behavior of the data in a time series can be made up of several components. Theoretically, any time series can be decomposed into:

Trend

The trend component accounts for the gradual shifting of the time series to relatively higher or lower values over a long period of time.

Cyclical

A regular pattern of sequences of values that go above and below the trend line is a cyclical component.

Seasonal

The seasonal component accounts for regular patterns of variability within certain time periods, such as a year. While seasons could be modeled using cyclical components. Often within-year, within-month, within-week or within-day cycles are treated as “seasonal” behavior.

Irregular

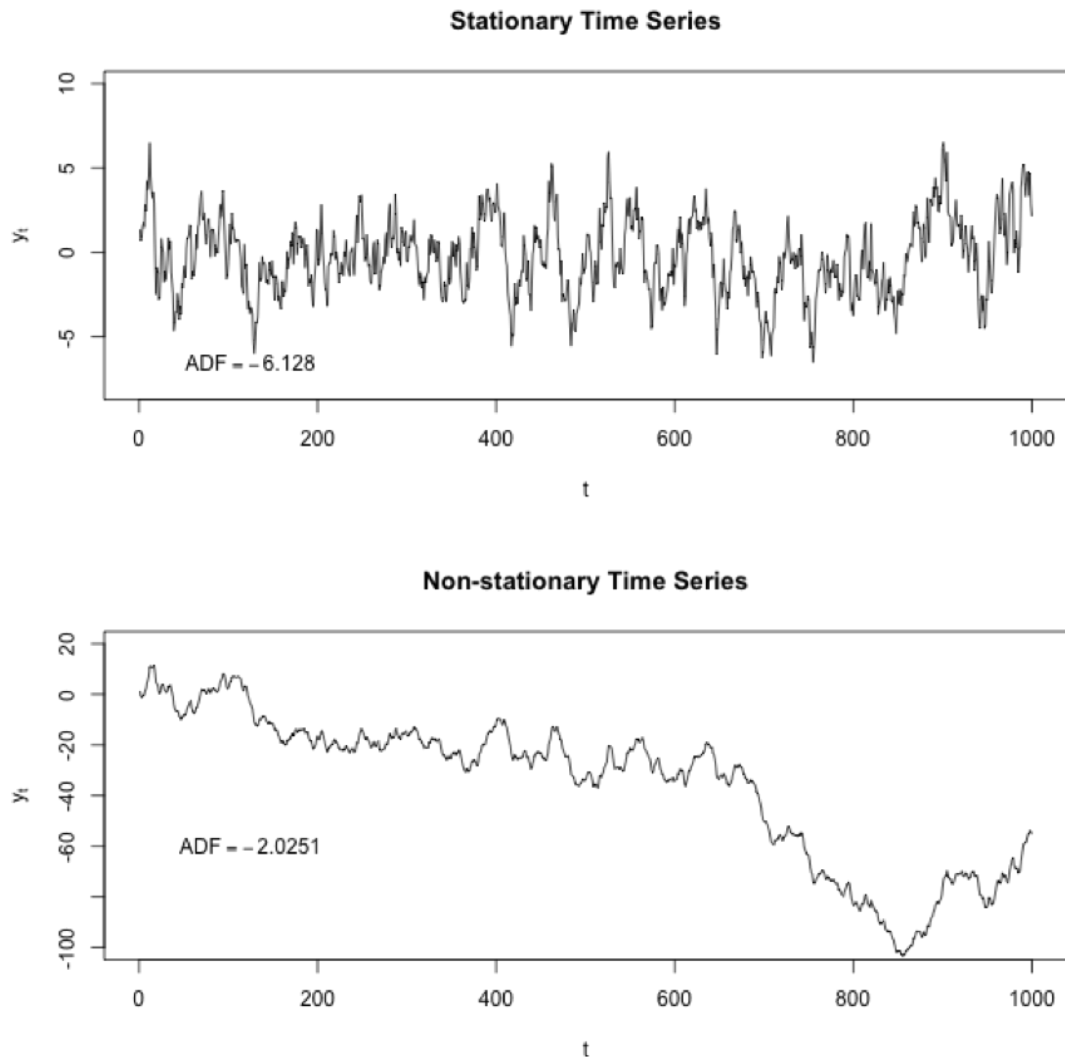
The irregular components represent erratic, unsystematic, ‘residual’ fluctuations. That is, noise.

Stationary versus Non-stationary time series

Non-stationary time series or have means, variances and covariances that change over time. Whereas while stationary time series have up and down fluctuations they have means, variances and covariances. That is, a [stationary process](#) is a stochastic process whose joint probability distribution does not change when shifted in time.

For example, we expect financial time series data, to be non-stationary since we would expect a deterministic upward trend in price.

A grown adults weight would more likely be stationary time series unless he or she fundenemtally changes their "state".



Two simulated time series processes, one stationary the other non-stationary. - from <https://commons.wikimedia.org/wiki/File:Stationarycomparison.png>

Smoothing Methods of Time Series

Smoothing a time series: to eliminate some of short-term fluctuations. Sometimes smoothing is done to remove cyclical or seasonal fluctuations. (i.e. "deseasonalize" a time series). Usually smoothing refers to methods for reducing or canceling the effect due to random variation. The simplest smoothing method is just the "simple average of all past data. Of course, the "simple" average or mean of all past observations is only a useful estimate for forecasting when there are no trends.

Time Series Decomposition

Time Series Decomposition is a procedure to identify the component factors of a time series. That is, to create a model that expresses the time series variable Y in terms of the components T (trend), C (cycle), S (seasonal) and I (irregular).

Time Series Analysis

Time series analysis generates a model that accounts an internal structure (such as autocorrelation, trend or seasonal variation) of a set of data points taken over time. This is very frequently used for forecasting. That is, predicting and future event in time based on a sequential historical sample.

Autocorrelation

Autocorrelation, also known as serial correlation or cross-autocorrelation, is the cross-correlation of a signal with itself at different points in time. It is the correlation between values of the process at different times. Informally, it is the similarity between observations as a function of the time lag between them. It is a mathematical tool for finding repeating patterns, such as the presence of a periodic signal obscured by noise.

Let X be some repeatable process, and i be some point in time after the start of that process. (i may be an integer for a discrete-time process or a real number for a continuous-time process.) Then X_i is the value (or realization) produced by a given run of the process at time i . Suppose that the process is further known to have defined values for mean μ_i and variance σ_i^2 for all times i . Then the definition of the autocorrelation between times s and t is

$$R(s, t) = \frac{E[(X_t - \mu_t)(X_s - \mu_s)]}{\sigma_t \sigma_s},$$

where "E" is the expected value operator. Note that this expression is not well-defined for all time series or processes, because the variance may be zero (for a constant process) or infinite (for processes with distribution lacking well-behaved moments, such as certain types of power law).

Naive & Simple Averaging

Naive or "simple" averaging is just the mean (or median) of all past data. Of course, the "simple" average or mean of all past observations is only a useful estimate for forecasting when there are no trends.

For example, if we have prices are $p_M, p_{M-1}, \dots, p_{M-(n-1)}$ then the formula is

$$\text{SMA} = \frac{p_M + p_{M-1} + \dots + p_{M-(n-1)}}{n}$$

Moving Averages

A moving average (rolling average or running average) is an average that is updated for a window or history of n events. This is sometimes called the arithmetic moving average the most recent n data values. For an an equally weighted average of the sequence of n values x_1, \dots, x_n up to the current time:

$$\text{MA}_n = \frac{x_1 + \dots + x_n}{n}$$

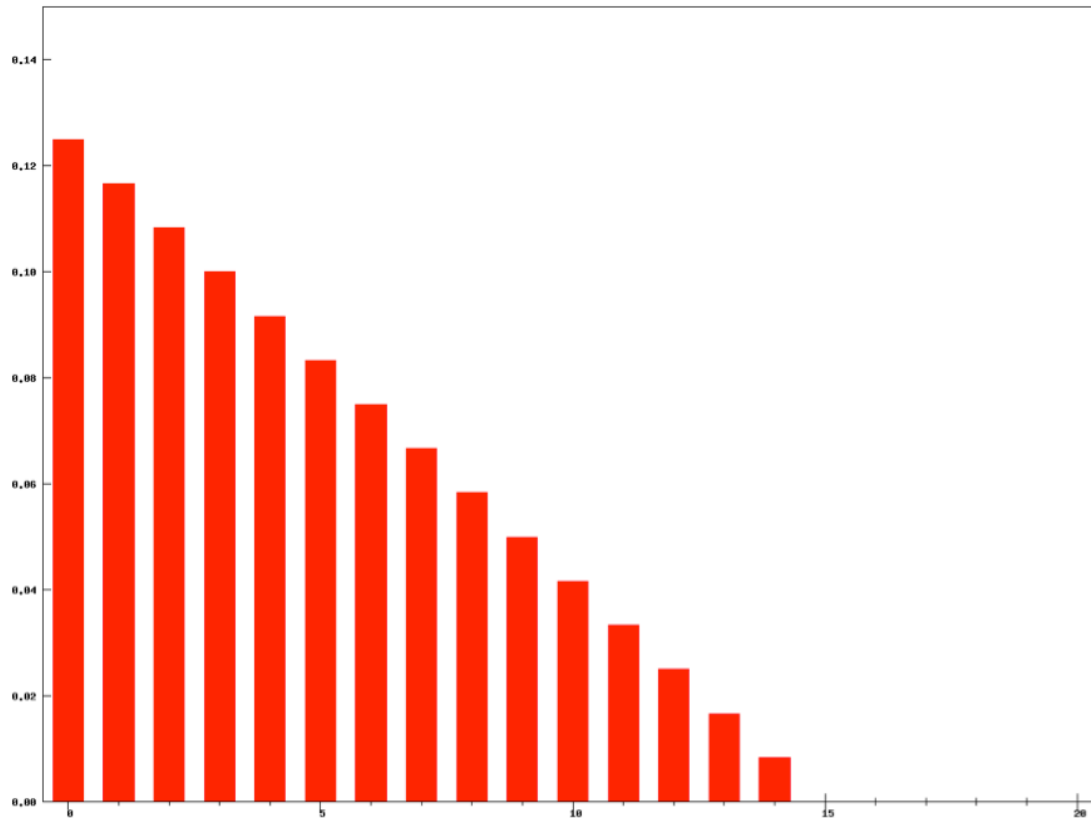
Weighted Moving Averages

A weighted moving average places more weight on recent observations. Sum of the weights needs to equal 1. For example, in technical analysis of financial data, a weighted moving average (WMA) often uses weights that decrease in arithmetical progression. In an n -day WMA the latest day has weight n , the second latest $n - 1$, etc., down to one.

$$\text{WMA}_M = \frac{n p_M + (n-1) p_{M-1} + \dots + 2 p_{(M-n+2)} + p_{(M-n+1)}}{n + (n-1) + \dots + 2 + 1}$$

That is, the denominator is a triangle number equal to $\frac{n(n+1)}{2}$.

There are many other methods for weighting the importance of more recent observations.



"Weighted moving average weights N=15". Licensed under CC BY-SA 3.0 via Commons

Exponential moving average

Exponential moving average (also called Single Exponential Smoothing) continually revises a forecast in light of more recent experiences. Averaging (smoothing) past values of a series in a decreasing (exponential) manner.

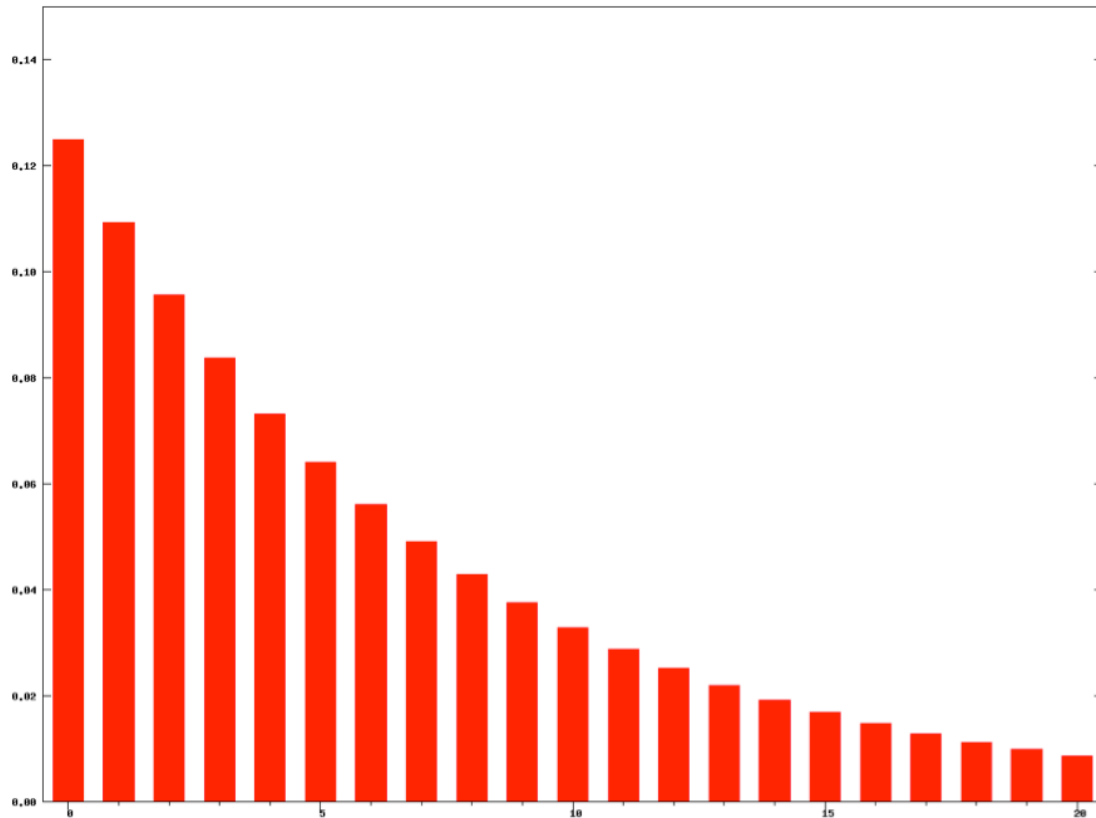
The EMA for a series Y may be calculated recursively:

$$S_1 Y_1 \quad \text{for } t > 1, S_t = \alpha \cdot Y_t + (1 - \alpha) \cdot S_{t-1}$$

Where:

- The coefficient α represents the degree of weighting decrease, a constant smoothing factor between 0 and 1. A higher α discounts older observations faster.
- Y_t is the value at a time period t .
- S_t is the value of the EMA at any time period t .

Since the base values of the recursive initial call is undefined. The priming of values such as S_1 may be initialized in a number of different ways, most commonly by setting S_1 to Y_1 , or by setting S_1 to an average of the first few observations.



"Exponential moving average weights N=15". Licensed under CC BY-SA 3.0 via Commons

Double exponential smoothing

Holt-Winters method for exponential smoothing extended Holt's method to capture seasonality. This is sometimes called "Double exponential smoothing." The Holt-Winters seasonal method comprises the forecast equation and two smoothing equations. We use $\{s_t\}$ to represent the smoothed value for time t , and $\{b_t\}$ is our best estimate of the trend at time t . The output of the algorithm is now written as F_{t+m} , an estimate of the value of x at time $t+m$, $m>0$ based on the raw data up to time t . Double exponential smoothing is given by the formulas

$$\begin{aligned} s_1 &= x_1 \\ b_1 &= x_1 - x_0 \end{aligned}$$

And for $t > 1$ by

$$\begin{aligned} s_t &= \alpha x_t + (1 - \alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1} \end{aligned}$$

where α is the data smoothing factor, $0 < \alpha < 1$, and β is the trend smoothing factor, $0 < \beta < 1$.

Triple exponential smoothing

Triple exponential smoothing account seasonal changes as well as trends using three smoothing equations — one for the level, one for trend, and one for the seasonal component denoted, with smoothing parameters α , β , and γ . We use a variable m to denote the period of the seasonality, i.e., the number of seasons in a year. For example, for quarterly data $m=4$, and for monthly data $m=12$.

There are two variations to this method that differ in the nature of the seasonal component. The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.

Autoregressive integrated moving average (ARIMA)

An autoregressive integrated moving average (ARIMA or ARMA) model combines an autoregressive component with a moving average component in to a single model.

An [autoregressive integrated moving average \(ARIMA or ARMA\)](#) model is a generalization of an autoregressive moving average (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to reduce the non-stationarity.

Non-seasonal ARIMA models are generally denoted $ARIMA(p, d, q)$ where parameters p , d , and q are non-negative integers, p is the order of the Autoregressive model, d is the degree of differencing, and q is the order of the Moving-average model. The he number of differences d is determined using repeated statistical tests. The values of p and q are then chosen by minimizing the AICc after differencing the data d times.

The ARIMA model uses an iterative three-stage modeling approach:

Model identification and model selection: making sure that the variables are stationary, identifying seasonality in the dependent series (seasonally differencing it if necessary), and using plots of the autocorrelation and partial autocorrelation functions of the dependent time series to decide which (if any) autoregressive or moving average component should be used in the model.

Parameter estimation using computation algorithms to arrive at coefficients that best fit the selected ARIMA model. The most common methods use maximum likelihood estimation or non-linear least-squares estimation.

Model checking by testing whether the estimated model conforms to the specifications of a stationary univariate process. In particular, the residuals should be independent of each other and constant in mean and variance over time.

(Plotting the mean and variance of residuals over time and performing a Ljung-Box test or plotting autocorrelation and partial autocorrelation of the residuals are helpful to identify misspecification.) If the estimation is inadequate, we have to return to step one and attempt to build a better model.

Linear Time-Series Model

Using regression we can model and forecast the trend in time series data by including the time as a predictor variable. Time series processes are often described by multiple linear regression (MLR) models of the form:

$$y_t = X_t\beta + e_t,$$

Note that the above equation differs from our usual regression in that the predictor variables are a function of time, t . As opposed to the more familiar form below:

$$Y = \beta X + e$$

where y_t is an observed response and X_t includes columns for contemporaneous values of observable predictors. The partial regression coefficients in β represent the marginal contributions of individual predictors to the variation in y_t when all of the other predictors are held fixed. A common feature of time series data is a trend. Regressing non-stationary time series can lead to spurious regressions.

The error term ε_i still has the following assumptions:

- have mean zero; otherwise the forecasts will be systematically biased.
- statistical independence of the errors (in particular, no correlation between consecutive errors in the case of time series data).
- homoscedasticity (constant variance) of the errors.
- normality of the error distribution.

Resources

- [Using R for Time Series Analysis](#)
- [Quick-R: Time Series](#)
- [R Time Series Tutorial](#)
- [Time Series Analysis - RDataMining.com: R and Data Mining](#)
- [Learning Time Series with R - Revolutions](#)
- [Using R \(with applications in Time Series Analysis\)](#)

