TIME SERIES

**The QQ plot**

The quantile-quantile plot is used to determine if a data is distributed a certain way.

It showcases how the data set fits a Normal Distribution

This takes all the values a variable can take, and arrange them in ascending order.

The Y axis takes the values, from low to high, and the X axis, the theoretical quantiles,

The theoretical quantiles measure how many standard deviations away from the mean,

these values are.

The Red line shows what the data point should follow, if they are Normally Distributed.

**White Noise**

A special type of time-series where the data doesn’t follow a pattern. Since no pattern c

could be found, we can’t predict white noise. For white noise, there should be a

a constant Mean =0

a constant Variance  is constant

no autocorrelation , no clear relationship of the past and

present values of the series

white noise is a sequence of random data, where every value has a time-period associated

with it.

**Random Walk**

A special type of time-series, where values tend to persist over time and the difference

between the periods are simply white noise.

, where , arbitrary and unpredictable.

* **Seasonality**
  + Appearance of trends on a cyclical basis
  + Several ways of handling seasonality, eg. Decomposition
    - Split into 3 effects

Trend…….Pattern

Seasonal …..Cyclical effect

Residual……Error of prediction

Naïve Decomposition: we expect a linear relationship between the three parts and the

observed time series.

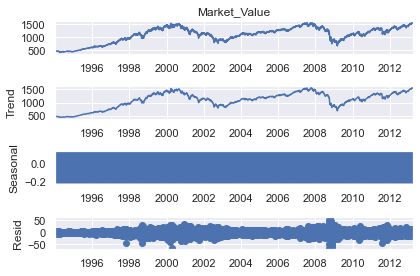
Naïve Decomposition has two main approaches:

Additive: at any time,

Observed values = trend + seasonal + residual

Multiplicative: at any time,

Observed values = trend x seasonal x residual



The trend closely mimics the Market\_Values or the Observed values or the series, because the decomposition function uses the previous period values as a trend-setter.

The trend explains most of the variabilities in the data.

The Seasonal part looks looks like a rectangle, because the values oscilating between

-2 and 1, indicating no concrete cyclical pattern, when using naïve decomposition.

The residuals are the errors for the model…representing the true values and the predictors. Clearly, you could observe the instability around 2000 and 2008…reasons could be assigned, empirically…housing prices etc.

Result of additive decomposition suggests no seasonality in the data.( Trend is almost the same as Observed, and the seasonal is rectangular)

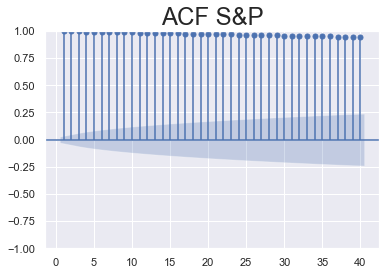
**Autocorrelation**

This measures the correlation between the sequence and itself. Measures similarities

Between previous data and current, in a series…maybe similarities in sales of yesterday

And today.

**Autocorrelation Function ACF**

****

Values on the X-axis represents the lags, and Y-axis, possible values of autocorrelation

Coefficients or the correlation coefficients of the present value and its respective lagged

Values…the correlation coefficients or the autocorrelation coefficient ranges from -1 to 1…since all the lagged values are closer to +1, it indicates that, there exist a very strong

positive correlation between the present value and all the 39 previous values, or a positively strong autocorrelation between the present value and the lagged values.

Clearly, there exist a stronger autocorrelation, with immediate lags, whiles autocorrelation reduces, with increasing lags or farther from the present value.

The greater the distance in time, the more unlikely it is that autocorrelation persists. Eg.

Today’s prices is closer to yesterdays, than a month ago; or recent prices best explain

current price than old prices.

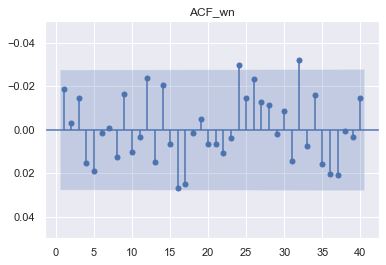
The blue point at the top gradually diminishes with lags…indicating that, autocorrelation

diminish with time or lags.

Shaded blue area is the level of significance…clearly, as the lags increases, the level of

Significance decreases….clearly, the immediate lag values have higher autocorrelation coefficient and are highly significant. As the lags increases, the autocorrelation decreases and the level of significance also decreases.

The ACF of a White Noise



Unlike that of the S&P, the spikes go either way…positive and negative, whiles the S&P has all positive values.

There isn’t any strong autocorrelation coefficient amongst the lags…none of the values or spikes has a strong autocorrelation coefficient, 0.8…not even 0.1…highest is 0.03…indicating no autocorrelation between the current value and any of the lagged or previous values. This indicates that, the present value or future values cannot be predicted by previous or lagged values…no predictive power whatsoever.

Also, all the lines in the S&P falls within the blue area.

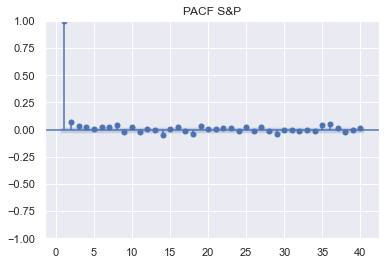
**Partial Autocorrelation Function PACF**

In autocorrelation, the present value is affected, directly and indirectly, by the previous

values. We us the PACF to measure the direct relationship. We use PACF to estimate,

for instance, lag (3), directly on the current value…unlike in the ACF where lags (1,2 and

3) all influence the current value. Thus, the PACF cancels all the effects of the other lags, but the reference lag, whiles the ACF estimates the combined effects of all the previous lags.



Clearly, only the present value is significant. Worth noting that, the first lag of both ACF and PACF are identical.

**Picking the Correct Model**

The Log-Likelihood Ratio Test (LLR test), to test the statistically significance of

predictions …applies to models with different degrees of freedoms

Information Criteria

The lowest AIC or BIC