**Time Series**

**white noise –**

A special type of time-series, where the data doesn’t follow a pattern.

Conditions for white noise – constant mean, constant standard deviation, no autocorrelation in any period.

White noise is an example of weak form stationary process.

If data is non-stationary, it can’t be considered white noise.

A white noise process produces completely random data and all the ACF coefficients should not be significantly different from 0.

**Random walk –**

All random walk processes are non-stationary.

**Picking the correct model**

We usually start with a simple model and then expand it as long as it follows several conditions –

1. Significant coefficients – The coefficients of the complex model need to be significantly different from zero. We must omit them as they hold no predictive powers and revert back to the simpler model.
2. Parsimonious (as simple as possible) –
   1. Generally – simpler model > complex model
   2. Significantly better predictions – simpler model < complex model

To determine whether the predictions are significantly better, LLR (log likelihood ratio) test can be done. It can only be applied to models with different degrees of freedom.

When comparing several models with the same total number of lags -> No LLR.

We need to compare the information criteria(AIC, BIC) of each one.

The lower the coefficient, the less data the model require to make accurate predictions.

We opt for model with **lower AIC and BIC coefficients** and **higher log-likelihood.**

LLR test – tests if the difference between log likelihoods of two models is significant or not. If the LLR is within 5% significance, then we should opt for the more complicated model (which is the second parameter in the LLR function).

We can repeat the same process of fitting a more complicated model and checking if it gives us distinguishably greater log-likelihoods until we reach a model where it doesn’t.

Moreover, the said model should also have its last lag not be significantly different from 0.

1. If our model fits well there should be no trend we have failed to account for. Therefore, Residuals of the model should resemble **white noise.**

**AR model**

AR models processes the stationary data the best. One way to deal with non-stationary time series in finance is to use **returns time series** instead of prices as it should fit the stationary assumptions.

Fitting higher Lag AR models – Usually, the more lags we include, the better our predictions become. However, we can run the risk of overfitting.

As a general rule, we prefer higher log-likelihoods but lower information criterion(AIC, BIC and HQIC).

2 conditions need to be satisfied for knowing how many lags to consider –

1. Non-significant p-value for the LLR test
2. Non-significant p-value for the highest lag coefficients.

Max no. of lags we are willing to include in AR model can be determined using PACF.

The value of each coefficient is between -1 and 1. (to prevent compounded effects exploding in magnitude)

**Residuals are usually unpredictable differences because if there’s a pattern it will be captured in the other incumbents of the model.**

**Residuals can be referred as unpredictable shocks.**

To determine the correct no of lags we should incorporate in our model, we rely on the ACF and PACF.

**Normalizing values –**

Normalizing does not affect stationarity.

Normalizing does not have any effect on model selection. The same model that failed for regular returns would also fail for normalized returns.

**MA model:**

Max no. of lags we are willing to include in MA model can be determined using ACF as MA models aren’t based on past period returns. Determining which lagged values have a significant direct effect on the present-day ones is not relevant. The total accumulated effects accommodate for these unexpected shocks.

The value of each coefficient is between -1 and 1. (to prevent compounded effects exploding in magnitude)

In ACF graph, we expect compounding effect to decrease further back in time we go. We can disregard the 18th lag (significant) and all significant periods that follow it. That’s because we expect their impact on today’s returns to be minimal.

White noise means the errors don’t follow a pattern.

MA models don’t perform well for non-stationary data.

An MA(1) model where all the coefficients are extremely close to 1 is just an approximation of any AR(n) model which only takes the error term from n periods back.

If the data is **non-stationary**, then it **can’t be** considered as **white noise.**

**Stationarity doesn’t mean data is white noise but vice versa is true.**

**A white noise process produces completely random data and all the ACF coefficients should not be significantly different from zero.**

**ARMA**

To apply log likelihood ratio test the simple model should be nested with the complex one(for ARMA models).

(for strict AR and MA models) Any AR(P) is nested in AR(P+1).

Conditions for nesting –

ARMA(P1, Q1) ARMA(P2, Q2)

Second model is nested if an only if 3 conditions are satisfied –

1. P1 + Q1 > P2 + Q2
2. P1 >= P2
3. Q1 >= Q2

If all these 3 conditions are not satisfied –

We manually compare the log-likelihoods and AIC’s of both models –

Our preferred model should yield a higher likelihood and lower AIC.

The volatility in returns might not be fully comprehendable if we use only ARMA models.

ARMA models also perform poorly for non-stationary data.

ARMA models can’t be tested with PACF and ACF for no. of lags

**ARIMA (p, d, q),**

**Best way to find a perfect ARIMA/ARMA model:**

1. **Start with the simple model.**
2. **Examine the ACF of the residual values to get a feel of which lags to use. (see 7th video in ARMA folder)**

p – AR, d – I, q – MA

d – I -> Integration – The number of times we need to integrate the time-series to ensure stationarity.

ARIMA(p, 0 ,q) == ARMA(p, q)

ARIMA(0, 0, q) == MA(q)

ARIMA(p, 0, 0) == AR(p)

ARIMA (p, d, q) -> An ARMA(p, q) model for a newly generated time-series which is stationary. This stationary may require 1,2,…. Or more integrations.

For any integration we lose a single observation.

We only rely on integration, when our time-series doesn’t come from a stationary process. If a single layer of integration(d=1) accomplishes this, any additional ones are unnecessary.

How do we know if the integrated data set is stationary ?

1. Manually create an integrated version of the original time-series.
2. Use the Augmented Dickey-Fuller test.

ARIMA(1,0,1) is equivalent to an ARMA(1,1)

If series is stationary, no need for additional layers of integration.

**ARIMAX**

X is any other variable (or variables) that can affect prices as long as we have the data available for every period.

Such outside factors are known as “exogeneous” variables.

exog = array\_type (values associated with every time period)  
  
**SARIMAX (p,d,q)(P,D,Q,s)**

s expresses how far away the seasonal components would be from the current period.

**ARCH**

PACF of rt and rt^2 can be used to decide on no. of lags to use.

R-squared is a measurement of explanatory variation away from the mean.

Log-likelihood measures goodness of a fit. Therefore, even the simplest ARCH model yields a better estimate than the complex multi-lag ARIMA models.

**GARCH**

It’s been mathematically proven that no higher-order GARCH models outperform the GARCH(1,1) when it comes to variance of market returns. To know the reason checkout 4th video of GARCH.