Contents

[What is a Time Series? 1](#_Toc521544705)

[Stationarity 2](#_Toc521544706)

[Trend 3](#_Toc521544707)

[Seasonality 3](#_Toc521544708)

[Seasonal Dummy Variables 3](#_Toc521544709)

[Serial Correlation (Autocorrelation) 4](#_Toc521544710)

[The Autocorrelation Function 4](#_Toc521544711)

[The Partial Autocorrelation Function 4](#_Toc521544712)

[White Noise and Random Walks 5](#_Toc521544713)

[The Dickey-Fuller Unit Root Test 5](#_Toc521544714)

[Autoregressive Models - AR(p) 6](#_Toc521544715)

[Moving Average Models - MA(q) 6](#_Toc521544716)

[Autoregressive Moving Average Models - ARMA(p, q) 7](#_Toc521544717)

[Autoregressive Integrated Moving Average Models - ARIMA(p, d, q) 7](#_Toc521544718)

[Steps to build a Time Series 8](#_Toc521544719)

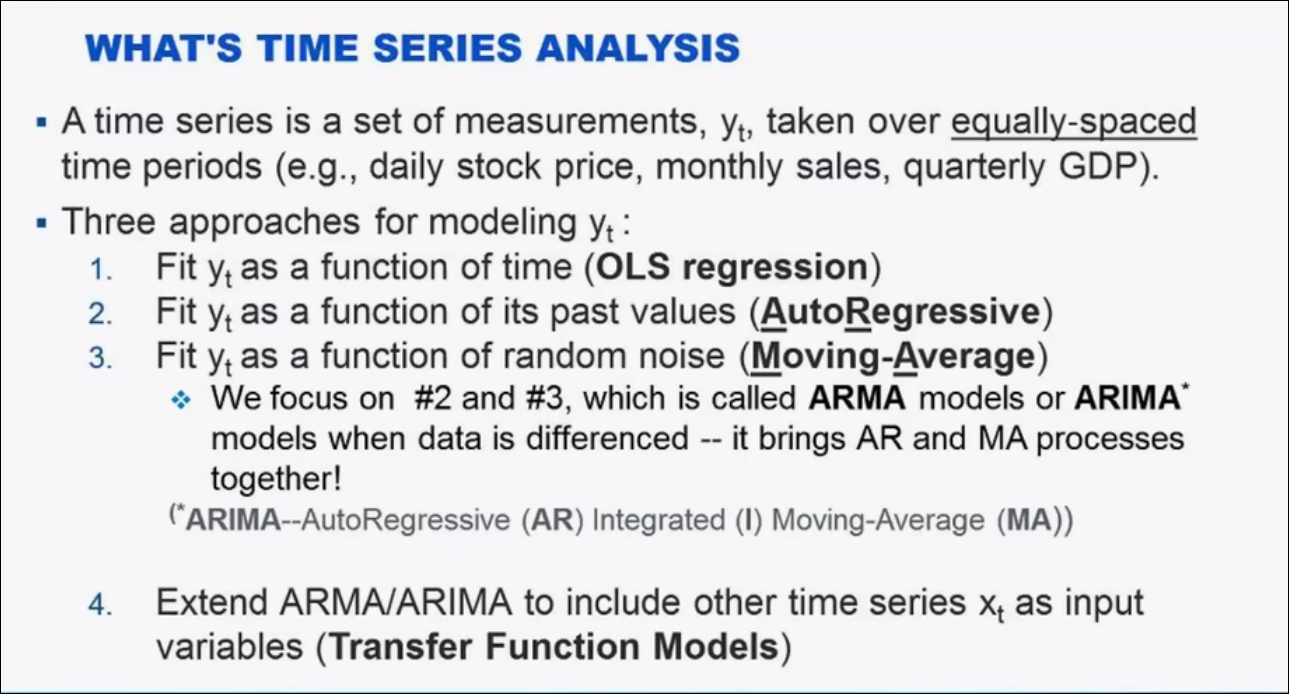
[Link 8](#_Toc521544720)

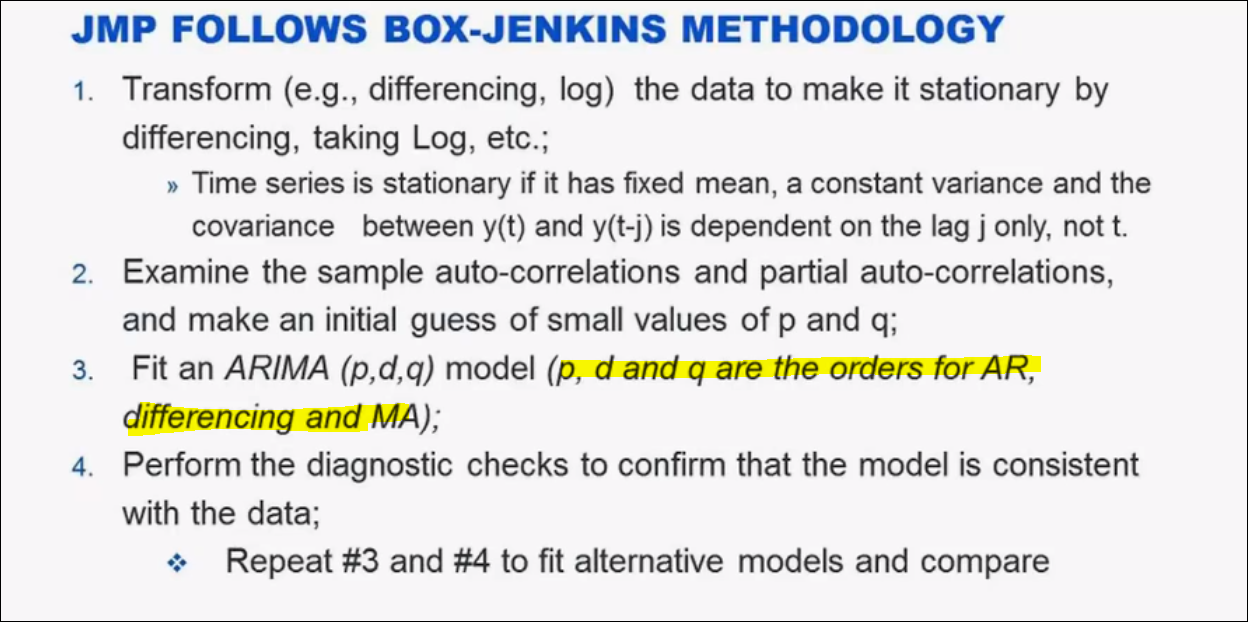
[Interview questions 9](#_Toc521544721)

[Have you used a time series model? Do you understand cross-correlations with time lags? 9](#_Toc521544722)

# What is a Time Series?

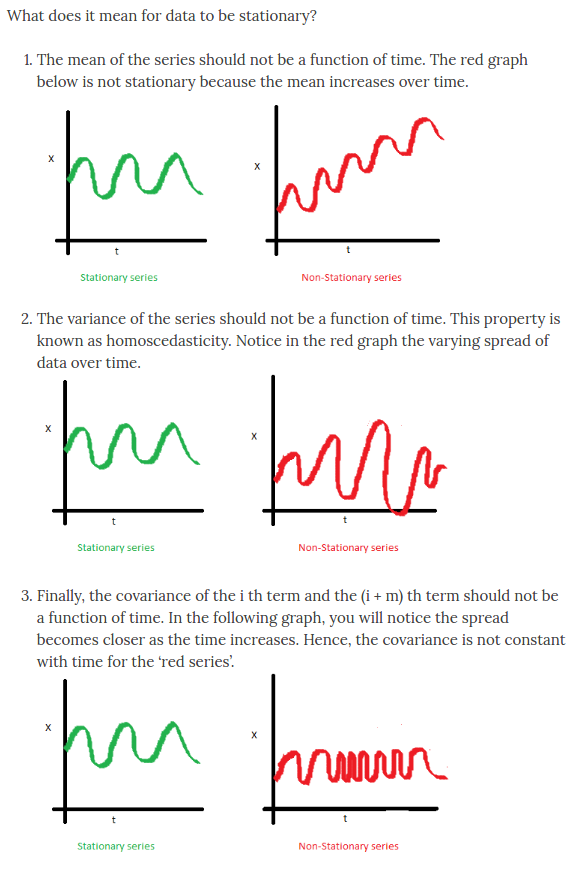
*A time series is a series of data points indexed (or listed or graphed) in time order.*





Box Jenkins methodology

# Stationarity



***So what?  Why do we care about stationarity?***

* A stationary time series (TS) is simple to predict as we can assume that future statistical properties are the same or proportional to current statistical properties.
* Most of the models we use in TSA assume **covariance-stationarity (#3 above)**. This means **the descriptive statistics these models predict e.g. means, variances, and correlations, are only reliable if the TS is stationary and invalid otherwise.**

*"For example, if the series is consistently increasing over time, the sample mean and variance will grow with the size of the sample, and they will always underestimate the mean and variance in future periods. And if the mean and variance of a series are not well-defined, then neither are its correlations with other variables."*

With that said, **most TS we encounter in finance is NOT stationary.** Therefore a large part of TSA involves identifying if the series we want to predict is stationary, and if it is not we must find ways to transform it such that it is stationary. (More on that later)

# Trend

*Trend* usually refers to a **deterministic** function of time.

􀂄Time series can be made of deterministic and **stochastic** components.

􀂄A stochastic component is subject to random variation and can never be predicted perfectly except for chance occurrences.

􀂄A deterministic component exhibits no random variation and can be predicted perfectly. Common deterministic trend functions include linear trend, curvilinear trend, logarithmic trend, and exponential trend.

# Seasonality

The *seasonal* component of a time series represents the effects of seasonal variation.

􀂄The most general meaning of *seasonality* is a component that describes repetitive behavior at known seasonal periods. If the seasonal period is integer S, then seasonal factors are factors that repeat every S units of time.

# Seasonal Dummy Variables

􀂄For a time series with S seasons, there will be S dummy variables, one for each season.

􀂄If the model has a constant term, only S-1 dummy variables are required.

# Serial Correlation (Autocorrelation)

Essentially when we model a time series we decompose the series into three components: trend, seasonal/cyclical, and random. The random component is called the residual or error. It is simply the difference between our predicted value(s) and the observed value(s).

We can calculate the correlation for time series observations with observations with previous time steps, called **lags**. Because the correlation of the time series observations is calculated with values of the same series at previous times, this is called **a serial correlation, or an autocorrelation.**

 Serial correlation is when the residuals (errors) of our TS models are correlated with each other.

**Why Do We Care about Serial Correlation?**

We care about serial correlation because it is critical for the validity of our model predictions, and is intrinsically related to stationarity. Recall that the residuals (errors) of a stationary TS are serially uncorrelated by definition! If we fail to account for this in our models the standard errors of our coefficients are underestimated, inflating the size of our T-statistics.

The result is too many Type-1 errors, where we reject our null hypothesis even when it is True! **In layman's terms, ignoring autocorrelation means our model predictions will be bunk, and we're likely to draw incorrect conclusions about the impact of the independent variables in our model.**

# The Autocorrelation Function

􀂄The autocorrelation function at lag *k*, ACF(*k*), represents the correlation of a time series with itself lagged by *k* time units.

􀂄The autocorrelation function is one of the primary tools for diagnosing trend, seasonality, and candidate forecast models.

􀂄A portfolio of autocorrelation function shapes may be used to determine an approximating time series model.

**Auto correlation coefficient – Moving average part**

Consider a time series that was generated by an autoregression (AR) process with a lag of k.

We know that the ACF describes the autocorrelation between an observation and another observation at a prior time step that includes direct and indirect dependence information.

This means we would expect the ACF for the AR(k) time series to be strong to a lag of k and the inertia of that relationship would carry on to subsequent lag values, trailing off at some point as the effect was weakened.

# The Partial Autocorrelation Function

A partial autocorrelation is a summary of the relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed.( *The partial autocorrelation at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags.)*

The autocorrelation for an observation and an observation at a prior time step is comprised of both the direct correlation and indirect correlations. These indirect correlations are a linear function of the correlation of the observation, with observations at intervening time steps.

It is these indirect correlations that the partial autocorrelation function seeks to remove. Without going into the math, this is the intuition for the partial autocorrelation.

We know that the PACF only describes the direct relationship between an observation and its lag. This would suggest that there would be no correlation for lag values beyond k.

􀂄The partial autocorrelation function at lag *k*, PACF(*k*), represents the correlation of a time series with itself at lag *k* adjusted for lags 1 through *k*-1.

􀂄PACF(*k*) is the coefficient of the *k*-th order autoregressive term in an autoregressive order-*k* model.

􀂄PACF(*k*) is calculated using a fast recursion equation.

􀂄A portfolio of partial autocorrelation function shapes may be used to determine an approximating time series model.

It will help in deciding auto correlation part

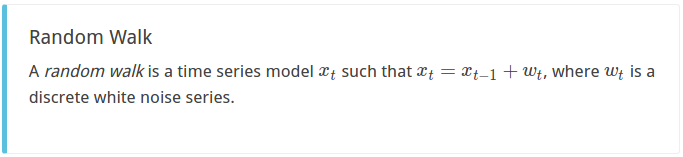
# White Noise and Random Walks

White noise is the first Time Series Model (TSM) we need to understand. By definition a time series that is a white noise process has serially UNcorrelated errors and the expected mean of those errors is equal to zero.

Another description for serially uncorrelated errors is, [independent and identically distributed (i.i.d.)](https://en.wikipedia.org/wiki/Independent_and_identically_distributed_random_variables). This is important because, if our TSM is appropriate and successful at capturing the underlying process, the residuals of our model will be i.i.d. and resemble a white noise process. Therefore part of TSA is literally trying to fit a model to the time series such that the residual series is indistinguishable from white noise.

Let's simulate a white noise process and view it. Below I introduce a convenience function for plotting the time series and analyzing the serial correlation visually. This code was adapted from the blog [Seanabu.com](http://www.seanabu.com/2016/03/22/time-series-seasonal-ARIMA-model-in-python/)

Random Walk



# The Dickey-Fuller Unit Root Test

􀂄The null hypothesis is that the time series is not stationary.

􀂄The alternative hypothesis is that the time series is stationary.

􀂄A significant value rejects the null hypothesis and provides evidence that the time series is stationary.

􀂄The name *unit root* reflects the fact that a time series with a certain characteristic polynomial having a unit root is not stationary. The mathematical details are

beyond the scope of this course.

􀂄Test statistics are available for lags 0 through at most

5. These reflect six different *approximating models*.

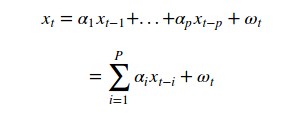
The six tests are not independent, but they may yield

conflicting results.

# Autoregressive Models - AR(p)

For a time-series auto correlation at lag L is defined as Pearson coefficient or correlation between **yt** and **yt+L**

When the dependent variable is regressed against one or more lagged values of itself the model is called autoregressive. The formula looks like this:



AR(P) MODEL FORMULA

When you describe the **"order"** of the model, as in, an AR model of order **"p",**the p represents the number of lagged variables used within the model. For example an AR(2) model or *second-order*autoregressive model looks like this:

 AR(2) model formula 

AR(2) MODEL FORMULA

# Moving Average Models - MA(q)

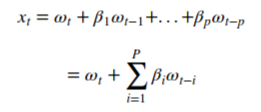
While Autocorelation is a useful property of time series , it does not always explain all the variation.

The unexpected component of variation (shocks) does impact future value of time series.This component can be captured with MA.

MA predicts subsequent series values based on the past history of deviation from predicted values.

A deviation from the predicted value can be viewed as white noise or shock.

MA(q) models are very similar to AR(p) models. The difference is that the MA(q) model is a linear combination of past white noise error(residual error) terms as opposed to a linear combo of past observations like the AR(p) model. The motivation for the MA model is that we can observe "shocks" in the error process directly by fitting a model to the error terms. In an AR(p) model these shocks are observed indirectly by using the ACF on the series of past observations. The formula for an MA(q) model is:



Omega (w) is white noise with E(wt) = 0 and variance of sigma squared. Let's simulate this process using beta=0.6 and specifying the AR(p) alpha equal to 0.

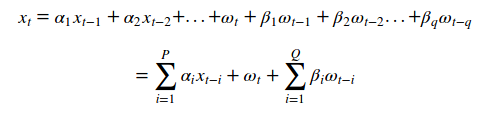
We would expect the ACF for the MA(k) process to show a strong correlation with recent values up to the lag of k, then a sharp decline to low or no correlation. By definition, this is how the process was generated.

For the PACF, we would expect the plot to show a strong relationship to the lag and a trailing off of correlation from the lag onwards.

# Autoregressive Moving Average Models - ARMA(p, q)

As you may have guessed, the ARMA model is simply the merger between AR(p) and MA(q) models. Let's recap what these models represent to us from a quant finance perspective:

* AR(p) models try to capture *(explain)* the momentum and mean reversion effects often observed in trading markets.
* MA(q) models try to capture *(explain)* the shock effects observed in the white noise terms. These shock effects could be thought of as unexpected events affecting the observation process e.g. Surprise earnings, A terrorist attack, etc.



# Autoregressive Integrated Moving Average Models - ARIMA(p, d, q)

# Steps to build a Time Series

Some important questions to first consider when first looking at a time series are:

* Is there a **trend**, meaning that, on average, the measurements tend to increase (or decrease) over time?
* Is there **seasonality**, meaning that there is a regularly repeating pattern of highs and lows related to calendar time such as seasons, quarters, months, days of the week, and so on?
* Are their **outliers**? In regression, outliers are far away from your line. With time series data, your outliers are far away from your other data.
* Is there a **long-run cycle** or period unrelated to seasonality factors?
* Is there **constant variance**over time, or is the variance non-constant?
* Are there any **abrupt changes** to either the level of the series or the variance?

# Link

<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/>

https://drive.google.com/file/d/0BwogTI8d6EEiaDJCRXd0dmU1ZDA/edit

# Interview questions

## Have you used a time series model? Do you understand cross-correlations with time lags?