Advanced Financial Instruments for Sustainable Business and Decentralized Markets

Prof. Abhinava Tripathi
Department of Management Sciences
Indian Institute of Technology, Kanpur

Week 7
Lesson 23

In this lesson, we start the discussion with a recap on the concept of stationarity. We discuss the types of non-stationarity and how to resolve each of these types namely stochastic non-stationarity and non-stationarity with deterministic trend. We also discuss the visual examination of security prices corresponding to different ARIMA processes and we also visualize different kinds of non-stationarity and the differences across these processes. Next we discuss the unit root test and test of stationarity including ADF that is Augmented K-Fuller, PP test and KPSS test. Lastly, we discuss the concept of cointegration, mean reversion and error correction models. We also discuss the Angel Grengell two-step approach to ECM and cointegration.

In this video, we will recap our understanding of time-series stationarity. We will also discuss some of the problems associated with dealing and modeling non-stationary series. Recall, in our earlier discussion we said a strictly stationary process is a time-series process Yt when we pick a set of sequence let us say Yt 1 to Yt n and the joint distribution of these observations remains same whether we pick another sequence from Yt 1 plus m to Yt n plus m. And similarly, we can move and take any such sequence and if the joint probability and this f is notation for joint probability distribution if it remains same for Yt 1 to Yt n a sequence of n terms to Yt 1 plus m to Yt n plus m, then this series is said to be strictly stationary.

But this strict stationarity condition is a rather stringent one and less practical. So, in a practical sense, we make use of what we call as weakly stationary process. What is weakly stationary process or covariance stationary process? If you recall, we said that it is a process for which the mean of the distribution is constant over time, its variance is constant as well as the auto covariance is also constant. So, it does not matter in the sequence which t1, t2 or tn we are looking at what matters is what is the difference between this t1 and t2. So, for example, if you are looking at two sequence let us say t1 to t5, their auto covariance will be same as t2 to t6.

So, what matters is not at what time they are only their difference and you can identify what would be the correlation. So, rather in simple terms, we are saying a weakly stationary process is a process where the process mean or expected value of the process, its variance which is the second moment and auto correlation structure or auto covariance remain constant across time. We also noted that it is difficult to make sense or make a lot of use of the auto covariance which is this expression expectation of Yt minus mu which is the mu is nothing but expected value of Yt itself. And similarly, Yt plus s this expression is auto covariance, but because of scale thing this auto covariance is not very difficult to make sense, it is slightly difficult to understand and therefore, what we do is we come to something called auto correlation by dividing this auto covariance with the variance term. What is variance? Variance is the auto covariance of Yt with its own self, with its own term neither lead nor lag same.

So, this tau s is the auto correlation, there is it is given that obviously, if when lag is equal to 0, then auto correlation will be 1, because we are computing correlation of Yt with its own self and for all the other lags, it will lie between plus 1 and minus 1. So, this is called auto correlation which is a slightly more reformed measure of auto covariance. Now, let us discuss some of the problems with time series stationarity. In a practical sense, when you have two random series, let us say Yt and Xt which are purely random with no connection with each other, when you will try to compute correlation or regression between them or any such statistical relationship, it should be 0. However, when you run the regression of these two Yt and Xt even though they are random, but if they are non-stationary and you sample, let us say you sample certain time periods for this process, let us say t1 to tn or tn plus 1 to tn and the series are stationary across different samples, you plot t statistic or adjusted R square stats which exactly a measure of the relationship statistical strength of this relationship, you will find a rather high value of adjusted R square.

Ideally, it should have been very small, maybe even 0, but when you sample these process maybe 1000 times or 500 times and plot them, this kind of adjusted R square measure will appear in fact, many times it is even more than 50 percent that is 0.5. Similarly, in a strict theoretical sense the t statistic from the regression should follow the distribution, which means almost 95 percent of the observations, 95 percent of the observation should lie between plus minus 2 close to plus minus 2. However, this kind of t statistics appears where the large number of values are beyond that plus minus 2 range, in fact, a lot of them even more than 90 percent, which indicates this is regression is rather spurious, that means our results are not correct and this is ascribed to non-stationary properties. So, to conclude we saw that in practice we rather follow a weak stationarity condition that is the mean variance and auto covariance structure should remain constant and if our series is non-stationary and did not do anything about it, our regression or any such measure of statistical strength is spurious in nature.

In this video, we will discuss the two key types of non-stationarity and their key characteristics. Let us introduce the two key types of non-stationarity in financial econometrics. The two key important non-stationarity are random walk model with drift, this is also called stochastic non-stationarity, it appears like this process yt equal to mu, mu is the drift plus yt minus 1 plus the error term mu t. Now, this mu is the drift and yt minus 1 comes with a correlation of 1. We will have a more detailed discussion in later subsequent slides and videos, but please notice the coefficient 1 has the implication that any shock to this series, historically shocks to the series do not die away.

$$y_t = \mu + y_{t-1} + \mu_t$$

We will elaborate this in more detail and that is why this is called stochastic non-stationarity, how to deal with it, what is the nature of this kind of non-stationarity, we will discuss shortly. mu t is the error term which is often considered to be a white noise IID process as we discussed earlier, that means identically and independently distributed and that identical distribution is often considered to be as normal. So, this notation is for a normal distribution with zero mean and constant variance and this is ascribed to this white noise mu t process. The second non-stationarity is called deterministic trend process or trending process, it appears like yt equal to alpha plus beta t plus mu t, the non-stationarity comes because of this part, the violation in non-stationarity comes because of this part because of this trending time term, the mean and other things are not constant over time. So, if you recall in the weak stationarity, we assumed a constant mean of the process, variance and auto-correlation structure because of this time trend, those terms are changing with time and therefore, this is deterministic trend process with non-stationarity component.

$$y_t = \alpha + \beta_t + \mu_t$$

Now, how to deal with these non-stationarities, we will discuss in subsequent videos, but one important concept will again recap the white noise IID process. So, if you recall, a white noise process is a process with no discernible structure. In financial econometrics, we generally assume that the mean of this process is zero. If you recall our previous discussion, particularly for security market returns, this is a very good assumption because it is difficult to believe that any particular stock will give you abnormal returns in some kind of this mean, which is greater than zero or if it is less than zero, then opposite argument can be applied with short sale logic. That means you will sell the stock short.

Also, for this white noise IID process, we said that its variance is constant and because there is no deterministic structure, its auto-correlation structure is zero, there is no auto-covariance, there is no discernible pattern for all the T's which are not same. So, this white noise process has a zero mean, finite constant variance and uncorrelated across time. We also put a normality assumption which is a good assumption to follow. Now, why this white noise IID modeling need for the error term u t? The idea here is that when you model a process y t, maybe some kind of ARIMA ARMA model and then you extract the residuals from the model. We assume that if the modeling has been good, then the residuals should be very clean, they should not have any pattern or some kind of structure that requires modeling.

And therefore, when we are doing the model diagnostics, it is rather more residual diagnostic wherein we extract the residual. The assumption is that these residuals asked theoretically should be white noise IID process and if they deviate from a white noise process, that means we find some kind of auto-correlation structure or some modeling possibility, some systematic pattern in this error term. Our residual diagnostics or model diagnostics suggest that our modeling is not very adequate and efficient and we need to put some more efforts in modeling our y t. So, the criteria of a good model in terms of white modeling y t time series is that the residual should be error free when we perform the residual diagnostics. They should have no discernible structure, there should be any, there should not be any systematic pattern in this mu t, it should be a white noise process that is what we assume.

So to summarize in this video, we saw the types of non-stationarity which is one is time series or deterministic non-stationarity and another is the stochastic non-stationarity. In subsequent videos, we will see how to tackle with these form of non-stationarity. In this video, we will discuss stochastic non-stationarity and how to account for it while modeling the data. Think of a generalized AR autoregressive AR1 process like this, y t equal to mu plus phi t minus 1 plus mu t, mu here represents the drift and y t minus 1 is the autoregressive term with its coefficient phi. Now what would happen if phi greater than 1? This would become an explosive process because each of the previous term has an accumulative impact which is increasing.

$$y_t = \mu + \phi y_{t-1} + \mu_t$$

For example, the impact of y t minus 1, let us say a shock came at time t minus 1 which is mu t minus 1. The impact of the shock on next period y t is a multiple of phi. In the next period phi square, as long as this phi is greater than 1, the impact of the shock on the process will explode which is not a very desirable property and a property which does not describe the real economic and finance data. So, generally we tend to ignore this part. However, there is a case for phi equal to 1, there are periods of extreme turbulence such as COVID, COVID period or similar crisis period where this phi becomes equal to 1.

What is the implication of this? The implication is that if there was a shock in t minus 1 which is mu t minus 1, this shock in future periods will not die down, it will keep on multiplying phi to the power 1, 2, 3 and if phi equal to 1, the shock will not go away and mu t minus 1 will always persist in the series. So, this kind of non-stationarity has been found to exist. Let us see more mathematically how it works. Let us recall our generic case of this AR1 process. For time being, we will assume that drift to be 0.

$$y_t = \phi y_{t-1} + \mu_t$$

As we said, we can assume that drift to be 0 for ease of modeling. Now, if you look at the term yt minus 1 in a similar way, we will subtract 1 from the subscripts to get this kind of expression that is yt minus 1 in terms of yt minus 2 and yt minus 2 in terms of yt minus 3. Now, let us try to express yt in terms of its lag. So, if we can substitute instead of yt minus 1, we can substitute this expression to get yt in the form of yt minus 2 and here as you will see, this term will become phi square yt minus 2. Now, for this yt minus 2, we can also go further back to express this in terms of yt minus 3 and then the term will be yt cube and so on up till let us say the first point is phi to the power t, which is the first term as y0.

$$y_{t-1} = \phi y_{t-2} + \mu_{t-1}$$

$$y_{t-2} = \phi y_{t-3} + \mu_{t-2}$$

$$y_{t-2} = \phi(\phi y_{t-2} + \mu_{t-1}) + \mu_t = \phi^2 y_{t-2} + \phi\mu_{t-1} + \mu_t$$

$$y_t = \phi^T y_0 + \phi\mu_{t-1} + \phi^2\mu_{t-2} + \phi^3\mu_{t-3} + \cdots + \phi^T\mu_0 + \mu_t$$

So, let us say first the series is oriented at y0 and the corresponding time period was t. So, this phi to the power t, y0 is the first term yt and then subsequent terms will be in the form of these white noise processes mu t minus 1, mu t minus 2 and so on. Now, a very interesting thing is happening here. If this phi is greater than 1, like we said the shock will, none of these shocks will die away, even till mu naught the shock will explode in fact. However, if this phi equal to 1, it has a very undesirable property, whether this is a more recent shock or a shock which was much further away in time that is mu naught, this will also become 1.

So, mu naught or mu t minus 1, mu t minus 2, whatever shock exists, it never dies down. However, for a financial or economic time series, a desirable property is that such shocks should decline over time, their impact from the series should decline. However, as long as the value phi equal to 1, it does not die. Only in the case where this phi is less than 1, in that case this phi to the power t tends to 0 as t increases and then only the shocks to the system die away. So, a desirable property that these shocks or these coefficients to mu t should be less than 1 in that or rather in this AR one process, if the phi is the autocorrelation coefficient for the first order process that is y t minus 1, it should be less than 1.

However, as we said there are, there is a type of data series that exists where phi equal to 1 often during crisis periods like COVID. In those cases, the shocks persist in the system and never die away and you have this kind of process y t equal to phi y naught plus this summation mu t into phi t and since phi equal to 1, it will be simply the summation of mu t s over time that these shocks will never die away, no matter how further away in time you are. Let us see how to deal with this kind of non-stationarity. This kind of non-stationarity, this is the generic expression we have taken the drift also. This is often referred to as random walk non-stationary process or a stochastic stationary process with drift component.

This is the drift component. The easiest way to remove this kind of non-stationarity is through taking first differences. What do we mean by this? So, we have this y t process. If I take this y t minus 1 on the LHS, we are left with what we call as first difference which is y t minus y t minus 1 which can be written as delta y t. So, delta y t is y t minus y t minus 1 which is also called first difference and you recall we also said that lag operator on y t which is like L to the power 1 by t which is nothing but y t minus 1. So, in lag notation this becomes L y t or we can take make it into the polynomial form if you recall 1 minus L y t equal to mu plus mu t.

Now, this is a stationary process. As you can see here, this is a stationary process that is delta y t is a stationary process. So, the first differencing has introduced non-stationarity and therefore, the process is called integrated to our order of 1. Why? Because we only needed to take first difference once. So, we only did the first differencing once.

$$\Delta y_t = y_t - y_{t-1} = y_t - Ly_t = (1 - L)y_t = \mu + \mu_t$$

So, this is called integrated of order 1. If you remember we said the process is ARIMA where I is this order of integration, I represents this order of integration. So, this is how you deal with this non-stationarity. You account for the non-stationarity by taking first difference. In this case, it is a integrated of order 1.

So, you have to take only first difference. It could have been second order non-stationarity, then you would have to take the first different two times and in that case this I value will change from 1 to 2. So, depending upon the order of integration, the I value changes and this is how you account for that non-stationarity. Also, if you recall this 1 minus L can be written in a polynomial form like 1 minus Z and this is the characteristic equation of the process. And it has a root at Z equal to 1 and that is why this is also called a unit root process or unit root non-stationarity.

So, the process, the characteristic equation of the process has a root at Z equal to 1 which also indicates non-stationarity and that is why this is also called unit root process because of the existence of this root which is equal to 1 or in technical terms we say that root lies on the unit circle, the root of the process lies on the unit circle. So, let us summarize what we discussed in this stochastic non-stationarity. We said that first differencing introduced stationarity in a process that was integrated of order 1 or I1. So, if a non-stationary series requires differencing d times, differencing of d times, then it is said to be order of integrated at d. So, it is of order d and it can also be written as Yt tilde Id that means it is an integrated process of order d.

So, the integration is order of d and it requires differencing d time that means delta d, this delta d represents d time differencing will when it is applied on Yt it will return an I0 process. That is if d time, if the process Yt is differenced d times it will result in an I0 process that is a process with stationary, it is a stationary process with no unit roots when difference d time if the process was integrated of order d. For example, if it is integrated of order 3, 3 times differencing that means delta 3, 3 times differencing will lead to a I0 process. What is the key difference between these processes that are non-stationary and a stationary process such as I0? So, non-stationary process such as I1 and I2 can wander very long away for very long times they can wander from their mean and if there is a drift then that wandering property will be even higher. In contrast, an I0 process will fluctuate around the mean very swiftly.

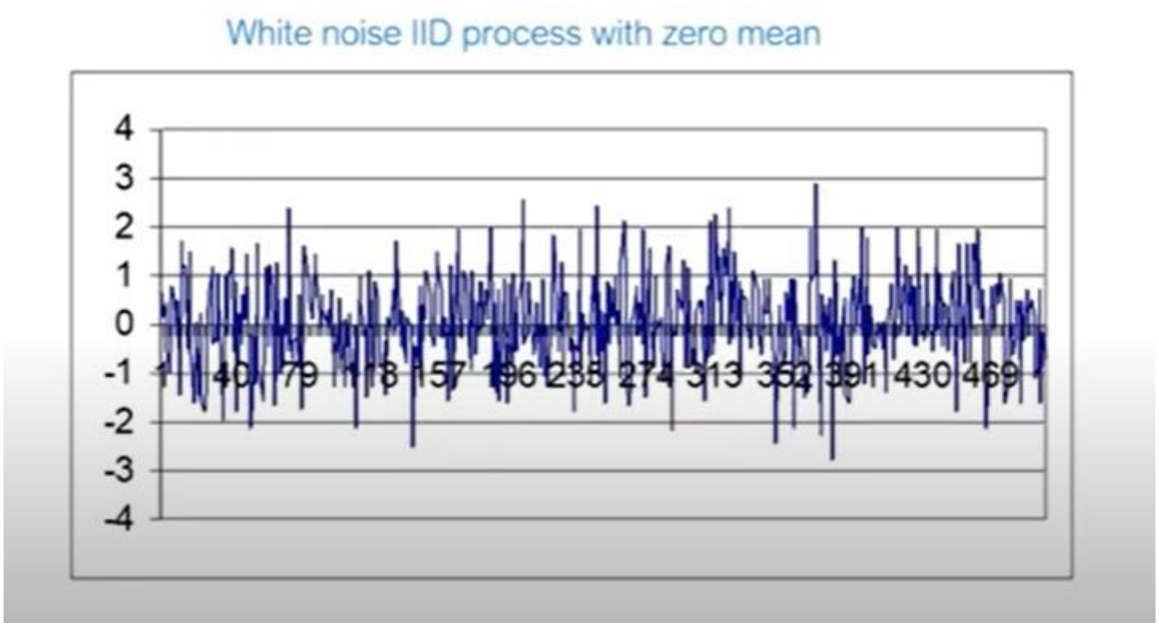
So, it will all very frequently cross the mean some if there is a if there exists a long term mean the I0 process will behave like this it will cross the mean very swiftly but if it is I1 or I2 process it will deviate and drift a lot from the mean and even if there is a drift further drift that new component is there then this deviations these deviations will be very large in magnitude and it will remain and so away from mean for very long times.

In this video, we will discuss deterministic non-stationarity and how to tackle it. Recall, we said that deterministic trend process appears like this where Yt is a function of time trend t. Now, this is rather simple case of non-stationarity because to deal with it only what is required is detrending what is called detrending. So, all you need to do is in the process Yt you model it with the time trend t like this and once you model out the time trend d here you will find the coefficient beta and constant term alpha the residuals from the process mu t will not have time t.

$$y_t = \alpha + \beta_t + \mu_t$$

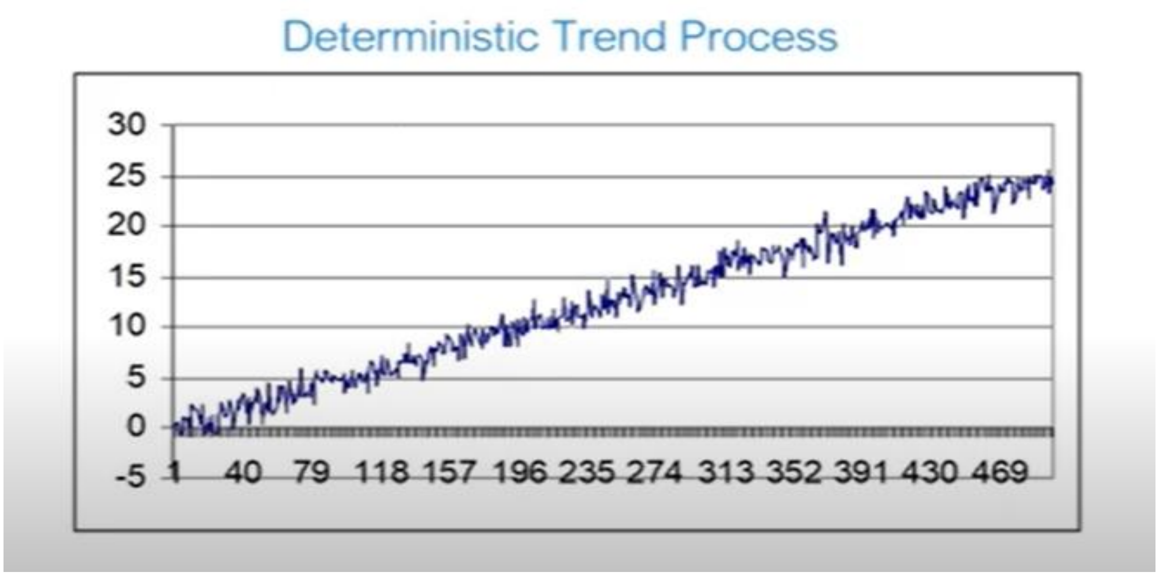
So, all the subsequent modeling we can do on mu t by removing the time trend through this process and subsequent estimation can be performed on these error terms for example testing for their white noise character or any such test would be done on mu t. So, to summarize in the deterministic non-stationarity all we need to do is model out the time trend and then work on the residuals. Please remember we have studied two kinds of non-stationarity stochastic and deterministic. If you perform the treatment which is not or opposite for example if you perform the treatment of deterministic non-stationarity on stochastic non-stationarity for example if it is stochastic non-stationarity and you add time trend or vice versa on deterministic non-stationarity you take first difference then the resulting process or the resulting model will not be error free it will be inefficient and most probably the non-stationarity will not be removed. So, it is better to first understand what kind of non-stationarity is causing the stochastic or deterministic trend and then accordingly treat this process in a different manner.

In this video we will conduct the time series and stationarity properties try to distinguish between different process through visual examination. Let us first look at this white noise process with a zero mean. Notice how frequently this process fluctuates around its mean. It quickly comes back and forth around this mean. So, this characterizes the white noise process with a zero mean.

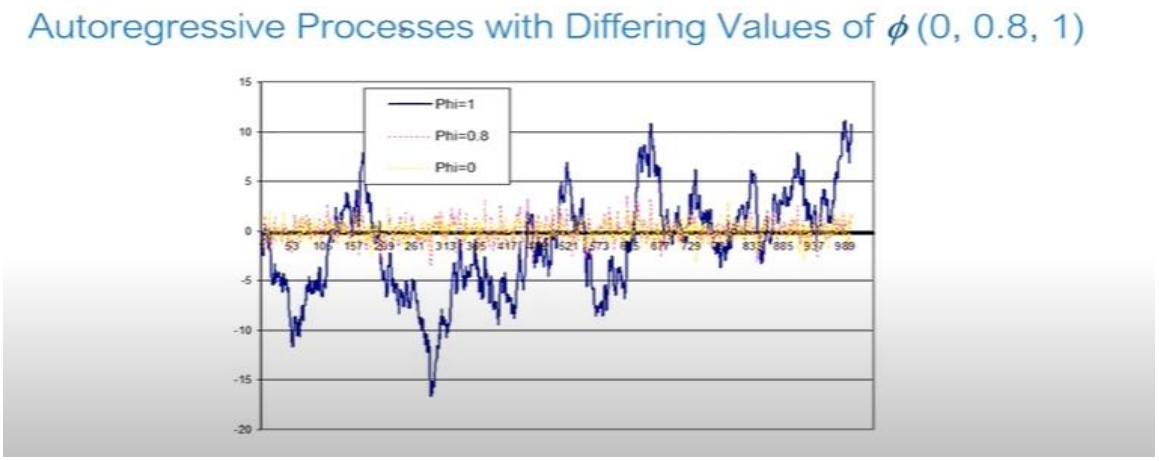


Next let us look at a random walk a process which has a unit root and the random walk process with a drift. Notice the process with the random walk this random walk process whenever it deviates from its mean it deviates much further away for longer times as compared to the white noise process. Similarly, when we add a positive drift to it the purple process it deviates on the positive side much frequently you can think of a process with negative drift also going in a very similar fashion on the opposite side. So, a random walk process with a positive drift is expected to deviate much further from its mean on the positive side while a negative drift will go much ahead on negative side.

Next let us look at a deterministic trend process. Now while this process fluctuates a lot like a white noise process but it fluctuates around that deterministic trend. So, it fluctuates very frequently around that deterministic trend. If we detrend the process if we detrend this process it will the residuals from the process will fluctuate like this because once the trend has been removed it becomes very close to a white noise process. Lastly let us examine three kinds of autoregressive process one with zero autocorrelation which is closer to what we call as white noise one with a higher 0.8 autocorrelation coefficient and another process with the autocorrelation coefficient of 1.



So, these are first order autocorrelation coefficient. Let us see how what kind of processes result from these correlation. Notice the process with autocorrelation coefficient of 1 which is of course a sort of unit root process or random walk process. The process deviates a lot from its mean and that too much further away. Any shocks to the system would persist much longer in this kind of process while if we compare it with the 0.8



which is this dotted pink line that also deviates though relatively less as compared to one but it still deviates as compared to a process with phi equal to zero where the deviations are much less and they are fluctuating around the mean which is almost as if the white noise process.

Now to summarize this video we studied that and we saw visually that a white noise process has no trending behavior and it frequently crosses its mean value which is generally zero. We also saw the random walk process and a random walk with drift process and they exhibit long swings around their mean value and they cross their mean value very rarely as compared to the white noise process. We also saw when we compared these two random walk and random walk with drift we found that the positive drift if in the process there was a positive drift the series was more likely to rise over time and deviate much further as compared to a simple random walk and this effect of this positive drift became greater and greater as the time passed. Similarly, we can think of negative drift as taking process on the negative side of the mean on the lower side of the mean.

Lastly, we also recall that we saw visually the deterministic trend process and we noted that it does not have a constant mean but those fluctuations they were around the upward trend. So there was the mean was sort of increasing secularly upward and those fluctuations in the deterministic trend process were around that upward trend and if this trend were to be removed from the series a plot very similar to white noise process would be obtained.

In a series of next two videos we will try to understand test of non-stationarity or what we often call as unit root test. To begin with let us understand with the help of a simple Dickey Fuller test of non-stationarity. We start with the simple autoregressive model which is Yt equal to phi times Yt minus 1 plus mu t which is a first order autoregressive model.

$$y_t = \phi y_{t-1} + \mu_t$$

The test of non-stationarity here is a simple objective simple basic objective to test whether the null hypothesis that phi here is equal to 1. We are not testing phi greater than 1 because those cases will result in explosive process which are generally not found in economics and finance. So by default the alternative is phi equal to 1 and the only one sided alternative becomes phi less than 1 because we are ignoring phi greater than 1. So we formulate our null hypothesis that series contains a unit root which means phi equal to 1 and alternative the only possible one sided alternative is that H1 is series is stationary. Now because the formulation of this kind of equation there is a possibility that phi equal to 1 and estimation of then this kind of model will not be feasible.

So we transform this model a little bit by subtracting both sides from Yt minus 1 both LHS and RHS. And then this Yt minus Yt minus 1 will become delta Yt because of first differencing and the right side will become psi times Yt minus 1 plus mu t where psi is nothing but equal to 1 minus phi because of that subtraction with Yt minus 1. So the resulting equation that is delta Yt which is the first difference of Yt equal to psi times Yt minus 1 plus mu t. So an equivalent test of phi equal to 1 which is the null that process stationary is equivalent to test of psi equal to 0 because psi is nothing but phi minus 1 here because we are ultimately subtracting Y Yt minus 1 minus Yt minus 1 which results phi minus 1 as coefficient. The question is why not use this original regression because original regression obviously by the null hypothesis of being non-stationarity and if you are going with the null and estimating this kind of model with phi being 1 that is not econometrically efficient and all those problems of spurious regression will follow.

$$\Delta y_t = \psi y_{t-1} + \mu_t$$

Now generally in this process we assume in financial market returns there is no drift or deterministic trend but let us say you believe you want the model to be very generalized kind of model to test for non-stationarity that is by allowing a drift or intercept along with the deterministic trend then you can consider a model like this where in addition to this phi into Yt minus 1 you also introduce mu which is the drift it can be positive or negative drift and a deterministic trend for t. So this will take care of your deterministic non-stationarity as well. Now again we repeat the same step we subtract by Yt minus 1 so we get delta Yt equal to psi times Yt minus 1 plus mu plus lambda t plus mu t. So this is a more generic expression where delta Yt is nothing but Yt minus Yt minus 1 and then we test this model for non-stationarity. So a test that is null non-stationarity which means phi equal to 1 and alternate if null of non-stationarity is rejected then process stationary with phi less than 1.

So equivalent model here psi then psi you test null with psi equal to 0 and alternate becomes that psi is less than 0 where psi is phi minus 1. So this again is a test of random walk against a stationary AR1 with drift and time trend. Why drift and time trend now because we have included these two terms. Now unlike the original model where we would have tested in a model like Yt equal to psi times Yt minus 1 plus error term we would have tested phi for its significance with phi the t value as phi upon standard error of phi because these are estimates we put a hat symbol but now because we have transformed the model and the resulting coefficient psi is tested and these are called tau test. Why these are called tau test because these are not your regular regressions there is a possible non-normality in the process there is a possibility of that.

So these the corresponding test statistic for psi is not the regular t statistic it is different we call it tau test and the value of tau its notation is similar which is psi hat estimated value of psi upon standard error of psi but the distribution is not similar to t. So the distribution is very different and it is provided through simulations and it has been found that it is different from t in fact this tau is slightly higher in values as compared to our conventional t value that means it requires more evidence if you study more about the t distribution and evidence in models you will find that t tau being higher reflects more evidence is required to reject the null. So these tau are rather non-standard unlike the conventional standard t distribution and like I said because we are rejecting the null when psi is less than 0 which means psi minus 1 is less than 0 or psi less than 1 in other words the tau values corresponding to this alternate hypothesis are larger negative in terms of magnitude they are larger than the regular t-strat that means they are larger negative because we are looking at negative values to reject the null and essentially in economics and finance sense these tau larger tau values indicate that we are requesting for more evidence while rejecting the null of non-stationarity.

To summarize this video here we saw the construction of simple Dickey Fuller or DF test of non-stationarity. In the next set of videos we will augment this test and study augmented Dickey Fuller test and few other tests of non-stationarity and how to use them in inferring whether a process is stationary or not.

In this video we will discuss augmented Dickey Fuller test of non-stationarity which is useful when the errors are serially auto correlated or there is some kind of serial auto correlation in the data. Let us go back to our simple Dickey Fuller model which appeared like this delta y2 in a simple form psi times yt minus 1 then you may have a drift and a trend plus error term. In this model the idea was that the null of unit root or non-stationarity will be rejected in favor of a stationary process or stationary alternative which means size less than 0. If this test is started tau what we call this tau is more negative is more negative as compared to our critical values or tau critical. So in terms of magnitude if tau observed or tau from sample is greater than tau critical then we would have rejected the null of non-stationarity and said the process is stationary.

However for this kind of model to work efficiently it is required that mu t here this mu t error term is a white noise process. White noise IID process but if there is some kind of auto correlation structure in delta yt process which is the dependent variable here then and if that is not well taken care of then mu t will also be auto correlated and if the mu t is auto correlated then the inferences and in fact are testing of this the significance of this psi variable would be inefficient and incorrect. The solution here is to augment this model with lags of delta yt that is delta yt minus 1, delta yt minus 2 and so on certain number of let us say p lags of this dependent variable delta yt which will soak up this auto correlation existing in the delta yt process. Let us discuss this in more detail. So the idea here is to augment the model with p additional lags notice this so we are introducing these 1 to p additional lags in the original model.

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \mu_t$$

You can here we are not showing the drift and time trend you can add them here if needed but more importantly we are adding the lags of this dependent variable. The idea is that these lags will soak up additional auto correlation in this delta yt time series process and this kind of model which is augmented with the lags of the dependent variable is called augmented Dickey Fuller or ADF test and these lags will help us absorb any serial correlation so that our yt will be our mu t the error terms will be more cleaner and closer to white noise ID process. Now there is a problem because we are introducing more lags the number of lags should be very clear how many lags we are going to choose that should be an optimum number. There are two ways to choose it either through the information criteria such as AIC, BIC criteria that we discussed in the previous videos or through practice knowing what is the information structure in the data. For example let's say we are looking at data or financial markets process where five week trading is going on then a lag of four to five days is sufficient to reflect or capture all the information assimilation process in the market.

So it depends on your idea of the market and various practical aspects to choose the lags selection why this is important for example if you are choosing two less very small number of lags then the entire serial correlation that you want to absorb in this delta yt process will not be absorbed and still mu t will be autocorrelated. If you are using too many lags it may absorb the autocorrelation very large number of lags but again it will require more parameters more alpha is to be estimated and this kind of unparsimonious or non-parsimonious model will introduce noise in the model and it soaks up lot of degrees of freedom that is number of observations or sort of leverage available to you the number of observations to estimate the model that is called degrees of freedom and that will be absorbed by these new parameters and essentially it will result in higher standard errors of estimates and a poor model estimation. So there is a trade-off whether you are choosing two less number of lags or too many so you need to choose optimum number of lags and then you run your augmented Dickey Fuller test of non-stationarity. Again similar to augmented Dickey Fuller test there are other tests such as Philips Perron and KPSS test though mathematics is not important but in subsequent videos we will discuss how to use all these tests in a more efficient way to arrive at accurate conclusion.

In this video we will discuss the approach to test for non-stationarity with the help of multiple test of stationarity and unit rule test. Recall previously we have discussed augmented Dickey Fuller test which was essentially a unit root test with its null hypothesis H0 of being non-stationarity while alternate if the null is rejected is of stationarity. Another test Philips Perron also has a similar null of non-stationarity and while rejection of null that is alternate hypothesis leads to stationarity of the process. In contrast there are tests of stationarity like KPSS test where null hypothesis is of process being stationary while rejection of null would suggest process being non-stationarity. Therefore if you put such test together and look for the outcome then ADFPP test would give you outcome such as rejection of H0 and KPSS do not reject H0 which are aligned to each other. So you have two cross two outcomes rejection and non-rejection four such outcomes with combinations of ADFPP giving you rejection, no rejection, rejection, no rejection.

So, you have to look at the KPSS giving you no rejection, rejection and so on. Now how to go about it? The approach here is called confirmatory analysis where you look for evidence which is consistent with each other. For example rejection of H0 and not rejection of H0 in KPSS this both these outcomes are consistent with each other. Similarly do not reject H0 for ADFPP and reject H0 are sort of consistent with each other because they both tell you the same thing to take process as non-stationary while the upper one will take process as stationary. However if you look out outcome three and four reject H0 and reject H0 there H0 being opposite this is a contradictory evidence. Similarly if you look at this do not reject H0, do not reject H0 this is also contradictory evidence.

So, in case your results appear to be from one or two you have a very solid and robust evidence where all the tests are converging with each other. So to summarize this video we discussed the approach test for non-stationarity with using multiple tests which include two classes of test. One class which is called unit root test where the null hypothesis is non-stationary and alternate is stationary and another set of test called test of stationarity where null is stationary and alternate is non-stationary.

In this video we will introduce a very fundamental and important concept called cointegration and mean reversion in time series properties. To begin with if two series let us say yt and xt have a property of coming together towards each other and even if in the short term if there are deviations but they come together towards each other that means if you take some combination of yt and xt and because of this fundamental property of staying with each other the combination of these two series has some kind of long-run mean value and like I said this long-term relationship or mean is driven by some fundamental factor.

Let me give you two examples. Think of a cash that is equity and futures prices of a same underlying let's say a stock of company ABC. So ABC stock and future obviously they will not deviate too much away. In short term they may move a little bit away but in long term they should come together. So if I were to take the difference of ABC's cash equity and futures the difference would remain somewhere around its long-term value even though on per se ABC equity price may move much up above and same about future it may also move much above or down but if you look at the difference of these values it will remain closer to some long-term average. Why? Because if the difference deviates arbitrage what we call as forces of market efficiency or arbitrages they will conduct buying and selling activity which will align these prices together because they are driven by the same underlying fundamental security and therefore the combination of these two series yt and xt these two time series will be stationary even though on their own they may not be stationary.

For example often it has been found that prices are I1. Prices are sort of I1 process that is integrated of order 1 and that is they are non-stationary on their own. So you need to take first differences like compute returns. Returns is also form of first differencing. So you have to compute returns or take first differences to make them stationary. However when we look at prices of cash equity of the same underlying ABC stock and futures and take the difference of cash and future the resulting series would be stationary in itself and this stationarity property results from the fact that these two series the cash and future are co-integrated.

Similarly think of ask and bid prices of a stock XYZ. Now this gap is called spread and it cannot increase infinitely why because there are forces of arbitrage for example if ask price go much above as compared to bid then arbitrage and similarly bid price may go too low then arbitrage will take direction and therefore this gap they will start buying or selling depending upon if it is overvalued or undervalued they start buying or selling and therefore if the gap in ask and bid is too high it will be driven back to it some kind of long term mean by these arbitrage so it cannot increase infinitely and this difference between ask minus bid will remain somewhere in a close range even though on their own ask and bid can increase a lot and come down as well but their difference the gap between them will not increase too much it may deviate in short term but in long term it will remain at some kind of stationary level and this is called cointegration between ask and bid which is driven by some kind of fundamental relationship between these two time series particularly in this case because they are ask and bid prices of the same security they will be aligned to each other that means their combination which in this case is the difference between ask and bid combination can be in any form a linear combination can be in any form in this particular case it will be the difference which is stationary.

Let us look at the mathematical notations and understand the concept of cointegration in more detail. Consider two or series of more variable that are I1 in nature their linear combination or rather what we call as error term for example if one series is y t and other series x t and you express their linear combination as y t plus alpha t y t as into alpha x t equal to mu t so mu t is a linear combination of these you can think of it as a error term or residual or some linear combination of these there exists some linear combination for a given value of alpha this mu t will also be I1.

$$y_1 = \alpha_0 + \alpha_1 * x_1 + \alpha_2 * x_2 + \mu_t$$

$$y_1 - \alpha_0 - \alpha_1 * x_1 - \alpha_2 * x_2 = \mu_t$$

So the idea here is that if you combine two or more time series variables with different orders of integration I d if you combine them the combination term that is mu t will be integrator of order d where integration of order d is the maximum integration so for example if you have n variables like y1 y2 and so on up till y1 so these are individually time series variables on their own different variables and if you combine them and let us say the highest order they some of them are integrator of order 1 some of them are integrator of order 2 and one of them the highest order of integration is order of d that is maximum integration which is our word order d then the resulting error term is expected to be of integrated of the largest in order that is d so the mu t will be integrated of order d which is the largest order of integration across all these variables for example let us take variable y1 x1 x2 and think of a model like this y1 equal to alpha naught alpha 1 x1 plus alpha 2 x2 plus this error term I can express mu t alternatively as a combination of these three variables y1 x1 x2 then whatever is the order of integration y1 x1 x2 mu t's order of integration will be same as the maximum order of integration across these three process for example if y1 is order of integrated 1 x1 is of 2 and x2 is of 3 then in general mu t is expected to be of integrated of order 3 similarly if all of them y1 x1 x2 are integrated of order 1 then chances are that error will be of order 1 however it is desirable to have an error term that is order of 0 and in case in that model that we express that y1 plus alpha 1 x1 plus alpha 2 x2 is mu t if mu t is found to be I0 if you find that mu t is I0 then this is called cointegration that means y1 x1 and x2 are cointegrated with each other what could be those economic and financial scenarios and influences that will lead to I0 and that is called cointegration so if a set of variables are cointegrated only and only you can find a linear combination like this there can be more such variables where you can find a linear combination resulting in the mu t which is integration of order 0 integrated of order 0 that means it is stationary and this is called stationarity this will result in a property called cointegration across all these variables in finance and economics many times series variables such as spot and futures rate exchange rates and like I said bid ask spread which is the gap in ask and bid or rather a linear combination of ask and bid is found to be stationary in finance and economics the idea is that such series where there is some fundamental economic and financial relationship in short term temporary disturbances in short term temporary disturbances can make them go apart but because of these fundamental influences that are providing that structural relationship in the long run these fundamental structural forces or the fundamental relationship will bring them together for example ask minus bid it cannot increase infinitely it has to be range bound similarly cash equity minus future it cannot the gap cannot increase it finitely it has to be arranged bound and this is this kind of long run long term equilibrium phenomena is called cointegration which reflects some kind of fundamental structural relationship between time series variables there can be two three or more variables and if we found their linear combination which we are calling as mu t or error term to be stationary or I0 then these variables would be called as cointegrated.

So to summarize in this video we understood the concept of cointegration we said that if because of some fundamental relationship a structural relationship if a set of time series variables irrespective of the fact that there are order of integration is more than zero order of integration can be one two three but the linear combination results in a error term which has a order of integration I0 which is stationary then these variables are called cointegrated variables and this cointegration results from some kind of fundamental economic and finance relationship.

In this video we'll introduce the error correction models or also referred to as equilibrium correction models. Typically when you are dealing with univariate series individually then a good approach to deal with non-stationary or integrated series is to convert them to I0 for example if it is I1 you take first differences and convert them to I0 generally when we are talking about financial economic series we are talking about stochastic non-stationarity. So in this manner when we construct I0 series often when you are looking at some kind of time series models of a multivariate nature then also you can follow this approach while converting individual series at I0 level and then try to find this relationship. But there is a problem as we will see in the next discussion this kind of approach where we convert a series such as I1 to I0 and then estimate the relationship between let's say a series like yt, xt or so on then we lose the long run relationship between these variables.

This we will discuss shortly but why are we discussing only I1 why not I2, I3 and so on. Generally in finance and economics we do not have series like I1 for example prices would be generally I1 or similarly other series such as GDP and other things would be I1 you convert them into I0 by taking a log return kind of form. So you when you convert them into first difference or return form they become I0. Now let's say there are two series yt and xt they are I1 of nature and you try to estimate a relationship which is changes in yt vis-a-vis changes in xt and try to estimate this relationship for example you may have cash or future data and you convert them into their first differences or return forms. The problem is that this is sort of short term relationship. Why? Generally when you're talking about long term relationship between two series yt and xt in the long run these variables they approach some kind of steady state equilibrium.

What do we mean by a steady state equilibrium? A state where yt is not changing much because it has reached equilibrium value the difference between yt and yt minus one which is Delta y is approaching zero that means there is some y steady state value which is approached in long run and therefore time by time there are no changes so Delta yt is approaching zero and same for xt as xt approaches its equilibrium level Delta xt also approaches zero so they are approaching some kind of steady state long term steady state value and therefore Delta yt and Delta xt are not changing becoming close to zero so if you want to estimate this kind of relationship that becomes meaningless because then it has no solution because this Delta yt and Delta xt not only they are not changing they are close to zero so there is no solution to such problem however there is one solution to this problem where you can get not only the short term but long term solution also only and only if the variables are co-integrated. Let us see how.The idea here is that if two series such as yt and xt are co-integrated then we can combine their lag values the lag values and their first differences to find this kind of model which is a combination of this first component short term and also their lag values which essentially captures their long term relationship so if they are co-integrated we can say that there exists a combination like this which is a long term relationship I repeat so if we assume that yt minus one and xt minus one are co-integrated there exists a long term relationship expressed in the form of this which is essentially nothing but the error term et which reflects the deviation this et reflects deviations in their equilibrium levels so this component is put here and the short term relationship is captured here so essentially what we are trying to determine through this relationship is what factors drive changes in yt and these are not only changes in xt but also deviations from long term equilibrium relationship captured by this et or this linear combination and assuming that there exists some linear combination because this yt and xt are co-integrated this et would not be of order I1 or higher if even if yt and xt are integrated but it will be I0 because of the co-integrated relationship and this is what we call as error correction model ECM or vector correction model ECM or often referred to as equilibrium correction model in this model as we discussed this yt minus one minus gamma times xt minus one is the error correction term et which captures the deviations from the long term relationship so this term gives you the deviations from long term relationship and this is supposed to be I0 because of the co-integration even though yt xt are I1 or maybe higher also also for models modeling perspective you can also have a constant term like alpha here you can in this relationship also you can have some kind of constant if required and because this entire relationship is purely stationary why because this relationship is I0 and this is anyway I0 therefore it is perfectly valid to use OLS procedure here so in this relationship where we are saying that changes in yt are a function of changes in xt plus some kind of let's call this term coefficient beta 1 and another coefficient beta 2 which captures the relationship between xt plus some kind of error term let us interpret this one by one first and foremost here in the short term error correction term et this is appearing with the subscript t minus one because it indicates deviations from equilibrium in the previous period so the current changes in delta yt which come with the subscript t are affected by the previous deviations from this long term equilibrium relationship so there we need a substrate t minus one that means changes at time t are driven by equilibrium deviations at t minus one that is logical second you have a gamma coefficient here which captures sort of long relationship between yt and xt minus one obviously depending upon sign you can get the impression about relationship but the magnitude of this coefficient gives you the long term relationship between x and y next you have beta 1 as a coefficient of delta xt because this is a coefficient of changes in xt it sort of captures the short term relationship between delta yt and xt so because we are using first differences here again the assumption is that this yt and xt are I1 generally price series and other economic and financial series are I1 in nature and their first differences are I0 generally that is conventionally that happens next this beta 2 which is a multiple of this error term or this long term relationship deviations from the long term relationship is called as speed of adjustment because it is the magnitude of this term essentially reflects how much error in the previous period which was ET or deviation from the equilibrium was corrected in the current period so it sort of gives you the correction in error or adjustment towards the equilibrium level so this beta 2 measures that adjustment towards the equilibrium or that speed of adjustment towards equilibrium now just to summarize this discussion here we use for two series this can be extended for a multiple series and where we can use a larger cointegrating vector with multiple variables such as you can have a vector like yt plus alpha 1 x1t plus alpha 2 x2t and so on you this is called cointegrating vector where instead of only having two variables you can have more variables so if you have a relationship like this yt plus alpha 1 t XP and so on plus alpha 2 x 2t and so on the cointegrating vector can be represented as 1 plus alpha 1 t plus alpha 2 t so it's just a vector of coefficients this is called cointegrating vector

To summarize this video we said that often the approach to deal with non-stationary relationships or sort of multivariate non-stationary relationships is to take changes differences in the variables or series and take them to i0 level and once you have those variables at i0 level then try to find some relationship if that exists however the problem with this kind of approach is that it only captures short relationship and if there is exists some short long relationship and the variables are equilibrium such approach may lead to problematic results so only and only if a system of variables is cointegrated then only there exists a modeling approach called error correction models where we use a combination of first differences or i0 at change level first differences or change levels i0 plus the long-term relationship which is cointegrated which is in the form of lag variables and combine them to get a class of family of models called error correction models this kind of approach helps not only estimate the long-term relationship short-term relationship but also adjustment towards equilibrium as time progresses however this approach can be only applied if the system of variables or the system of linear combination of these variables is cointegrated that is driven by some kind of fundamental long-term relationship

In this video we'll discuss the Angel Granger approach to cointegration and error correction models.

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + \mu_t$$

So what is the intuition here behind the Angel Granger approach recall if a set of variables such as yt x2 t x3 t and so on up till xkt are cointegrated so you have k variables that are cointegrated so a system of variables with cointegrated values this model represents a sort of equilibrium where mu t can be i0 it can be i0 only if these set of k variables are cointegrated and therefore we can find the linear combination of these variables which is mu t which can be i0 and then all you need to do is set up test of stationarity or mu t to confirm this and that is what lies behind the Angel Granger approach as we are going to see the two-step Angel Granger approach to cointegration and error correction model is as follows in the first step we proceed with the variables such as yt x1 t x2 t in their original i1 form assuming these are financially converse variables so they would be i1 nature then in this model let's say model like yt equal to beta 2 x2 t plus beta 3 x3 t and so on up till beta k xkt in this kind of model we estimate the residuals from this model that is mu t at this stage because this relationship not necessarily be cointegrated we cannot say anything about the hypothesis behind beta beta 3 and beta k and their statistical significance so the next step would be to test these mu t the error terms for their stationarity or whether they are i0 or not through test such as ADF test TP test and KPSS test as we discussed earlier in the next step we approach the next step only only if the residuals are stationary if you find that residuals are not stationary indeed they are i1 then you have to only find the short term relationship by estimating the first differences in yt delta yt delta xt and so on so you have to find the first difference in that only short term relationship will exist however if you be if you find that mu t are stationary then you can approach with the VECM or what we call as error correction model which is a combination of these changes in delta yt and a short term relationship is represented by this delta xt along with alpha 1 and mu t which is the error correction term now please note delta xt not be only one variable it can be a system of variable such as delta xt x1t delta x2t and so on multiple such variables it can have and the second term which is mu t represents that long term relationship such as we discussed earlier for example it can be yt minus 1 we are using t minus 1 because it represents values at lags then minus beta 2 x2 t minus 1 minus beta 3 x3 t minus 1 and so on. For example if this is the nature of relationship as we saw then your co-integrating vector would be 1 minus beta 2 minus beta 3 you can generalize this relationship for more number of variables and this captures this mu t so this is your mu t when you again estimate this relationship then in that case alpha 1 captures your short term relationship alpha 2 captures your correction in error it is like speed of adjustment towards equilibrium because it is getting multiplied to mu t and this is your co-integrating vector 1 minus beta 2 minus beta 3 which captures the long term relationship and now this model can be estimated with OLS however all those problems that are conventionally associated with OLS such as autocorrelation, heteroscedasticity and so on are also applicable here as well.

To summarize this video we discussed the Angel-Vengel two-step approach to estimation of error correction models with co-integrating relationships. In the first step we found the relationship in the long term model and in that long term relationship we extracted the residuals and examined them for their stationarity. The ECM kind of model will only exist if these error terms were stationary and i0 and if they were then we would go ahead with the ECM model which is a combination of first differences and lack terms of the time series variables. If the error terms are not stationary then we have to restrict ourselves to only with the short term relationship.

To summarize in this video we will recap the time series stationarity properties. We consider a weak stationary process which has a constant mean, constant variance and a constant auto covariation structure.

If a process is non-stationary any kind of modeling and estimation efforts may result in spurious regressions. There are two kinds of non-stationarities. First the stochastic non-stationarity resulting from autocorrelation coefficient of 1. This kind of non-stationarity is removed by taking first differences. The process is said to be of order D if it requires first differencing D times. In finance and economics most of the time we observe stochastic non-stationarity of order 1. The second is due to a time trend in the process. This kind of non-stationarity is dealt by incorporating a time trend in the process and extracting the residuals. When examined visually a white noise process has no discernible structure and fluctuates around its mean frequently. A random walk process deviates from the mean value and keep wandering for long times. A random walk process with a drift remains further away from its mean for very long horizons. A deterministic trend process fluctuates around the mean that has a secular increasing or decreasing linear trend with time. Autoregressive process with zero autocorrelation coefficient behave as white noise processes. As this autocorrelation coefficient approaches 1 the process comes closer to a random walk process.

The test of stationarity are of two kinds. First the ADF test with a null of non- stationarity and second the stationarity test as KPSS test with null of stationarity. Both types of tests are employed in conjunction and their results are used in a complementary manner to make inferences. Often a system of time series process is integrated of the order which is the maximum order of integration across the all of the components of the system. However in case this system is cointegrated then a linear combination exists which is I0 that is integrated of order 0 and stationary. The cointegration in a system of time series reflect some fundamental structural relationship across these time series such as cash equity and futures or bid-ask prices and so on.

Therefore such system has a long-term mean reversion towards which the system reverts over long term. In the short term the system can deviate from this long term mean. To model such cointegrating relationships we employ error correction models or what we call as VECM. These models employ a system of first difference to capture the short term relation and lack terms in the form of cointegrating vector to capture the long term relationship. If there is no cointegrating relationship then only short term relationship exists.

One of the most employed approach to this cointegrating relationship is two-step Engel-Garengell approach. In the first step we run the long term model to obtain the error term. These errors are tested for stationarity to confirm the presence of a cointegrating relationship. If found stationary then this error term captures the long term cointegrating relationship and is employed along with first differences short term model to construct the ECM or VECM model.

Thank you.