**AN6902 Time Series Exam Notes**

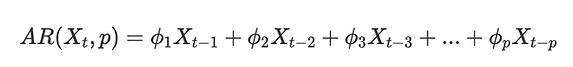
1. What is ARMA

**ARMA - Baseline Model**

ARMA stands for Autoregressive Moving Average. As the name suggests, it is a combination of two parts - Autoregressive and Moving Average.

**Autoregressive Model - AR(p)**

Autoregressive model makes predictions based on previously observed values, which can be expressed as AR(p) where p specifies the number of previous data points to look at. As stated below, where *X* represents observations from previous time points and *φ* represents the weights.



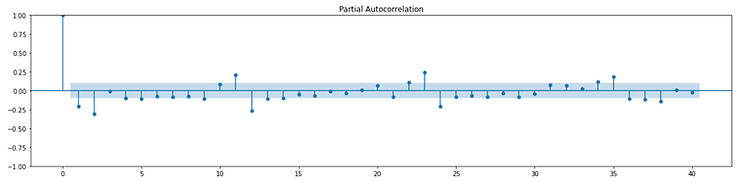
For example, if p = 3, then the current time point is dependent on the values from previous three time points.

**How to determine the p values?**

PACF (Partial Autocorrelation Function) is typically used for determining p values.

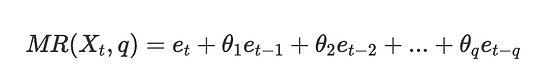
For a given observation in a time series Xt, it may be correlated with a lagged observation Xt-3 which is also impacted by its lagged values (e.g. Xt-2, Xt-1 ).

PACF visualizes the **direct** contribution of the past observation to the current observations. For example, the PACF below when lag = 3 the PACF is roughly -0.60, which reflects the impact of lag 3 on the original data point, while the compound factor of lag 1 and lag 2 on lag 3 are not explained in the PACF value. The p values for the AR(p) model is then determined by when the PACF drops to below significant threshold (blue area) for the first time, i.e. p = 4 in this example below.



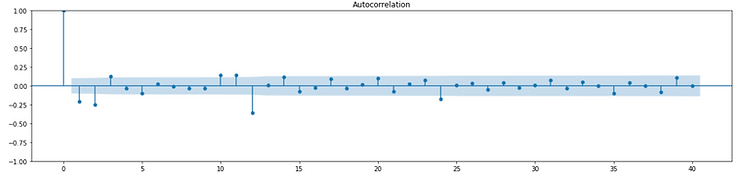
**Moving Average Model - MR(q)**

Moving average model, MR(q) adjusts the model based on the average prediction’s errors from previous q observations, which can be stated as below, where e represents the error terms and θ represents the weights. q value determines the number of error terms to include in the moving average window.



**How to determine the q value?**

Autocorrelation Function (ACF) can be used for determining the q value. It is typically selected as the first lagged value of which the ACF drops to nearly 0 for the first time. For example, we would choose q=4 based on the ACF plot below.



Source:

<https://www.visual-design.net/post/time-series-analysis-arma-arima-sarima>

1. Unit Root Test

## Unit Root

A unit root is a unit of measurement to determine how much stationarity a time series model has. Also called a unit root process, we determine the stochasticity (randomness) of the model using statistical Hypothesis testing. ‘These are statistical hypothesis tests of stationarity that are designed for determining whether differencing is required.’ Although there is a myriad of ways to check for presence of a unit root process, I used an Augmented Dickey Fuller test.

Why is this important? In a model that has a unit root, spikes and shocks to the model will happen. Meaning that a stock price might make a big jump or a big fall that has nothing to do with seasonality. If there is stochasticity in the model the effect of this shock will disappear with time. An important thing to take into consideration when building a broader business model.

**Types of Unit Root Tests:**

* The Dickey Fuller Test/ Augmented Dickey Fuller Test
* The Elliott–Rothenberg–Stock Test, which has two subtypes:

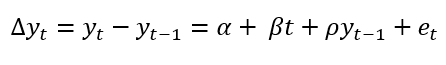
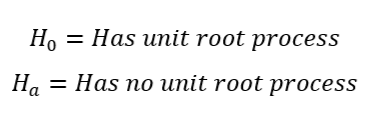
1. The P-test takes the error term’s serial correlation into account.
2. The DF-GLS test can be applied to detrended data without intercept.

* The Schmidt–Phillips Test: Subtypes are the rho-test and the tau-test.
* The Phillips–Perron (PP) Test is a modification of the Dickey Fuller test, and corrects for autocorrelation and heteroscedasticity in the errors.
* The Zivot-Andrews test allows a break at an unknown point in the intercept or linear trend (4).

**Dickey Fuller Test**

The Dickey Fuller Test is a statistical hypothesis test that measures the amount of stochasticity in a time series model. The Dickey Fuller Test is based on linear regression.

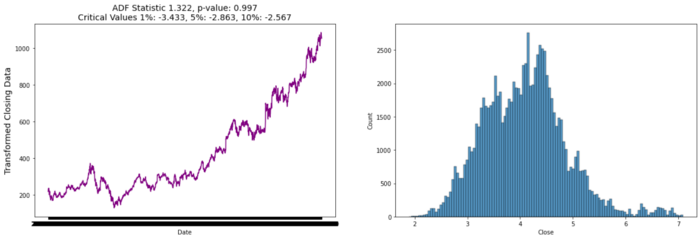
The Dickey Fuller test above actually creates a *t-statistic* that is compared to predetermined critical values. Being below that critical statistic (p-value) allows us to reject the null hypothesis and accept the alternative. If we are above this test statistic we fail to reject the null hypothesis. Critical values are tabulated based on relevant data. The formula for calculating critical values is long and arduous so I won’t cover it here, but know that they are tailored to each new data set.



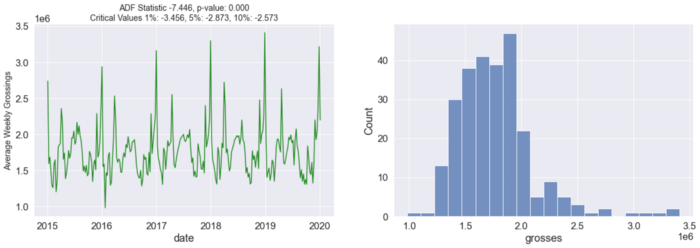
## Augmented Dickey Fuller Test

Serial correlation can be an issue, in which case the Augmented Dickey-Fuller (ADF) test can be used. The ADF handles bigger, more complex models. It does have the downside of a fairly high Type I error rate. The key take away for an Augmented Dickey Fuller is listed in the title, it deals with augmented data. It is a modified algorithm that is equipped to handle data with high dimensionality.

Below are the results from an Augmented Dickey Fuller Test from two different data sets. One being stochastic in nature, the other naught. The first test color coded in purple has a high p value and a test statistic well above the highest critical value. This means it has a unit root process and therefore is stochastic in nature. We fail to reject the null hypothesis.



The second test is color coded in green and has a low p value and a test statistic well below the lowest critical value. This means it *does not* have a unit root process and therefore is non-stochastic in nature. We reject the null hypothesis and accept the alternative.

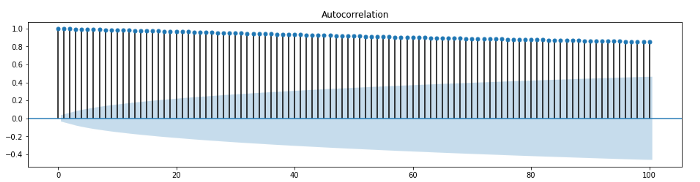


# **Adjusting the Unit Root for Stochasticity**

## Autocorrelation

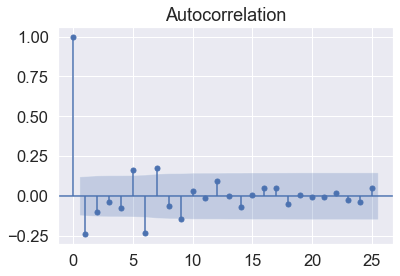
‘Just as correlation measures the extent of a linear relationship between two variables, autocorrelation measures the linear relationship between lagged values of a time series.’(3) Autocorrelation means that the linear model is self aware, it is constantly taking into account a past version of itself. A time lag is when the model measures its current performance in comparison to past performance after a given set of time. This is yet another way of measuring the stationarity of a model. It is a simpler approach than trying to color code your graphs based on the outcome of your AD Fuller results.

Below is the correlogram for Google stocks over a period of 5 years. The lines being extremely close to one and in a slight descending visual pattern indicates high amounts of autocorrelation.



Correlogram for Google stock prices.

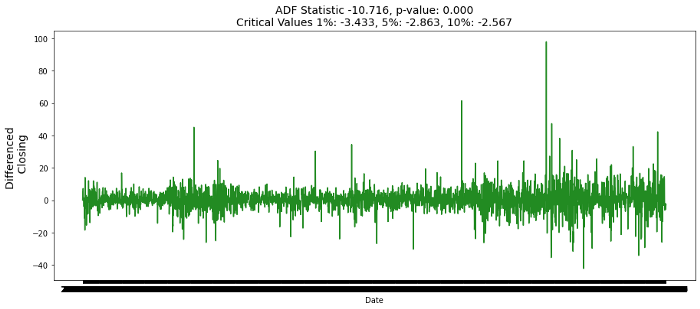
Looking at the second correlogram below, we can see that autocorrelation is very low and random meaning our model is stable. Usually correlograms start with the first lag being fully correlated with itself at 1. The closer to 0 each lag lies, the less autocorrelation present in the model.



**Differencing**

*Differencing* is a technique that can be applied to a data set in order to remove any sort of stochasticity. It is a method to make a non-stationary time series stationary — compute the differences between consecutive observations. This is a technique applied after a unit root test and autocorrelation test have been run.

‘Transformations such as logarithms can help to stabilize the variance of a time series. Differencing can help stabilize the mean of a time series by removing changes in the level of a time series, and therefore eliminating (or reducing) trend and seasonality.’



Source: <https://medium.com/codex/unit-root-in-time-series-38d451d742ce>

1. What is ARCH

Autoregressive Conditional Heteroskedasticity, or ARCH, is a method that explicitly models the change in variance over time in a time series.

Specifically, an ARCH method models the variance at a time step as a function of the residual errors from a mean process (e.g. a zero mean).

A lag parameter must be specified to define the number of prior residual errors to include in the model. Using the notation of the GARCH model (discussed later), we can refer to this parameter as “q“.

The approach expects the series is stationary, other than the change in variance, meaning it does not have a trend or seasonal component. An ARCH model is used to predict the variance at future time steps. In practice, this can be used to model the expected variance on the residuals after another autoregressive model has been used, such as an ARMA or similar.

Source: <https://machinelearningmastery.com/develop-arch-and-garch-models-for-time-series-forecasting-in-python/>

<http://www.cmat.edu.uy/~mordecki/hk/lecture13.pdf>

**Why an ARCH model?**

Autoregressive models can be developed for univariate time-series data that is stationary (AR), has a trend (ARIMA), and has a seasonal component (SARIMA). But, these Autoregressive models do not model a change in the variance over time.

The error terms in the stochastic processes generating the time series were homoscedastic, i.e. with constant variance.

There are some time series where the variance changes consistently over time. In the context of a time series in the financial domain, this would be called increasing and decreasing volatility.

Volatility in Finance: Degree of variation price series over time as measured by the standard deviation of the series. Suppose that Si is the value of a variable on a day ‘i’. The volatility per day is the standard deviation of ln(Si /Si-1).

In time series where the variance is increasing in a systematic way, such as an increasing trend, this property of the series is called heteroskedasticity. This means changing or unequal variance across the series.

If this change in the variance can be correlated over time, then it can be modelled using an autoregressive process, such as ARCH.

**When to apply an ARCH Model?**

In practice, this can be used to model the expected variance on the residuals after another autoregressive model has been used, such as an ARMA or similar.

Since we can only tell whether the ARCH model is appropriate or not by squaring the residuals and examining the correlogram, we also need to ensure that the mean of the residuals is zero.

Crucially, ARCH should only ever be applied to series that do not have any trends or seasonal effects, i.e. that have no (evident) serially correlation. ARIMA is often applied to such a series (or even Seasonal ARIMA), at which point ARCH may be a good fit.

**ARCH Model of Order Unity:**

ARCH(p) model is simply an AR(p) model applied to the variance of a time series.

ARCH(1):

A time-series {ϵ(t)} is given at each instance by ϵ(t) = w(t)\*σ(t)

where w(t) is the white noise with zero mean and unit variance.

Var(x(t)) = σ²(t) = ⍺0+⍺1 \* σ²(t-1)

where ⍺0, ⍺1 are parameters of the model and ⍺0 > 0, ⍺1 ≥ 0 to ensure that the conditional variance is positive. σ²(t-1) is lagged square error.

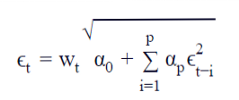
We say that ϵ(t) is an autoregressive conditional heteroskedastic model of order unity, denoted by ARCH(1).

ϵ(t) = w(t)\* σ(t) = w(t)\* ⎷(⍺0 + ⍺1 \*ϵ²(t-1))

similarly ARCH(2):

ϵ(t) = w(t)\* σ(t) = w(t)\* ⎷(⍺0 + ⍺1 \* ϵ²(t-1) + ⍺2 \* ϵ²(t-2))

similarly ARCH(p):



ARCH(p) formula

ϵ(t) = w(t) \* ⎷(⍺0 + ⍺(p) \* ∑ ϵ²(t-i)

where:

* 1. p is the number of lag squared residual errors to include in the ARCH model.
  2. i = (1,2,3,-,-,-, -, p) tells us the number of logged periods of the square error.

Interpretation:

1. If the error is high during the period (t-1), it is more likely that the value of error at the period (t) is also higher.
2. vice versa — If the error is low during the period (t-1) then the value inside sqrt will be low which results in a decreased error in (t).
3. Remember, ⍺1 ≥ 0 for the positive variance.
4. For the stability condition to hold, ⍺1 < 1, otherwise ϵ(t) will be explosive (continue to increase over time).

Note: As mentioned earlier ARCH(1) should only ever be applied to a series that has already had an appropriate model fitted sufficient to leave the residuals looking like discrete white noise.

**Why does this model volatility?**

From variance formula, we can derive the below equation:

Var(ϵ(t)) = ⍺0 + ⍺1 \* Var(ϵ(t-1))

We can say that the variance of the series is simply a linear combination of the variance of the prior element of the series.

Source: <https://medium.com/@ranjithkumar.rocking/time-series-model-s-arch-and-garch-2781a982b448>

1. Extension of ARCH

(Don’t know what is this)

This article provides an overview of two time-series model(s) — ARCH and GARCH. These model(s) are also called volatility model(s). These models are exclusively used in the finance industry as many asset prices are conditional heteroskedastic.

ARCH — Autoregressive Conditional Heteroskedasticity

GARCH — Generalized Autoregressive Conditional Heteroskedasticity

These models relate to economic forecasting and measuring volatility.

Some of the techniques adopted in the finance sector — ARCH, ARCH-M, GARCH, GARCH-M, TARCH, and EGARCH.

ARCH model is concerned about modelling volatility of the variance of the series.

These model(s) deals with stationary (time-invariant mean) and nonstationary (time-varying mean) variable(s).

Some of the real-time examples where ARCH model(s) applied: Stock prices, oil prices, bond prices, inflation rates, GDP, unemployment rates, etc.

* GARCH
* NGARCH
* Nonlinear Asymmetric GARCH(1,1) (NAGARCH)
  + For stock returns, parameter theta is usually estimated to be positive; in this case, it reflects a phenomenon commonly referred to as the "leverage effect", signifying that negative returns increase future volatility by a larger amount than positive returns of the same magnitude
* IGARCH
  + Integrated Generalized Autoregressive Conditional heteroskedasticity (IGARCH) is a restricted version of the GARCH model, where the persistent parameters sum up to one, and imports a unit root in the GARCH process.
* EGARCH
  + The exponential generalized autoregressive conditional heteroskedastic (EGARCH) model by Nelson & Cao (1991) is another form of the GARCH model.
  + This is particularly useful in an asset pricing context.
* GARCH-M
  + The GARCH-in-mean (GARCH-M) model adds a heteroskedasticity term into the mean equation.
* QGARCH
  + The Quadratic GARCH (QGARCH) model by Sentana (1995) is used to model asymmetric effects of positive and negative shocks.
* GJR-GARCH
  + Similar to QGARCH, the Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model by Glosten, Jagannathan and Runkle (1993) also models asymmetry in the ARCH process.
* TARCH
  + Threshold ARCH (TARCH) model by Zakoian (1994) is similar to GJR GARCH. The specification is one on conditional standard deviation instead of conditional variance.
* fGARCH
  + Hentschel's fGARCH model, also known as Family GARCH, is an omnibus model that nests a variety of other popular symmetric and asymmetric GARCH models including APARCH, GJR, AVGARCH, NGARCH, etc.
* COGARCH
  + continuous-time generalization of the discrete-time GARCH(1,1) process
* ZD-GARCH
  + Unlike GARCH model, the Zero-Drift GARCH (ZD-GARCH) lets the drift term in the first order GARCH model.
  + ZD-GARCH model is always non-stationary, and its statistical inference methods are quite different from those for the classical GARCH model.
* Spatial GARCH
  + spatial equivalent to the temporal generalized autoregressive conditional heteroscedasticity (GARCH) models. In contrast to the temporal ARCH model, in which the distribution is known given the full information set for the prior periods, the distribution is not straightforward in the spatial and spatiotemporal setting due to the interdependence between neighbouring spatial locations.

Source: <https://medium.com/@ranjithkumar.rocking/time-series-model-s-arch-and-garch-2781a982b448>

<https://en.wikipedia.org/wiki/Autoregressive_conditional_heteroskedasticity>

1. What is GARCH

Generalized Autoregressive Conditional Heteroskedasticity, or GARCH, is an extension of the ARCH model that incorporates a moving average component together with the autoregressive component.

Specifically, the model includes lag variance terms (e.g. the observations if modelling the white noise residual errors of another process), together with lag residual errors from a mean process. Essentially, by comparison, in the ARCH(q) process the conditional variance is specified as a linear function of past sample variances only, whereas the GARCH(p,q) process allows lagged conditional variances to enter as well.

The introduction of a moving average component allows the model to both model the conditional change in variance over time as well as changes in the time-dependent variance. Examples include conditional increases and decreases in variance.

As such, the model introduces a new parameter “p” that describes the number of lag variance terms:

p: The number of lag variances to include in the GARCH model.

q: The number of lag residual errors to include in the GARCH model.

A generally accepted notation for a GARCH model is to specify the GARCH() function with the p and q parameters GARCH(p, q); for example GARCH(1, 1) would be a first order GARCH model.

A GARCH model subsumes ARCH models, where a GARCH(0, q) is equivalent to an ARCH(q) model. The main advantage of GARCH over ARCH is that, ARCH(p) modelling requires relatively high values of p for good fitting, while GARCH(1,1) is usually enough for fitting financial data.

|  |  |
| --- | --- |
| ARCH | GARCH |
| AR model with conditional heteroskedasticity | Model variance with AR(p), generalised ARCH |
| Does not consider volatility of the previous period | Considers volatility of the previous period |
| Does not model change in variance over time | Models conditional change in variance over time |

**What is a GARCH model?**

Generalized Autoregressive Conditional Heteroskedasticity, or GARCH, is an extension of the ARCH model that incorporates a **moving average** component together with the **autoregressive** component.

Bollerslev (1986, Journal of Econometrics) generalized Engle’s ARCH model and introduced the GARCH model.

Introduction of moving average component allows the model:

1. To model the conditional change in variance over time.
2. Changes in the time-dependent variance.

Examples include conditional increases and decreases in the variance.

Thus GARCH is the “ARMA equivalent” of ARCH, which only has an autoregressive component. GARCH models permit a wider range of behaviour more persistent volatility.

**GARCH Model of Order p, q — GARCH(p,q):**

GARCH(1,1):

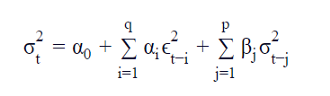
Here we are going to consider a single autoregressive lag and a single “moving average” lag. The model is given by the following:

ϵ(t) = w(t) \* σ(t)

ϵ(t) =w(t) \* ⎷(⍺0 + ⍺1 \*ϵ²(t-1)) + β1 \* σ²(t−1)

Similarly GARCH(p,q):

A time-series {ϵ(t)} is given at each instance by ϵ(t) = w(t)\*σ(t) and σ²(t) is given by:



where α(i) and β(j) are parameters of the model.

⍺0 > 0, ⍺i ≥ 0, i =1, … q, β≥ 0, j = 1, … p imposed to ensure that the conditional variances are positive.

Here we are adding moving average term, that is the value of σ² at t, σ²(t), is dependent upon previous σ²(t-j) values.

**Interpretation:**

1. The large value of β1 causes σ(t) to be highly correlated with σ²(t−1) and gives the conditional standard deviation process a relatively long-term persistence, at least compared to its behavior under an ARCH model.
2. For p = 0 the process reduces to the ARCH(q) process.
3. For p = q = 0, ϵ(t) is simply white noise.

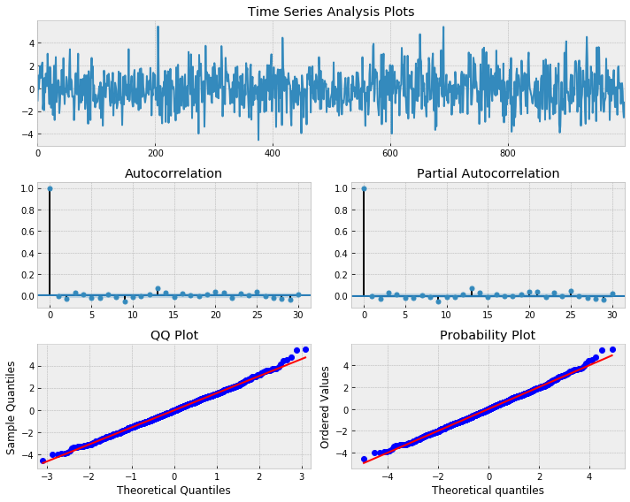
Source: <https://medium.com/@ranjithkumar.rocking/time-series-model-s-arch-and-garch-2781a982b448>

1. ARCH-GARCH Condition Model

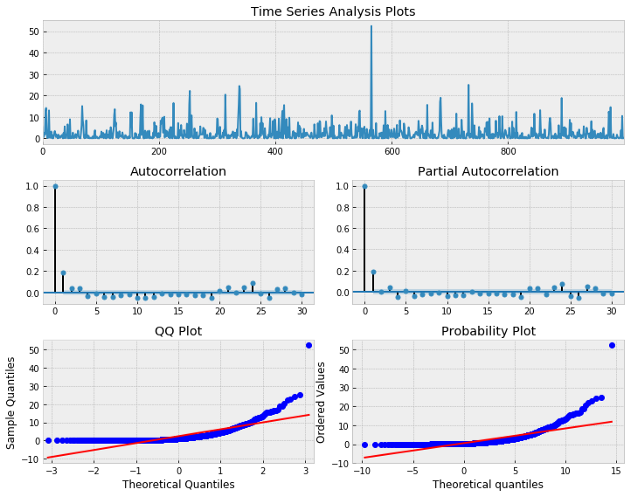
**How to configure the ARCH and GARCH Model(s):**

1. The configuration for an ARCH model is best understood in the context of **ACF and PACF plots of the variance** of the time series.
2. This can be achieved by subtracting the mean from each observation in the series and squaring the result, or just squaring the observation if you’re already working with white noise residuals from another model.
3. The ACF and PACF plots can then be interpreted to estimate values for p and q, in a similar way as is done for the ARMA model.

ARCH(1) model:

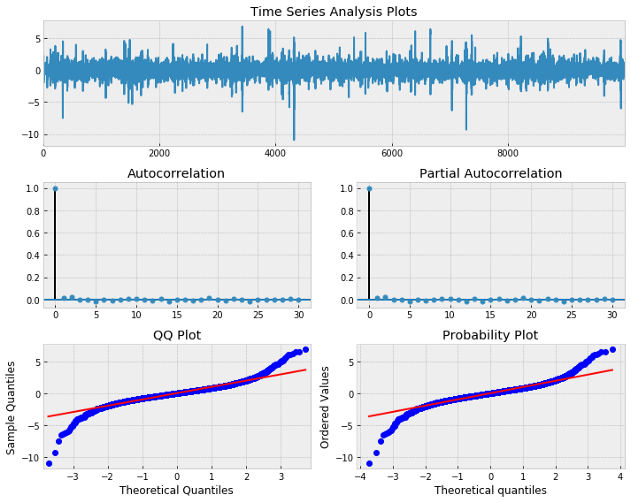


ARCH(1) Squared model:



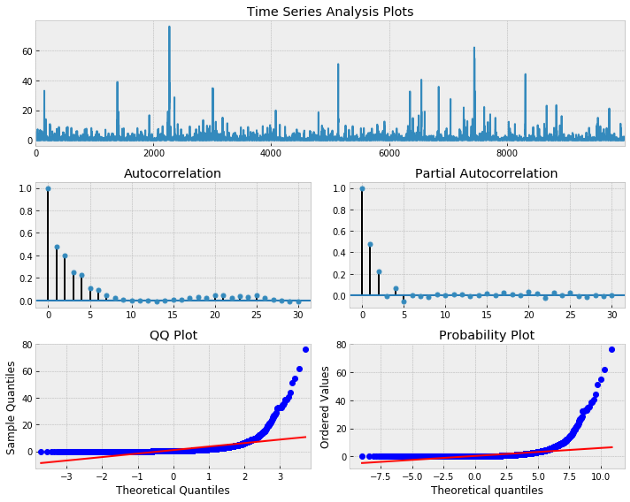
Observation: ACF and PACF seem to show significance at lag 1 indicating an AR(1) model for the variance may be appropriate.

GARCH(1,1) model:



Observation: We have noticed that overall this process closely resembles white noise, however, take a look when we view the squared eps series.

GARCH(1,1) Squared model:



Observation: we can observe clearly autocorrelation present and the significance of the lags in both the ACF and PACF indicates we need both AR and MA components for our model.

Source: <https://medium.com/@ranjithkumar.rocking/time-series-model-s-arch-and-garch-2781a982b448>

1. Why do we use these models

In time series analysis for forecasting new values, it is very important to know about the past data. More formally, we can say it is very important to know about the patterns which are followed by the values with time. There can be many reasons which cause our forecasted values to fall in the wrong direction. Basically, a time series consists of four components. Variation of those components causes the change in the pattern of the time series. These components are:

* **Level**: It is the main value that goes on average with time.
* **Trend**: The trend is the value that causes increasing or decreasing patterns in a time series.
* **Seasonality**: This is a cyclic event that occurs in time series for a short time and causes the increasing or decreasing patterns for a short time in a time series.
* **Noise**: These are the random variations in the time series.

The combination of those components with time causes the formation of a time series. Most time series consists of the level and noise/residual and the trend or seasonality are the optional values. They may take part or they may not.

The combination of the components in time series can be of two types:

* Additive
* Multiplicative

**Additive time series**

if the components of the time series are added together to make the time series. Then the time series is called the additive time series. By visualization, we can say the time series is additive if the increasing or decreasing pattern of the time series is similar throughout the series. The mathematical function of any additive time series can be represented by:

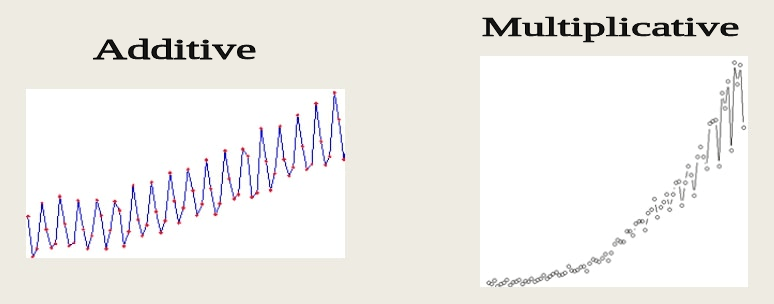
y(t) = level + Trend + seasonality + noise

**Multiplicative time series**

If the components of the time series are multiplicative together, then the time series is called the multiplicative time series. By visualization, if the time series is having exponential growth or decrement with time then the time series can be considered as the multiplicative time series. The mathematical function of the Multiplicative time series can be represented as.

y(t) = Level \* Trend \* seasonality \* Noise

The image below represents the additive and multiplicative time series.



In the above image, we can see the difference in the growth of values. In additive, it is quite slower and has a proper trend but on the other hand, we can see that the time series is growing exponentially with the time

Instead of knowing the type of time series, it is better to know about the component of the time series. As we earlier discussed, the time series is the composition of level, trend, season and residuals. It is much better to know all the components in a time series for better understanding and making a model more accurate for better forecasting values.

In time series if the components are available in uneven amounts then it can cause the model to predict wrong values.

Source: <https://analyticsindiamag.com/why-decompose-a-time-series-and-how/>

1. Why cointegration is important

Stationarity is a crucial property for time series modelling. The problem is, in practice, very few phenomena are actually stationary in their original form. The trick is to employ the right technique for reframing the time series into a stationary form. One such technique leverages a statistical property called cointegration.

Cointegration forms a synthetic stationary series from a linear combination of two or more non-stationary series. Cointegration tells you that, although two series move independently, the average distance between them remains relatively constant. Essentially, cointegration aims to uncover causal relationships among variables by determining if the stochastic trends in a group of variables (e.g. spot vs. futures price) are shared by the series. Cointegration essentially means two variables have a long-run relationship.

Two series are cointegrated if they are both individually unit-root nonstationary (integrated of order 1: I(1)) but there exists a linear combination that is unit-root stationary (integrated of order 0: I(0)). If any of the individual series are already stationary, then cointegration would be redundant since the linear combination would heavily emphasize the stationary series.

**Cointegration vs Correlation**

Although the correlation coefficient and cointegration both describe some underlying relationship between variables, the two properties are not synonymous. It is very possible for two time series to have weak/strong correlation but strong/weak cointegration.

**Cointegration Testing**

Engle-Granger Procedure

This is the original procedure for testing cointegration developed by Robery Engle and Clive Granger in their seminal paper Engle and Granger [1987].

The 2-step procedure is easy to follow and paints a good picture of the general idea behind cointegration. The idea is to first verify that the individual series are indeed integrated of order I(1) (being non-stationary).

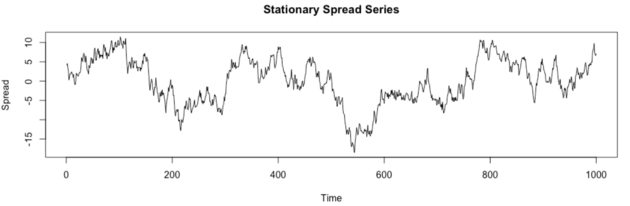
Then we regress one on the other using standard OLS and check if the residual series is integrated of order I(0) (suggesting stationarity).

If the residuals are stationary, then we can extract the coefficients from the OLS model and use those to form a stationary linear combination of the two time series.

We’ll run this on the first two cointegrated series. First let’s test each series to confirm that they are indeed non-stationary (there exists a unit-root) using the popular ADF test. We next fit a linear model using OLS and check the residuals for stationarity.

We are comfortable with assuming that we now have a stationary series to work with. Because we are unable to reject the null hypothesis of stationarity, the Engle-Granger test suggests a cointegrating relationship exists. The cointegrating equation is given by:

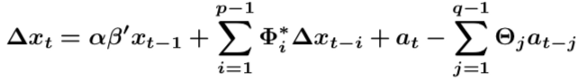




Johansen Test

There are a few shortcomings of the Engle-Granger procedure. The first being that it is incapable of simultaneously testing for cointegrating relationships among multiple series. Additionally, regressing y2 on y1 instead of y1 on y2 (as we did above) would yield a different cointegrating vector. Thus, the choice of which series to regress on the other is somewhat arbitrary. The other shortcoming that it overlooks the underlying error-correction model influencing the spread series.

Conveniently, all of the above shortcomings can be addressed through the Johansen procedure. This procedure estimates cointegrated VARMA(p,q) in the VECM (vector error-correction model) form for m cointegrating relationships between k different series in xₜ.

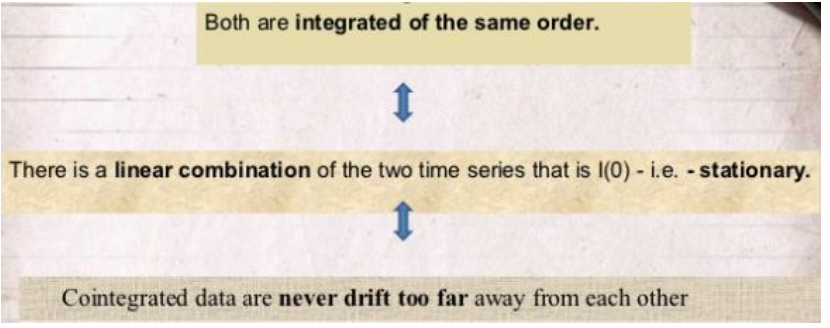


Source: <https://medium.com/analytics-vidhya/cointegration-for-time-series-analysis-1d0a758a20f1>

When estimating an econometric model with nonstationary variables, this variables have to be cointegrated in order for the model to be meaningful

The Granger Theorem states that if two series are non-stationary (i.e. I(1)), there can be a linear combination of the two series that is stationary, in that case we can say the two variables are cointegrated.

Economically speaking, two variables will be cointegrated if they have a long-term or equilibrium relationship between them.



Examples of possible Cointegrating Relationships in finance:

• spot and futures prices

• ratio of relative prices and an exchange rate

• equity prices and dividends

• Short and long term interest rate

No cointegration implies that series could wander apart without bound in the long-run.

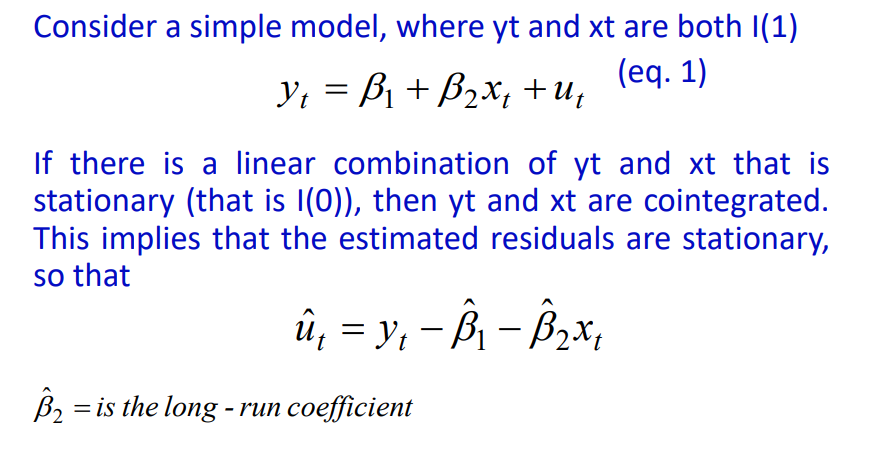
Source: <http://www.ecostat.unical.it/Algieri/Didattica/Financial%20Markets/Tutorials/Lecture_L5_Time%20Series%20ECM.pdf>

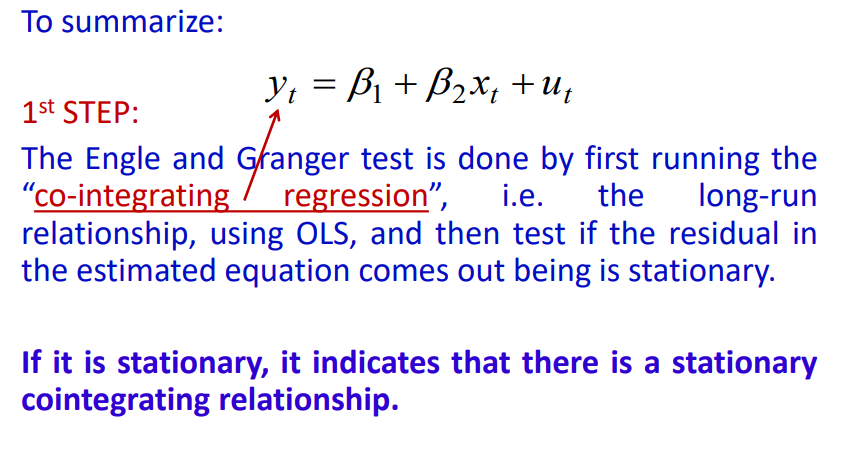
1. Purpose of Error Correction Model (ECM)

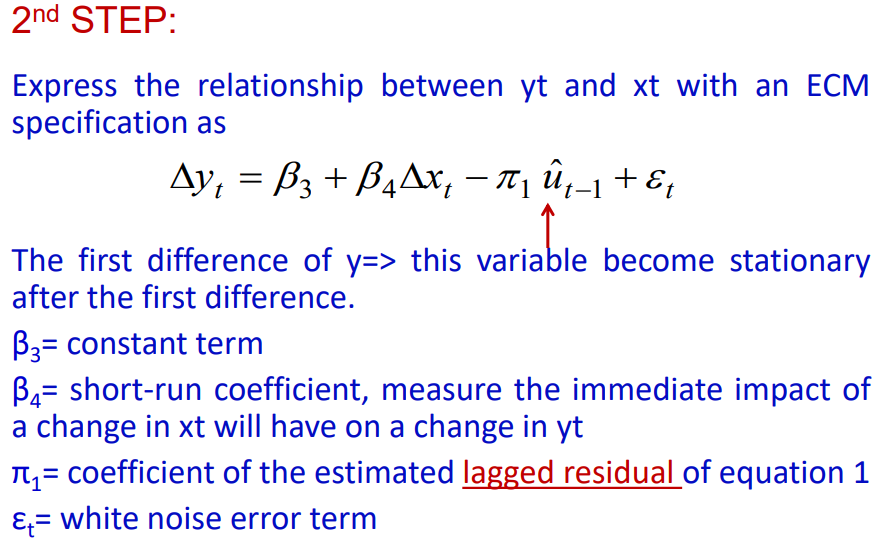
If two series are cointegrated, an error correction model (ECM) could be appropriate rather than a model in pure first difference (i.e. I(1) form) because it would enable us to capture the long-run relationship between series as well as the short-run. The error-correction infers that the last-period’s deviation from a long-run equilibrium, called the error, influences its short-run dynamics.

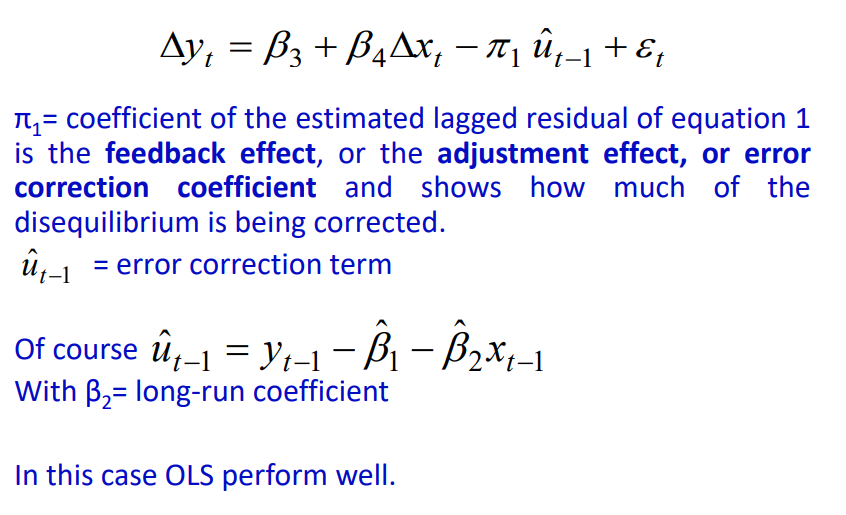
An Error Correction Model (ECM) is the standard way to model time series equations. The ECM makes it possible to deal with nonstationary data series and separates the long and short run. ECM models make no ad hoc assumptions of how the variables change over time.

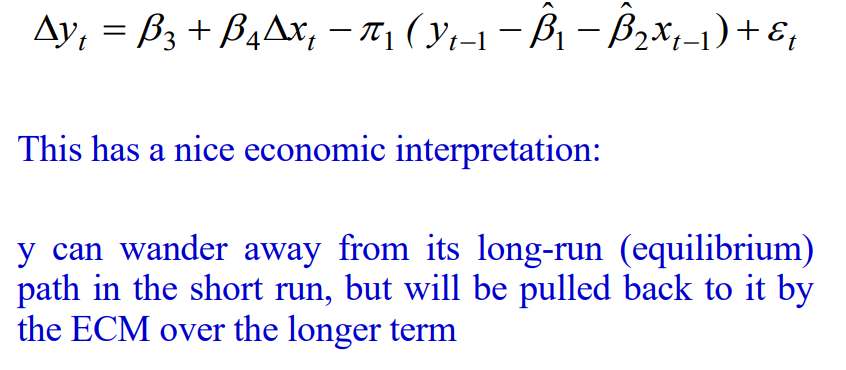
Two Steps Engle and Granger:











Source: <http://www.ecostat.unical.it/Algieri/Didattica/Financial%20Markets/Tutorials/Lecture_L5_Time%20Series%20ECM.pdf>

**Other notes:**



**Stationarity**

Time series data can be classified into stationary and non-stationary. Stationarity is an important property, as some models work well when the data is stationary. However, time series data often possesses the non-stationary property. Therefore, we need to understand how to identify non-stationary time series and how to transform it through various techniques, e.g. differencing.

Stationary data is defined as not depending on the time component and possesses the following characteristics, constant mean, constant variance overtime and constant autocorrelation structure (i.e., the pattern of autocorrelation does not change over time), without periodic or seasonal component.

**Techniques to Identify Stationarity**

The most straightforward method would be examining the data visually. For example, the visualization above indicates that the time series follows an upward trend and its mean values increase over time, suggesting that the data is non-stationary. To quantify it stationarity, we can use following two methods.

Firstly, **ADF (Augmented Dickey Fuller) test** examines stationarity based on the null hypothesis that data is non-stationary and alternative hypothesis that data is stationary. If the p-value generated from the ADF test is smaller than 0.05, it provides stronger evidence to reject that data is non-stationary. (i.e. if p<0.05, it is stationary)

Secondly, ACF **(Autocorrelation Function)** summarizes the two-way correlation between the current observation against past observations. For example, when the lag=1 (x-axis), ACF value (y-axis) is roughly 0.85, meaning that the average correlation between all observations and their previous observation is 0.85. In the later section, we will also discuss using ACF to determine the moving average parameter.

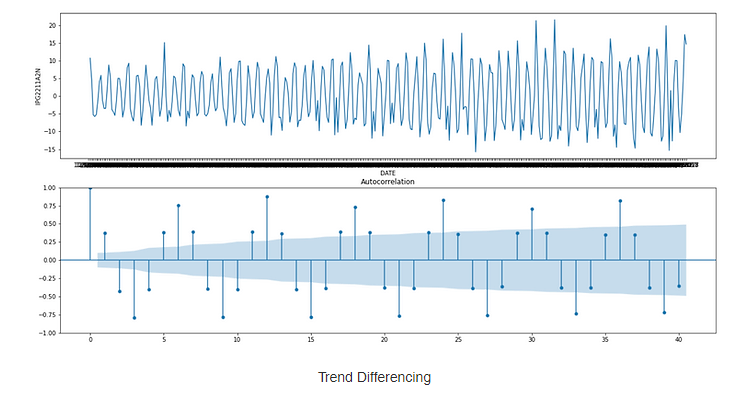
For non-stationary data, ACF drops to 0 relatively slowly, because non-stationary data may still appear highly correlated with previous observations, indicating that time component still plays an important role. The diagram above shows the ACF of the original time series data, which decreases slowly thus very likely to be non-stationary.

### Stationarity and Differencing

Differencing removes trend and seasonality by computing the differences between an observation and its subsequent observations, differencing can transform some non-stationary data to stationary.

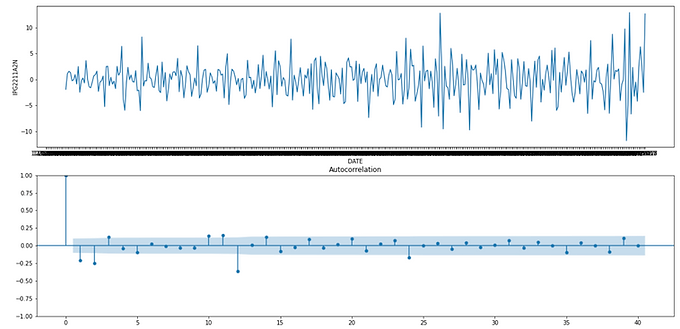
**1. remove trend**

We can plot the time series chart as well as the ACF plot after applying trend differencing. As shown below that the trend has been removed from the data and data appear to have constant mean. The next step is to address the seasonal component.



**2. remove seasonality**

From the ACF plot above, we can see that observations are more correlated when lag is 12, 24, 36 etc, thus it may follow a lag 12 seasonal pattern. After removing the seasonal pattern, the time series data below becomes more random and ACF value drops to a stable range quickly.



## Stochastic

A fancy word for randomness, for the most part. Stochasticity (Also a Random Walk with a Drift) can be defined as a variable or process that has uncertainty in it. The features lack dependence between one another. Stochasticity would be used over the word randomness when the probability of a feature is important. Whereas a term such as ‘random sampling’ just refers to having a lack of bias but does not inference an outcome.

**Serial Correlation (H0 = no serial correlation)**

If untreated, serial correlation can lead to a number of issues:

* 1. Reported standard errors and t-statistics are invalid (even asymptotically)
  2. Coefficients may be biased
  3. In the presence of lagged dependent variables, OLS estimates are biased and inconsistent

Overall, serial correlation has a large impact on standard errors and efficiency of estimators.

According to the correlogram of residuals, if there are no serial correlation, the AC and PACF at all lags should be near zero and all Ljung-Box Q-statistics should be insignificant.   
Clearly, this is not the case in the correlogram which shows substantial and persistent autocorrelation in residuals.

There are two other test statistic methods to test for serial correlation – the Durbin Watson and Breusch-Godfrey tests.

To test the hypothesis of no serial correlation, compare the reported ***Durbin-Watson stat*** to a table of critical values. E.g. DW = 0.027, we soundly reject the null hypothesis of no serial correlation. For **Breusch-Godfrey test,** the null hypothesis of no serial correlation is easily rejected.

To resolve serial correlation, the data can be differenced. Taking the first-order differences addresses a number of issues that arises in the time series data:

1. It eliminates most serial correlation
2. It de-trends the data
3. It transforms an I(1) process to an I(0)

After differencing,

1. The time trend now is not significant (Prob > 0.05). Taking the first-differences has detrended the data.
2. The R-squared value is much lower now reflecting the fact that it is harder to fit differences data (compared to levels).

**Heteroskedasticity and Autocorrelation (H0 = no heteroskedasticity)**

In many financial time series, the conditional variance of the error term depends on past values of the error term. This is also known as autoregressive conditional heteroskedasticity (ARCH).

When testing for heteroskedasticity, **residuals should not be serially correlated.** Any serial correlation will generate invalidate tests for heteroskedasticity. Since we have identified that there is no serial correlation, we can then test for heteroskedasticity.

You can correct both heteroskedasticity and autocorrelation of unknown form using the HAC Consistent Covariance (Newey-West).

We can perform the White test (i.e. test of H0 = no heteroskedasticity, against heteroskedasticity of unknown, general form).

Based on the test statistics, we reject the null hypothesis of no heteroskedasticity, which means that the error term is heteroskedastic and standard errors should be adjusted.