Review

Modelling objectives:

Consider a regression model:

Inference

Goal: Discovery of possible causes, i.e. find the confidence interval of

Prediction

Goal: Estimate , i.e. find the confidence interval of

For all regression model, the goal is the minimize the residuals, such that their mean is zero.

Ordinary least squares(OLS):

,

where

First of all, we need to check if the given time series is a random walk.

Random walk type

Pure Random walk:

where

Random Walk with Drift:

where is a constant called drift

Random Walk with Drift and Deterministic Trend

where is a constant called drift and is a deterministic trend.

We cannot forecast a random walk.

Define AR(1):

If and , then the price is called mean reversion.

If the price is not random walk, the price will revert to a long term mean .

Stock price model:

Geometric random walk with trend

where is the continuously compounded rate of return of the stock, is the implied volatility of the stock and is discretized to

Stock prices are not always random walks, so it makes sense to try to predict selected price data with machine learning or deep learning.

Assumptions:

1. (Gaussian White Noise) , and independent from the input variables
2. Stationarity: mean, variance and covariance do not vary with time

All models below assume that the series be modeled is a stationary series.

Statistical models

1. AR(p):
   * p is the number of target lag inputs in the regression
   * p depends on the partial auto correlation graph of the target
2. ARMA(p, q):
   * adds q, the number of residual inputs in the regression
   * the residuals are generated by the previous application of ARMA(p,q) regression
   * q depends on the auto correlation graph of the target
3. ARIMA (p,d,q):
   * adds d, the number of times the target is differenced
4. SARIMA (p,d,q)x(P,D,Q,m):
   * adds m, the seasonal period length, D, the seasonal differencing, P, the seasonal lag input no. & Q, the seasonal regression residual input no.
5. SARIMAX: adds exogenous inputs

Graphical user interface, text, application

Description automatically generated

What Time series decomposition do

1. Remove Trend
2. Remove Seasonality
3. Remove Correlation

Workflow:

1. Test for Mean Reversion
   1. The Dickey Fuller Test for Mean Reversion

Test:

* 1. The Augmented Dickey Fuller Test for Mean Reversion
  2. Test for stationary: <https://towardsdatascience.com/detecting-stationarity-in-time-series-data-d29e0a21e638>

ACF: Autocorrelation gives you an idea of how data points at different points in time are linearly related to one another as a function of their time difference

Property: the ACF of the sum of two periodic series is the sum of the individual ACFs

PACF: The partial autocorrelation of a time series for a given lag is the partial correlation of the time series with itself at that lag given all the information between the two points in time. It means you need to compute a number of conditional correlations and subtract these out of the total correlation. The PACF, on the other hand, reveals which correlations are “true” informative correlations for specific lags rather than redundancies.

Property: it can determine the seasonal lag even though the data has trend.