Advanced Financial Instruments for Sustainable Business and Decentralized Markets

Prof. Abhinava Tripathi

Department of Management Sciences

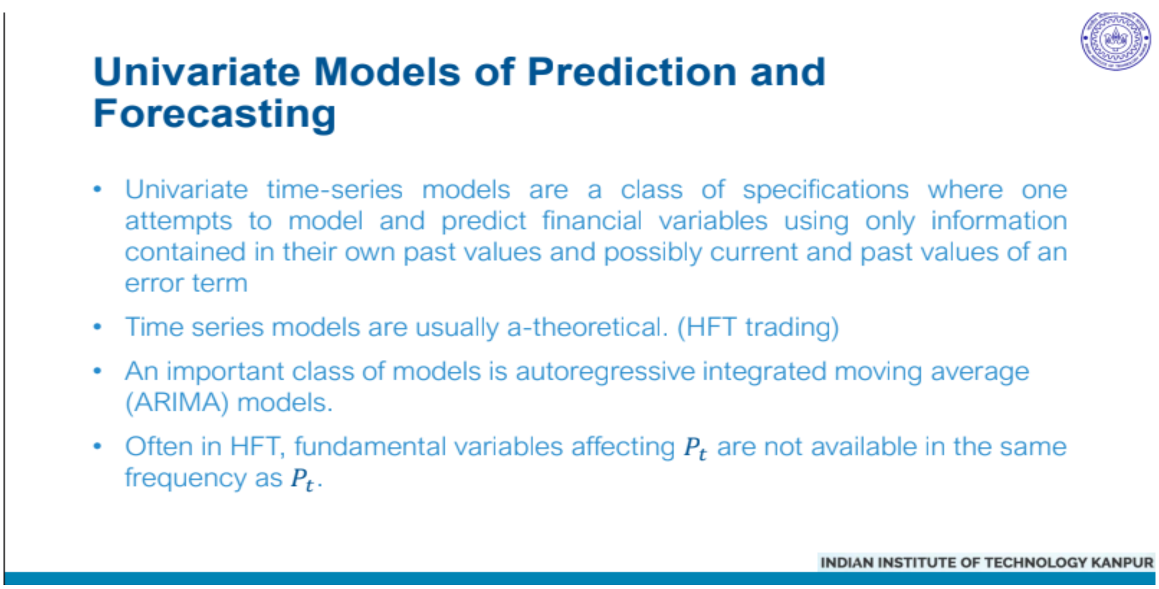
Indian Institute of Technology, Kanpur

Lecture 21

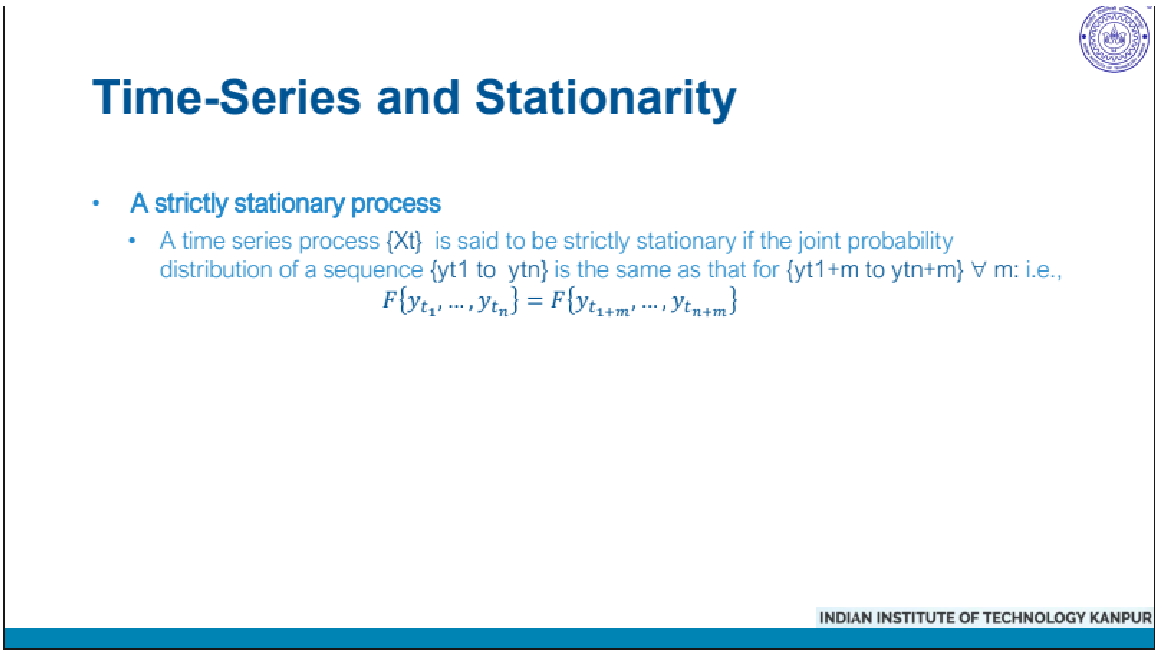
Week 7

In this lesson, we will discuss the univariate models of prediction and forecasting with the help of our Marima class of models. We will start with the concept of time-sealing stationarity and white noise process. Then we introduce autoregressive AR processes and their structure with the help of a simple example. We also discuss its stationarity property. We will also discuss the autocorrelation coefficient and autocorrelation function and partial autocorrelation function in the context of AR processes. Next, we introduce the moving average or MA processes and its structure with the help of a simple example.

We also discuss the ACF and PACF plots for an MA process. In this video, we will discuss univariate models of prediction and forecasting. Univariate time series models of prediction and forecasting are an extremely important class of models. They are specifications where one attempts to model and predict financial variables, in particular for example financial market prices, using only the information contained in their own past values and possibly current and past values of an error term.

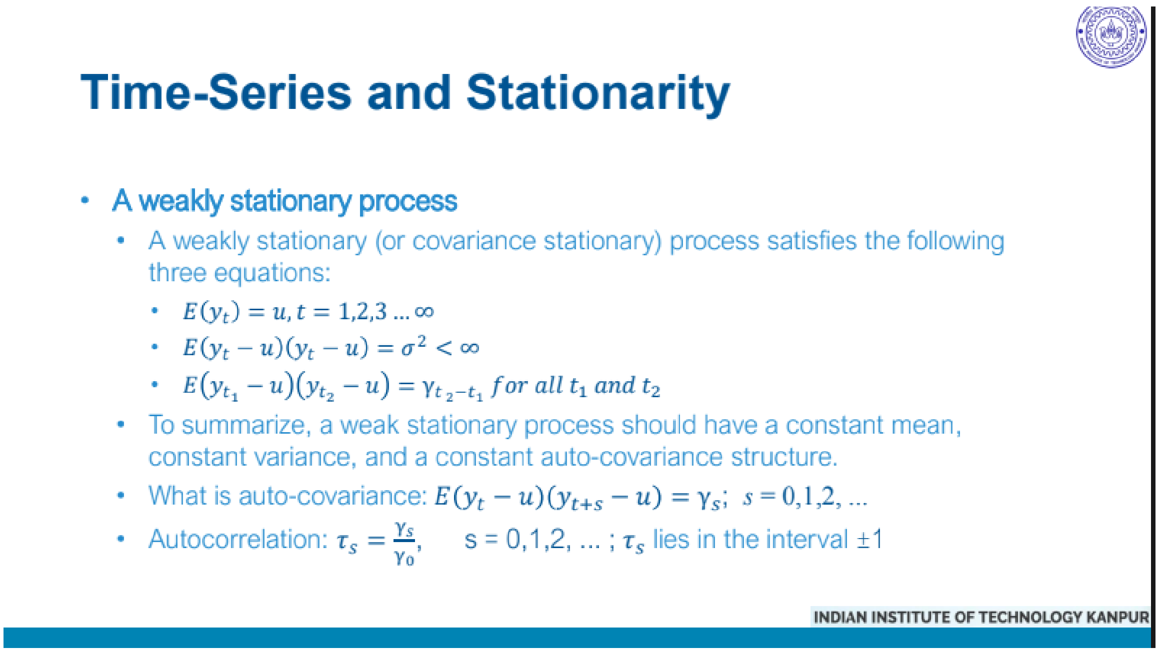


We will have more to say about this error term but generally this is used to mean or denote the information that has just arrived in the market. Unlike the structural models that include some theoretical underpinning and a number of variables, these time series models are at theoretical such as modeling high frequency trading data. What do we mean by this? In general, any such model would derive or take away from some kind of structural model. However, we do not care what that structural model is as long as we have the variable of interest such as security market returns and its own past values. It is assumed that any such structural model that we do not observe or know about at play, automatically that information that is contained in that structural model is incorporated in the historical time series values of the variable of interest.



So for example, you may have some idea that GDP of a country may affect the returns of Nifty-fifty. Let's say Indian GDP affects Nifty-fifty returns. Even though I may not have the knowledge about this model but I know as long as this GDP affects Nifty-fifty and I know that as per the time series model, the historical values of Nifty-fifty will help me predict the future values of Nifty-fifty, I can be rest assured that this relationship, this underlying structural relationship will be captured while I model this time series model of Nifty-fifty with its own previous values in time series model, which is rather more at theoretical nature but by this background, it automatically incorporates the information content of the structural model which incorporates the relationship between GDP and its impact on Nifty-fifty. One of the most important class of such models includes auto-regressive integrated moving average ARIMA class of models. We will have more to discuss about this model.

Why such time series models are important and particularly the univariate models is because often when you are modeling the financial market data with modern financial market instruments that are part of this course. Structural variables are not available at high frequency and most of the modern day trading is happening at high frequency domain. So whenever you are modeling any such model where high frequency trading or any such relationship on daily or weekly basis where more structural variables are not available such as GDP or central bank policy variables, then you have to rely on the historical values of the variable that is available for example prices p_t or returns r_t that are available on higher frequencies and you have to rely on such time series model that employs past values of the variable itself. This is one important motivation of using these time series models. To summarize, in this video, we discussed the importance of univariate time series models.

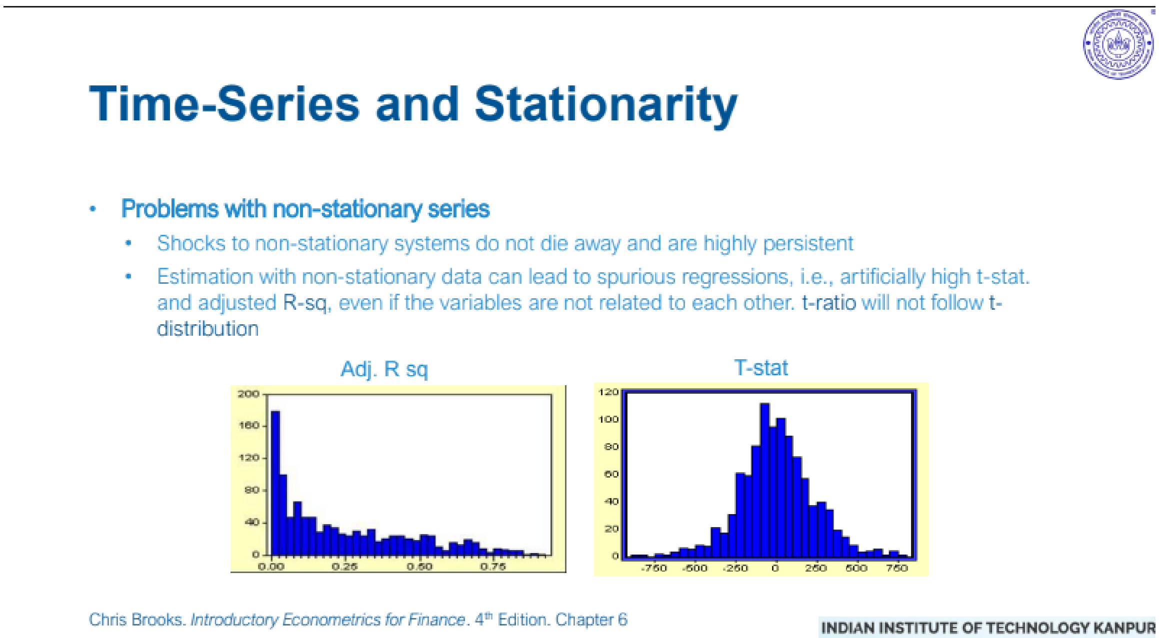


We noted that in modern financial markets, the trading activity and various other dynamics play at extremely high frequencies such as intraday level where various structural model variables such as GDP and monetary policy variables are not available at those frequencies and therefore we have to rely on the historical values of the variable itself such as returns and prices to model the data using models such as ARIMA models. And therefore, in this condition, the univariate models, univariate time series models for prediction and forecasting becomes an extremely important class of variables. And even more so it has been found that while examining the prediction accuracy through methods such as out of sample prediction, these univariate class of models often work even better than the structural models as well. And also the data requirements of these models are extremely parsimonious and they are extremely easy to model and that voids us the intuition and motivation to work with these and learn these univariate time series models. Introduction to time series stationarity.

In this video, we will discuss weak and strict form of stationarity and also the problems that happen when we employ non-stationary series. Let's start with the strict form of stationary process. In a strictly theoretical sense, A time series process $\{x_t\}$ is said to be strictly stationary if the joint probability distribution of a sequence {yt1 to ytn} is the same as that for {yt1+m to ytn+m} ∀ m: i.e.,

$$F\{y_{t_1}, \dots, y_{t_n}\} = F\{y_{t_{1+m}}, \dots, y_{t_{n+m}}\}$$

That means as long as we have the distribution of observations, the different movements starting from first moment that is mean, second moment which is change or second moment which is sort of you can say variance and higher moments are same across as the time progresses across all the observation. For example, whether you are looking at from y_t +1, y_{t1} to y_{tn} or M steps ahead or k steps ahead, the set of observations that you have, its joint distribution for a given set of observations remain same or put it more simply for any observation if I pick for any time Tk, if I pick observation y_k or y_{tk}, the probability that would lie in certain interval will remain constant as time progresses.



This is called stationarity, strictly stationary process which has lot of desirable properties. However, in a practical sense, we make use of something called weak stationary process. What do we mean by weakly stationary or covariance stationary process? We mean the following. This kind of process has the constant mean that means the first moment which is the mean, it is constant. Second moment or variance is also constant sigma square.

So, the variance is also constant. For example, if the time gap is 5 lakhs or 6 lakhs it is what only matters. It does not mean what time you are standing.

A weakly stationary (or covariance stationary) process satisfies the following three equations:

$$E(y_t) = u,\ t = 1,2,3 \dots \infty$$

$$E(y_t - u)(y_t - u) = \sigma^2 < \infty$$

$$E\left(y_{t_1} - u\right)\left(y_{t_2} - u\right) = \gamma_{t_2 - t_1}\ for\ all\ t_1\ and\ t_2$$

So you can compute the autocovariance for a s lag period like this expected value of YT minus mu into YT plus s minus mu. This is for s period difference. So this is for s legs gamma s and this autocovariance structure remains constant as time progresses. While this is an important measure because of scale across different processes is different to compare. So what we focus on is called autocorrelation that means correlation which is a standardized or normalized version of autocovariance.

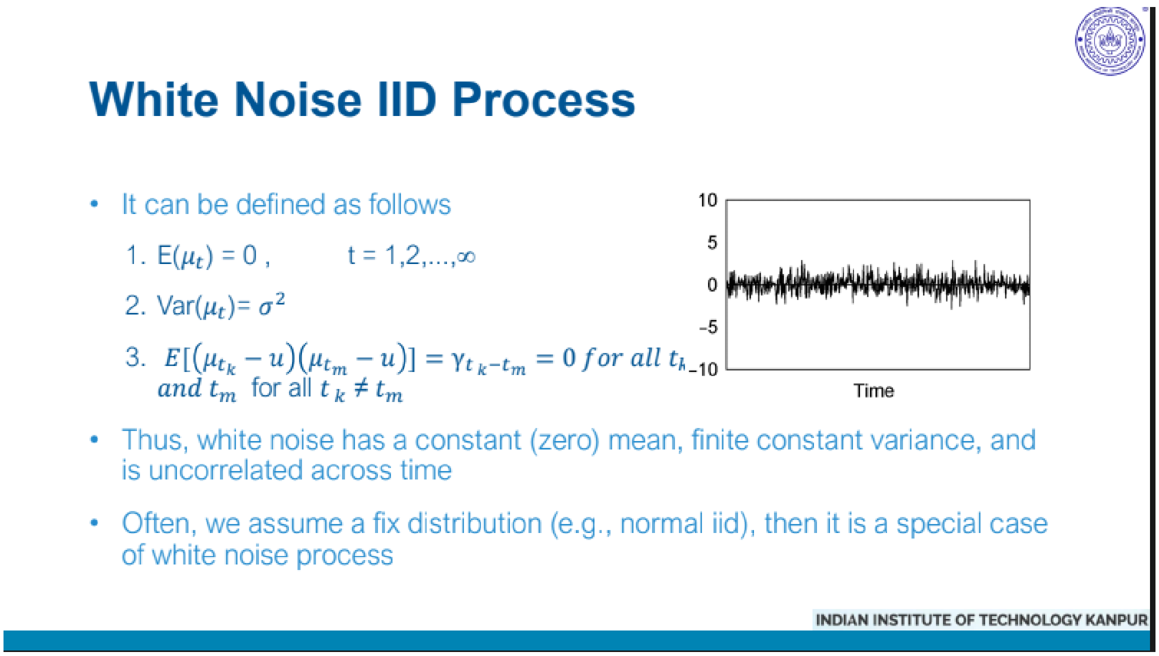
$E(y_t - u)(y_{t+s} - u) = \gamma_s;$ s = 0,1,2, ...

Autocorrelation: $\tau_s = \frac{\gamma_s}{\gamma_0},$ s = 0,1,2, ... ; τ_s lies in the interval ±1

So if you divide the autocovariance at different lags maybe lag s whereas can be 0, 1, 2 with autocovariance at 0 lag that means autocovariance at 0 lag is variance itself that means autocovariance with its own term which is essentially nothing but gamma naught which is variance. This scaled version, this scaled value is called autocorrelation. So at s equal to 0 this value will be 1. In general this value varies from minus 1 to plus 1 and this autocorrelation value it measures from minus 1 to plus 1 and an extremely important measure to define a weakly stationary process the autocorrelation structure of a weakly stationary process. So if you plot this autocorrelation on a graph starting from 0 lag, first lag and so on at 0 lag you get a value of 1 and next whatever autocorrelation structure is there you can plot this at different lags 1, 2 and 3 this kind of diagram is called autocorrelation function autocorrelation function ACF or in short ACF diagram which we will be using very heavily and it is also called correlogram.

So this will be called correlogram or autocorrelation plot or ACF plot. So let us discuss what are the problems of working with non-stationary series. Like we have defined the non-stationary series is a series where different moments do not stay constant with time particularly the first moment which is mean the second moment which is variance and the autocorrelation structure tau s does not remain constant with time. Such non-stationary series has various problems while working with them on econometric estimation. For example shocks to such non-stationary series do not die away.

A very desirable property of any econometric series or econometric series time series is that any shock dies away naturally over as time progresses. However for non-stationary series because of their trending nature and drastically varying different moments the shocks to the series do not die away and they are highly persistent this is very important point. The second important point is that estimation with non-stationary data can lead to spurious regressions. For example even though the two series maybe y1 and x1 they are not very related but still if you run some kind of regression on them they appear to be very spurious resulting in high t statistics and adjusted R square. For example if you sample from two non-stationary series and two different period samples and keep on sampling the process over and over and run regressions you will find a large number of hard adjusted R square like plotted here a very high values for the sample adjusted R square and also very high t statistic values which should not be the case.



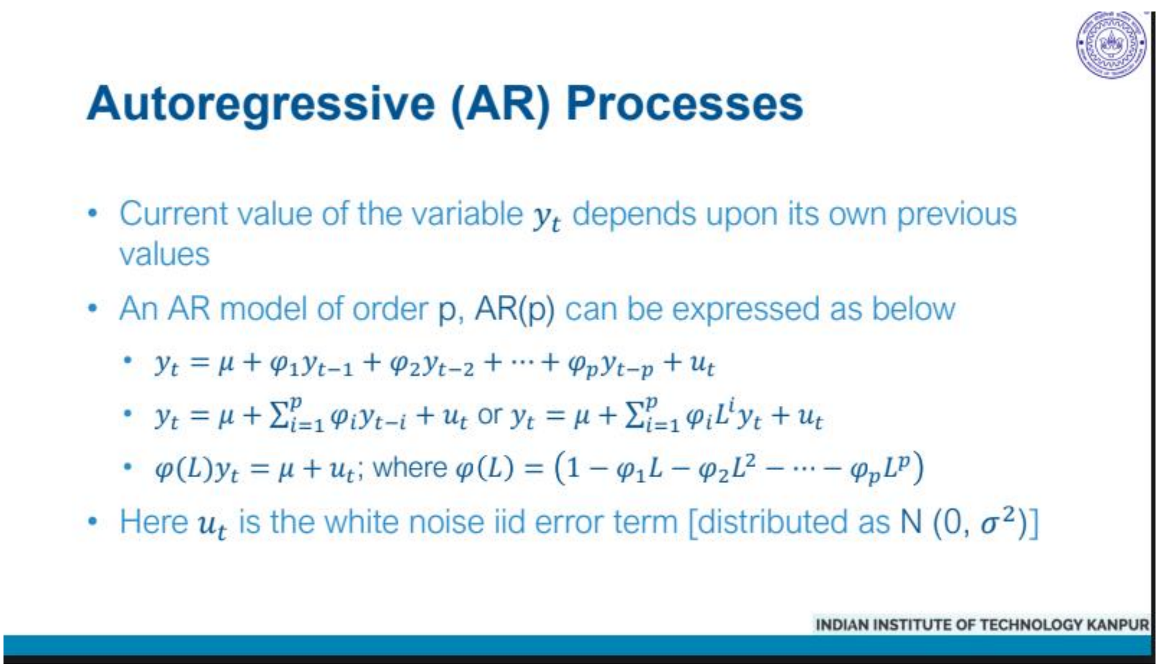
Ideally with two series y1 and x1 are independent any sample and subsequent regression should give you poor adjusted R squares and t-stats but because of this trending nature of non-stationary series for example non-stationary series shocks to the system do not die away and therefore they appear to be moving or trending like this the y_1 and x_1. So, because of this trending nature at whatever point you take a sample y_1 dash x_1 dash or y_2 dash x_2 dash they are always trending and this results in some kind of positive or negative spurious relationship when you estimate some kind of correlation or regression measure on them. To summarize in this video we discussed a strict stationarity process and a weakly stationary process that are that is often employed in a practical sense. We also saw some of the shortcomings and problems with estimation of these non-stationary series with econometric methods such as regression. In this video we will introduce white noise IID or identically independently distributed process.

We will also understand what is the importance of modeling a process as white noise process. To begin with white noise means a process which has no discernible structure.

It can be defined as follows:

1. E (μ_t) = 0, t = 1,2,...,∞
2. Var (μ_t)= σ^2
3. $E[(\mu_{t_k} - u)(\mu_{t_m} - u)] = \gamma_{t_k - t_m} = 0\ for\ all\ t_k$ $and\ t_m$ for all $t_k \neq t_m$

To put in more detail a process with the constant mean generally that mean or expected values taken as 0 but can be a constant number we generally take it to be 0 and a constant variance sigma square which is not changing across the process because there is no discernible structure or there is no reason to expect that this sigma square or variance of the process will change. But most importantly there is no auto correlation or auto covariance structure that means for any two values in the process this t1, t2 they need not be immediate preceding term they can be any two values taken from a time series of value may be t1, t2, t3 or t1, t2 or t1, tk any such two generic values the correlation or covariance between them is 0 for different values of t1, t2 and so on where as long as t1 is not equal to t2. If t1 is equal to t2 then this becomes variance or sigma square and in auto correlation terms this value tau 0 will be equal to 1 as we have already seen.



So, to put or summarize this we are saying a white noise is a process which has a constants or rather 0 mean a finite constant variance and it is not correlated across time that there is no auto correlation. Sometimes you also put some assumption about the distribution and generally this is a normal distribution. So, this becomes a white noise normal IAD process that means when you say independently and identically distributed that identical distribution is normal in most of the cases, but a more generic term is IAD which there one I is for identical another I is independent and these for distribution. So, a normally distributed white noise IAD process is a special case of white noise process. Now just to summarize this what is the importance of modeling a process or when I say process it can be financial market returns and so on as a white noise IAD process.

Many times you model the prices given a certain model it can be various advanced models as we will see like GART kind of model ARIMA model in this case, but once you model it you tend to believe now that you have taken out everything every discernible structure in the process and model it. So, if you have modeled it correctly the residual or error term in the process after you model the series let us say your modeling return series you model out with some kind of ARIMA process a generic term and extract the residuals from the model let us call them mu’s then the assumption here is that if you have done everything correctly these residuals that are affected from the ARIMA process should have no discernible structure because if we have modeled correctly any discernible any kind of pattern that should be modeled will be is modeled through parameters of this ARIMA. So, in the residuals or the remaining part there should be more discernible structure and therefore, it is argued that this mu’s or error term should be as close as possible to a white noise ID process with no discernible structure and that is why it is very important once you model a given process or return or financial market series a time series model then you make an assumption about residual being white noise and then you test the residual for white noise and that testing of white noise residuals is a sort of verification of your modeling accuracy or how good your model has been. So, to summarize we are saying that a white noise process in process with a constant generally 0 mean constant variance and no auto correlation structure no discernible structure. Generally, we use such modeling or modeling for white noise we model residuals from an ARIMA mod class of model or any other similar model where the assumption is that there is no discernible pattern in the structure because we have modeled out anything possible in our model efficiently.

So, generally there should be no discernible structure in our mu’s or put it more specifically there is no auto correlation structure in our mu’s or error terms if there is then there is some problem with our model. In this video we will introduce the AR or autoregressive process which is a very important component of ARIMA. So, this AR in ARIMA stands for autoregressive process. We will also discuss the motivation for modeling a process through autoregressive process. To begin with autoregressive or AR processes assume that a time series process y_t depends on its own previous values. The structure of AR model can be given as follows. Current value of the variable y_t depends upon its own previous values. An AR model of order p, AR(p) can be expressed as below

- $y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + u_t$
- $y_t = \mu + \sum_{i=1}^{p} \varphi_i y_{t-i} + u_t$ or $y_t = \mu + \sum_{i=1}^{p} \varphi_i L^i y_t + u_t$
- $\varphi(L) y_t = \mu + u_t$; where $\varphi(L) = \left(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p\right)$

Here u_t is the white noise iid error term [distributed as N (0, σ^2)]

So, a lag operator with the subscript i would mean Y series with the lag i. So, this term can be further simplified with a lag operator like this. We will see later what the importance is of denoting a process like this. Again, mu's is the error term. This mu's can be written as a white noise.

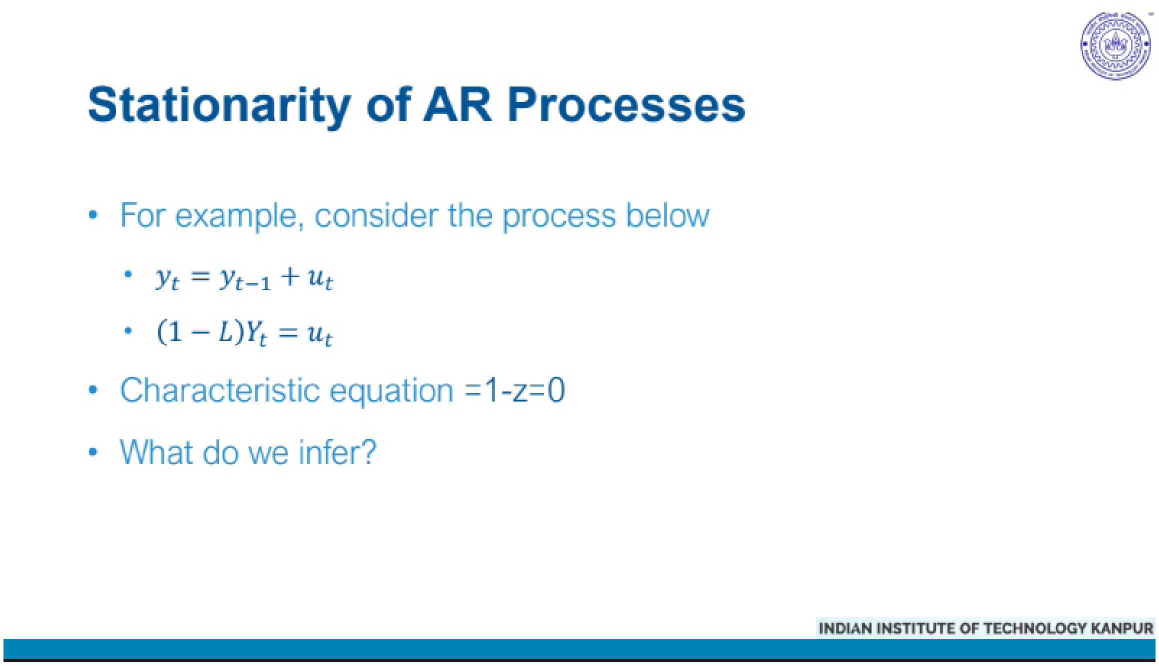
This is assumed to be a white noise iid term. Ideally, if the process is purely an AR process, autoregressive process and we have modeled it appropriately then mu's should be purely white noise. Generally, we put another assumption of normality. So, it is considered to be distributed as normally distributed with the mean 0 and constant variance. Once we have modeled this process in this form, we can take out this phi i, this particular part on LHS and therefore the left side becomes some kind of function of y_t series with phi l where phi l denotes the set of parameters.

Now, why I am using the word l to the power p because often when you characterize this AR process, this is called characteristic of equation of this AR process. I repeat, this is called characteristic equation of AR process. On RHS we have the constant term. Generally, in financial market, when you are modeling financial market returns and those kind of financial market process, this mu's tends to become 0 often, sometimes it not.

And nu t is this white noise IID error term which is normally distributed. Now, later we will see that in algebra, the roots of this characteristic equation are very important to understand this AR process. So, often these roots are very important in determining properties such as stationary and non-stationary AR processes. So, this characteristic equation becomes very important and therefore, these subscripts become powers. For example, this l1 becomes l to the power 1 or lag operator at second level l2 becomes l square and then this will be solved like in any algebraic equation.

So, this is become that similar algebra kind of problems where this equation is called characteristic equation of this ARP process and its root which will be very important to this process which will be able to define the properties of this AR process are employed. Now, in summary and conclusion, we will try to understand what the motivation is to model a process like this AR autoregressive process. Process in financial markets like AR, it is often believed and for that you need to read a little bit of financial market literature like sequential information SIH, sequential information arrival hypothesis, mixture of distribution hypothesis and information asymmetry hypothesis that as information arrives in financial markets, in first place, complete information never releases in one go. So, information itself is released in a piecemeal manner in chunks of information pieces like delta 1, delta 2 over time and as these information pieces arrive, they are not immediately incorporated in prices, it takes some time. So, this arriving information or the information that has arrived incorporates gradually over time in financial market prices.

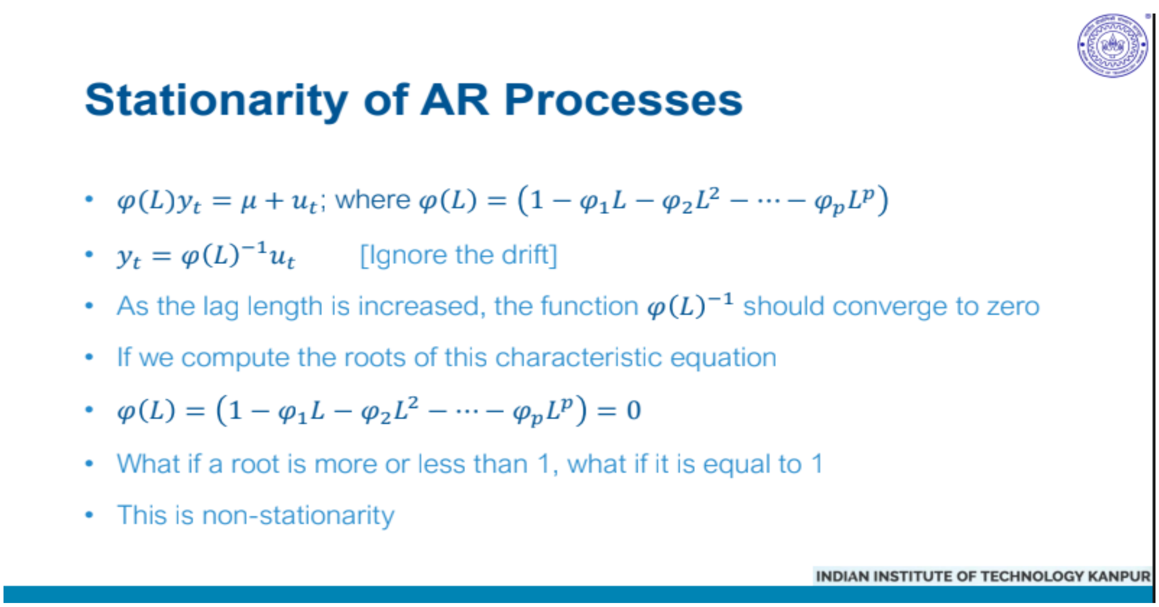
Therefore, by nature, this first in first place this information is autocorrelated because it is being released in the piecemeal manner and also because it is being gradually incorporated in prices by design there is an autoregressive structure in financial market returns that is to suggest that financial market price returns are autocorrelated and that is why theoretically AR process becomes extremely important to model financial market returns. The significance of this muted term is also very important. In econometrics or in modelling statistics, this is often treated as error term with not much value error term or innovation term. However, in financial economics, this mutie has a very important role. Generally, it is considered that when you model this mutie, this is taken to be as the information shock or innovation or information shock that has just arrived at time t.



So it is believed that once you model out the previous historical information structure which has been autocorrelated, like I said, information is arriving in a piecemeal manner, so it is bound to be autocorrelated. Once you model out this autocorrelated information, what you are left with an information shock which is completely unexpected. So this mutie is unexpected information shock. And this information, the information shock has no information content in the previous process. So this is unexpected. Had it been the case that some of this information was already understood or explained by the historical terms, it would have been modelled out. So whatever left unmodelled in this mutie is called unexpected information shocks to this YT process. And that is why modelling these financial market processes such as returns, it is extremely important, the AR becomes theoretically extremely important process. One last point, very important point that often some kind of statistical model may give you a very good fit, which is not an AR. However, that model may not be theoretically justified.

So always remember in modelling financial market and financial economics or any kind of economic series, theory is given preference over statistical and goodness of fit. So even if a mathematical model is giving a very good fit, but it is not theoretically justified, we should not use that we should use a model which is theoretically justified. This is important because ultimately, you are trying to model this process from some policy perspective. Maybe you want to do some kind of policy experiment, maybe you want to look at the impact of GDP on financial market index returns like Nifty Fifty. In that case, theoretical relationships are important.

And even though we are using time series process, but is still a theoretically justified structure is very important to have. And the second implication of that is often such mathematical models, good models, which work very well in in sample, when you model them out of sample in new data, they work very poorly. In contrast, a model which are theoretically justified, which have some theoretical underpinning, even though they are in sample fit in a given sample in they were modeled may not be as good as some kind of very complex mathematical model. When you look at out of sample fit, these models generally work these models which have some theoretical underpinning, they work much better than any statistical model which may not have some kind of underpinning and maybe a complex model as well. In this video, we'll discuss the stationarity of an AR process.



Stationarity for an AR process is very important. We have already discussed what are the problems with a non-stationary process. Also, if a process an AR process is nonstationary, that means the impact of some of the previous values on some of the current and future values does not die away. What essentially means if there was a shock historically in a system, it does not die away and it persists for very long times, which is not very desirable property in modeling any such AR or ARIMA ARMA process. Let's understand what stationarity is mathematically.

- $\varphi(L)y_t = \mu + u_t$; where $\varphi(L) = \left(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p\right)$
- $y_t = \varphi(L)^{-1}u_t$ [Ignore the drift]

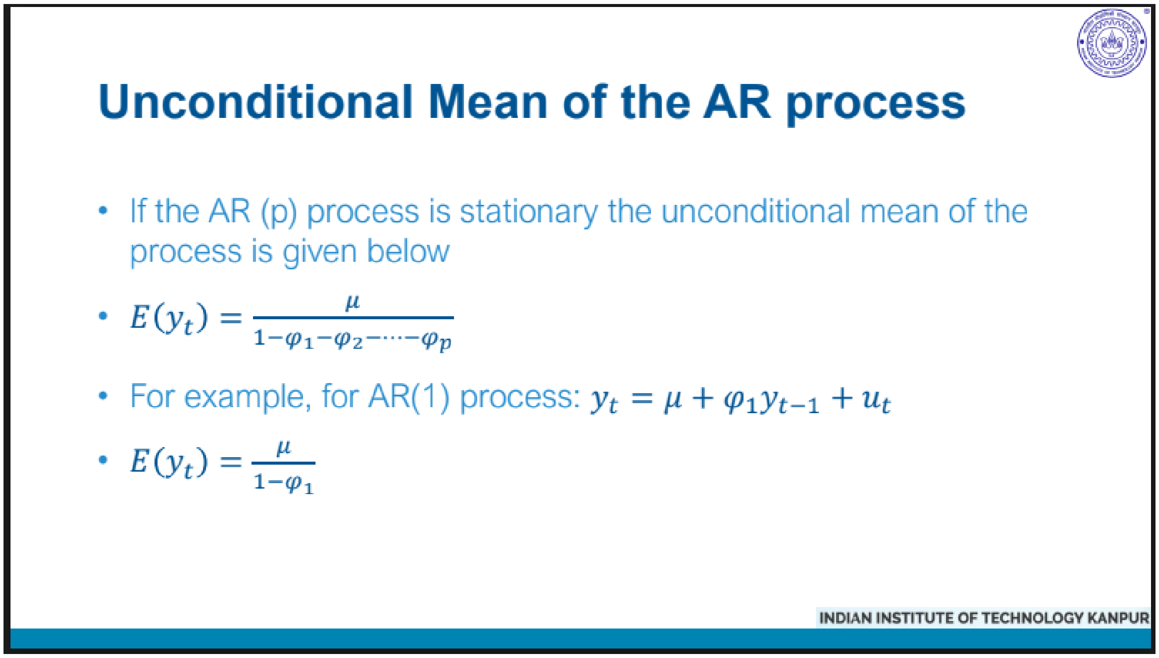
As the lag length is increased, the function $\varphi(L)^{-1}$ should converge to zero

- If we compute the roots of this characteristic equation

$$\varphi(\mathrm{L}) = \left(1 - \varphi_1 \mathrm{L} - \varphi_2 \mathrm{L}^2 - \cdots - \varphi_p \mathrm{L}^p\right) = 0$$

Generally, it is considered to be 0 in financial market process such as return for the following reasons. If this mu takes on some kind of costate value, that means there is some kind of arbitrage to be made to make excess abnormal returns and if it is negative vice versa, that is short positions can be taken to make abnormal returns and therefore, on average this mu should be insignificant or close to 0 and that is why a more practical assumption is to take this as 0, Now, this kind of transformation taking inverse of this $\varphi(\mathrm{L})$ is possible because of something called wold's theorem, which is out of the scope but the point is for this kind of transformation to exist, it is needed that $\varphi(\mathrm{L})$ be a stationary function that means this equation must represent a stationary process. What do we mean by stationary process financially or economically? Economically, if the process stationary that means as you keep on increasing lag length, these φ ones the magnitude of these φ ones should become smaller and smaller and converge to 0. Intuitively, it should be the case because as you go further previously or passed in time, the impact of events or shock that have arrived in historically they should die away gradually over time as time passes and this is what exactly we mean by this stationary condition.

More importantly, that also means that none of these φ should be equal to one or greater than one. If they are greater than one, then not only shock increases, but it explodes. If it is equal to one that means even though it does not explode, it does not die away infinitely ahead in the future, which is not possible in economics and finance or in general in any case. So, we say that if this process is non-stationary or this sequence is non-stationary, then this process is not feasible to model. So, in general, we rely on this characteristic equation and find its roots, this characteristic equation we find its roots and we say if the roots lie outside the unit circle, then the process stationary that means if the roots of this equation, think of this as a polynomial and the root of this equation are greater than zero then this process is stationary.



What essentially mean? Think of one of the root as maybe z1 or z2 and so on, some of the roots and if I put those roots 1 minus phi1 z equal to zero, then essentially z equal to 1 upon phi1, which means there is just a simple, a very simplistic example, we'll have a numerical example in the following discussions in next set of videos, but just think of it, if this z value is greater than 1, then only the φ1 will be less than 1 in magnitude, but if this z value in terms of magnitude is inside the unit circle that means either 1 or less than it, then phi1 will become greater than 1 or equal to 1 which is, which leads to non-stationarity and this is what is not desirable. So, this is the property of non-stationarity which is a non-desirable property and to summarize, we expect the roots of this characteristic equation, this is the characteristic equation of an ARP process, we expect the root of this equation to be outside unit circle to the process to become stationary that means if the process stationary, then all the autocorrelation coefficients, first they will be smaller and smaller and converge to zero as we go ahead past in time that means $\varphi(p)$ will be in terms of magnitude will be very small and second none of these autocorrelation coefficients should be either equal to 1 or greater than 1, none of them should be like that as long as the process is stationary which means all the roots, characteristic roots of this equation lie outside the unit circle. Now that we have understood the theory behind the stationarity of an ARP process, let us try to work out a simple example. Let us consider a simple ARP process For example, consider the process below

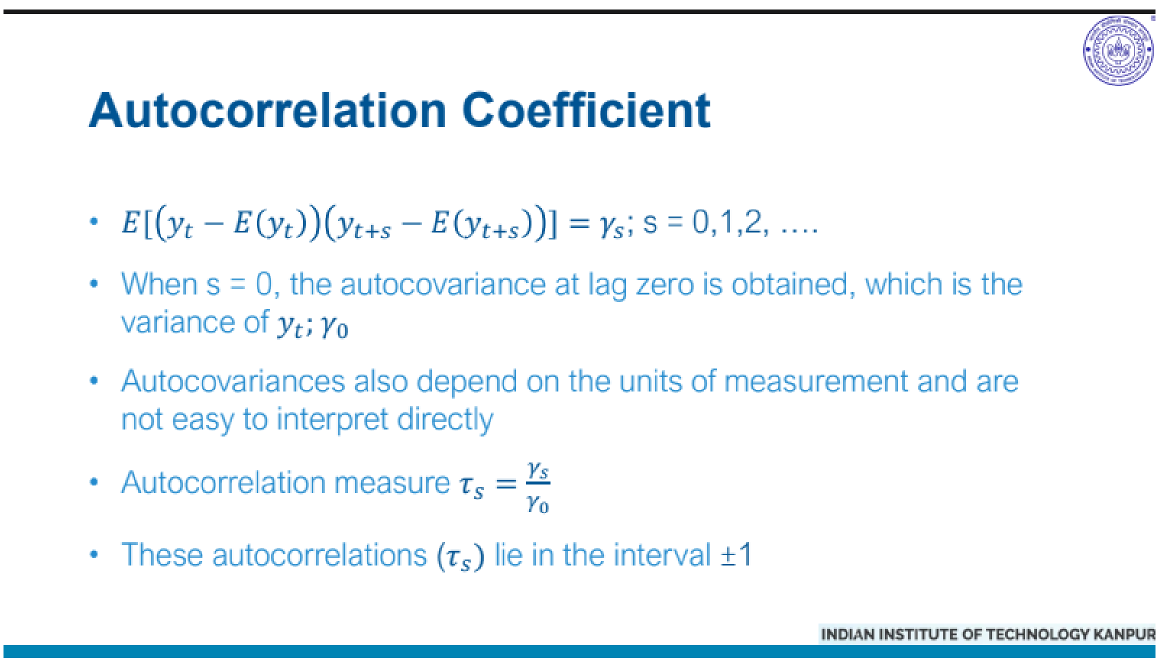
- $y_t = y_{t-1} + u_t$

- $(1-L)Y_t = u_t$

Characteristic equation =1-z=0

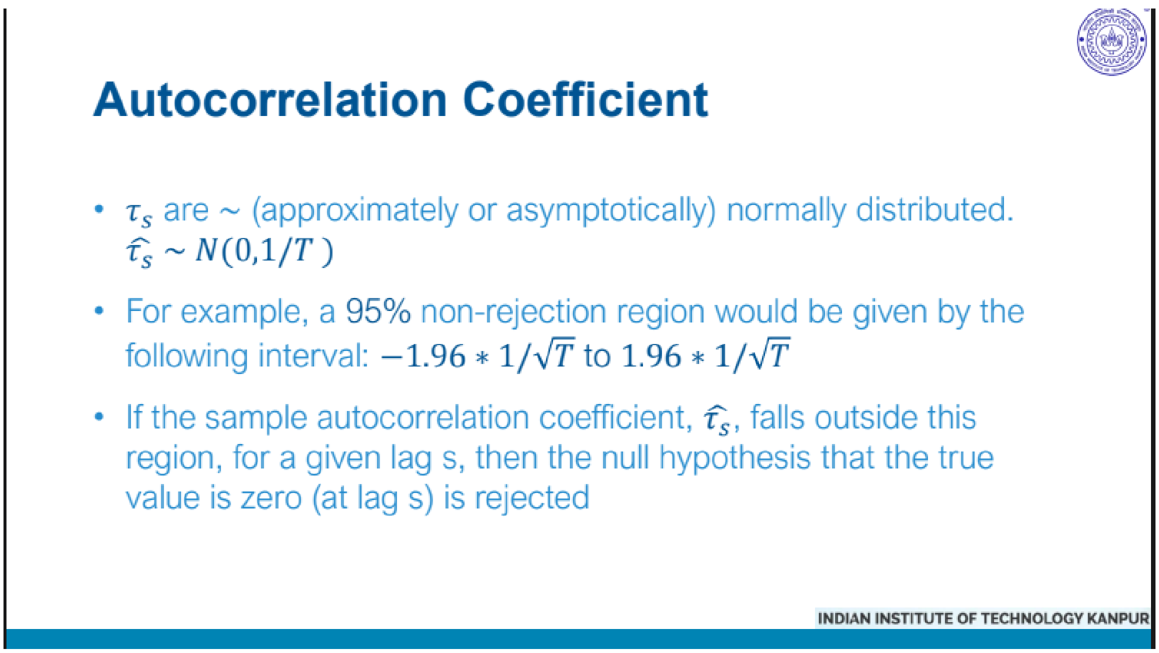
What do we infer? Essentially, this process has a mean of equal to zero as we said in general, we will take the mean as zero and it is only affected by its immediate previous lag that means only the immediate previous term in the series Y_t , Y_{t-1}, Y_{t-2}.

So, for example, Y_{t-1} will affect Y_t, Y_{t-2}will affect, Y_{t-1}but none of the second lag Y_{t-3}. will have no relation with Y_{t-1}or Y_{t-2} will not affect Y_t So, for this kind of process recall we can write Y_{t-1} as L of Y_t and we can put it in a more compact lag notation as $(1-L)Y_t = u_t$. Now with Woltz theorem we have the characteristic equation as $(1-L)$or its algebraic or polynomial notation as 1-z=0. Solving for this equation we get Z =1 which indicates that there is only one characteristic root which is =1 or lies on the unit circle. Recall the implication of Z =1 means that one of the autocorrelations there is only one autocorrelation which is equal to 1 indicating a process which is non-stationary or having what we call as unit root.



Unit root because its root equal to 1 although even without doing this calculation you could have already seen that the autocorrelation here is directly 1 which is not desirable. Why? Because autocorrelation of 1 means that whatever shock that has arrived or was there in the series Y_{t-1} it will never die away as process progresses it will always let us say there is a shock of magnitude m it will always have a multiple of m so in second series m square m cube so the shock will keep on multiplying every time further and it will not die away the shock will always persist with the power of 1 into 1. So, the shock will never die away and it will always persist. Now and therefore the process is non-stationary, and its inverse is also not possible because the inverse would not be possible as per the Wold's theorem if you recall what we said as φ(L) inverse which essentially in this case is $\frac{1}{1-L}$ and the root characteristic root being 1 this is not defined. So, and therefore this process is not a stationary process it has a unit root which is exactly that is Z=1 or indicating then process where the shocks do not die away.

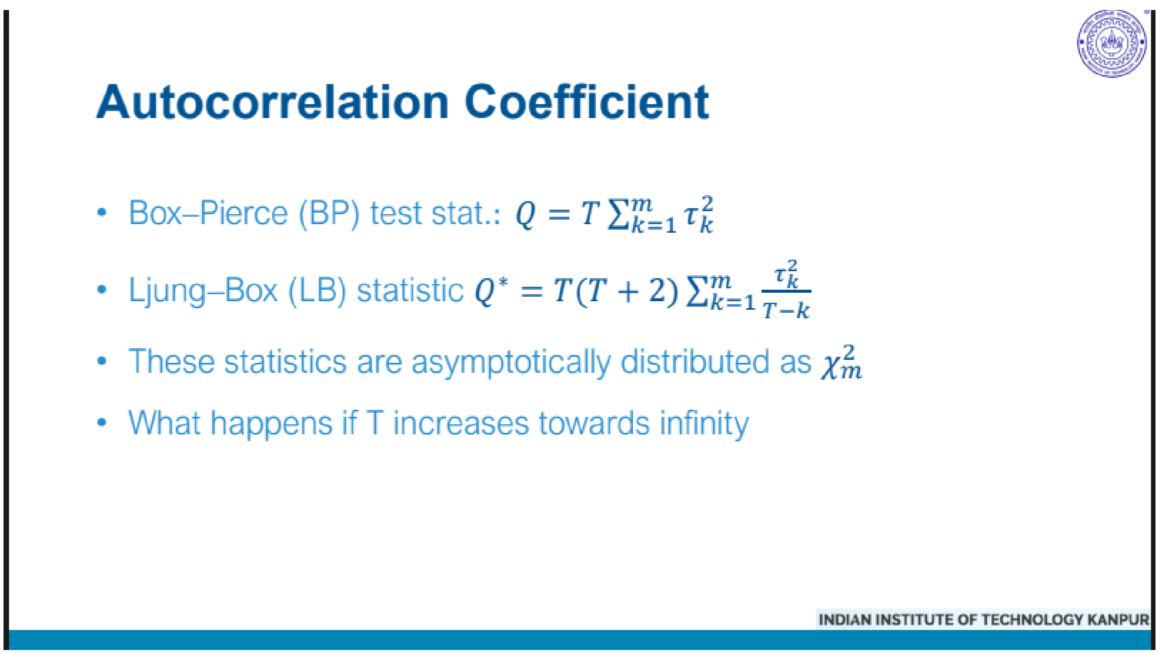
Moreover, the unconditional mean of an AR process is given by this expression below. Unconditional mean you can think of as a long term mean if there is no shock to process and we keep on moving ahead in future without any shock in the long term the process mean will converge to this although in the case of financial economics this is not a very relevant or important result for us why because in general we tend to assume this μ or the mean of the process is 0. So, in that case whatever the shocks may be the unconditional mean in long term will always be 0 by the simple property this mu is already considered as 0 but still it is important to note that there exists some kind of unconditional mean. In general if we go outside the financial economics region, in general in economics an unconditional mean is something which is called long term mean. So, if we keep on projecting the process and there are no new information incorporated in the long term the process converges the projections or forecast from the process will converge to its unconditional or long term mean which is a desirable property.



However, like I said in our case financial economics this μ= 0 the mean is said to be 0 and therefore this long term unconditional mean also becomes 0. Now in case of let us say this AR1 process let us say this while we are not putting this μ =0 and there is only one autocorrelation which is φ1 please recall this expression will become $E(y_t) =$

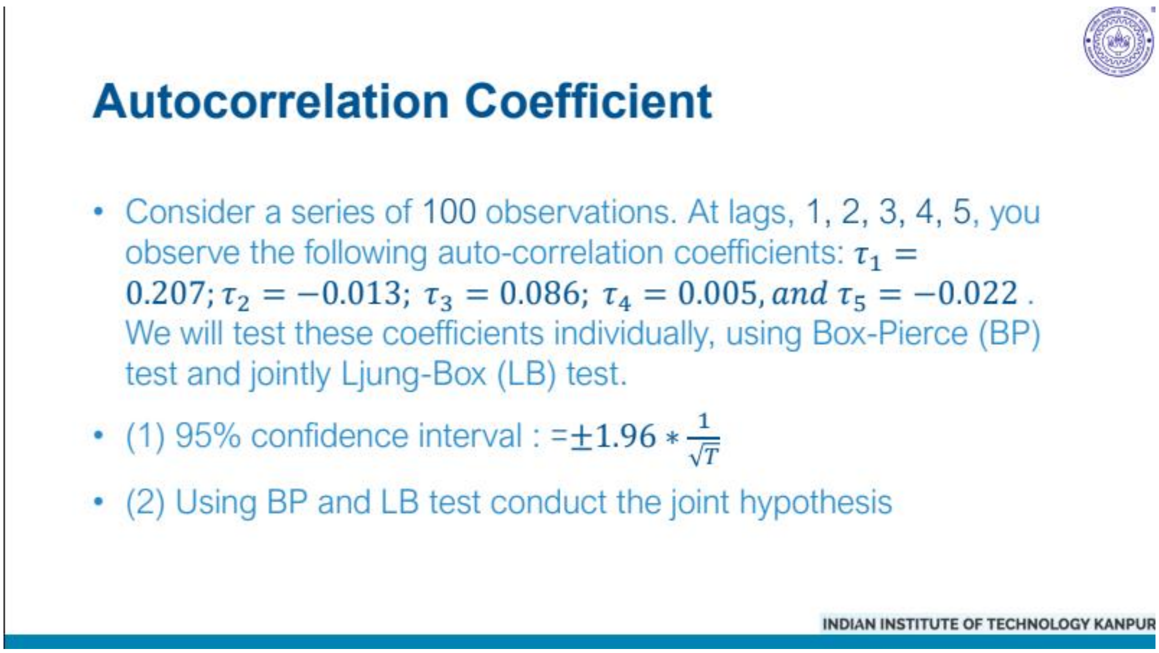
$\frac{\mu}{1-\varphi_1-\varphi_2-\cdots-\varphi_p}$ .So, recall the φ_1 our expression is φ_1 would be 1- $\varphi_1$and essentially when you take the inverse this 1- $\varphi_1$will phi inverse will be $\frac{1}{1-\varphi_1}$. Now in this case as long as this φ_1 if you recall if this as long as this is less than 1 in terms of magnitude then this process stationary but as long as as soon as this φ_1 becomes 1 which means the autocorrelation is 1 this process becomes non-stationary as we saw in the previous example. Also, if it is more than 1 then also this inverse is not possible the wolff's theorem will not result in this kind of inverse and again the process will not be stationary.

For this process to be stationary this φ_1 has to be less than 1 in terms of magnitude then only the characteristic root of this equation which is $1-\varphi_1$ z will be equal will be available that means they will not be 1. For example if $\varphi_1$then characteristic root equal to 1 or if $\varphi_1$is let us say more φ_1 is greater than 1 then z will be less than 1 that means again the root of this characteristic equation will be either on the unit circle or inside it which is which leads to non-stationary process. So to summarize in this video we discussed what and with a simple example what would be the implication of non-stationary non-stationarity of an AR process and how do we model an AR process and convert it into a more amenable form like this and this. We also saw with the simple AR 1 process how to convert into through wolff's theorem convert into a polynomial and obtain its characteristic root characteristic equation and then once the characteristic equation is available through simple programs even with a calculator or a simple program you can obtain its characteristic root and as long as those characteristic root lie either lie outside the circle the process is stationary. However, if the characteristic root lie on the circle or inside the circle the process is non-stationary.



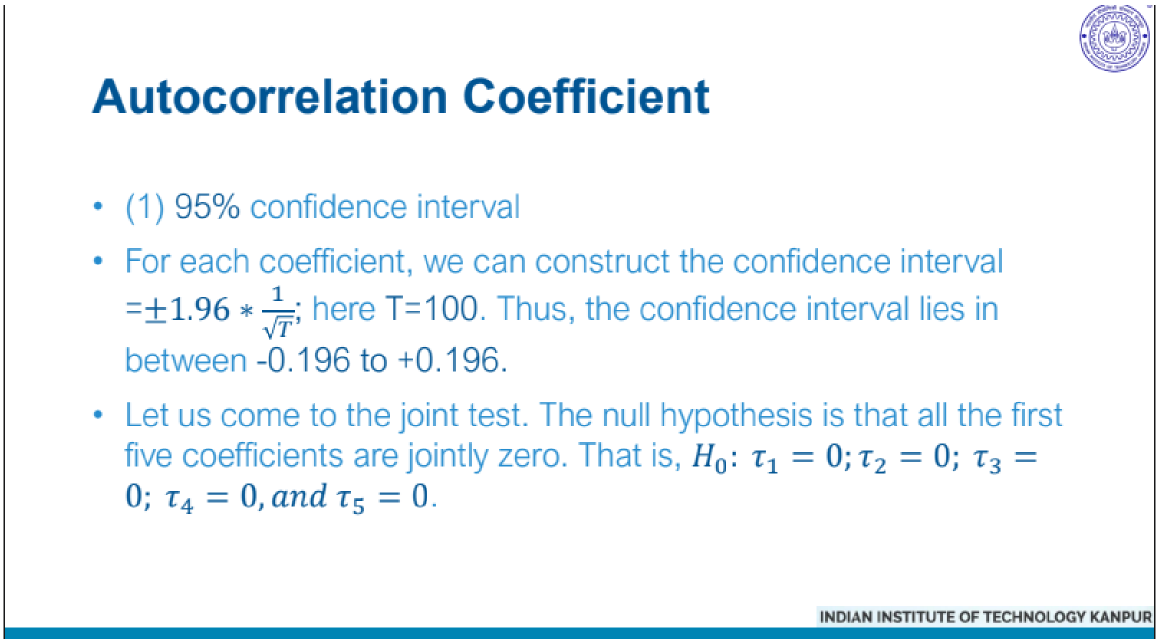
What is the implication of characteristic root lying inside the circle or on the circle? It means that either the autocorrelation is equal to 1 which means shocks to the series will not die away or if it is inside 1 then they not only not die away but they explode which is not a desirable or feasible property in finance and economics. In this video we will again introduce the autocorrelation coefficient and its role in identifying the structure of a time series process. If you recall the basic expression for autocovariance for a time series process $y_t$at lag s was $E[(y_t - E(y_t))(y_{t+s} - E(y_{t+s}))] = \gamma_s$; s = 0,1,2, ….

This gamma s is autocovariance at different lags of 0, 1, 2 and so on. When this s equal to 0 the autocovariance at lag 0 is obtained which is simply nothing but the variance of y_t itself which can be noted as gamma naught.



One problem with these autocovariances that they are they have different units of measurement if you are different measuring the process different process and therefore it is difficult to compare and interpret them directly because of this scale issue and therefore a good way is to scale or standardize them with the variance gamma naught. So, if you believe the process has a constant variance that is gamma naught which is nothing but the autocovariance of a process with itself that is $y_t$at s = 0 which means this will also become y_t this becomes variance term and assuming if for a process variance is constant that is gamma naught. If I divide any autocovariance gamma as the gamma naught it is called autocorrelation measure tau s as we have seen earlier. Now this tau s has a very interesting property that it lies between -1 to +1 where -1 means completely opposite movement opposite lock step movement while +1 means completely in one direction generally in like in practical scenario some the number will be somewhere in between. Now what is the importance here? Before we explain the importance one thing to note as s = 0 τ_s becomes τ_0 which is nothing but gamma naught upon gamma naught equal to 1.

So, the autocorrelation of a process with itself contemporary is term is 1. Now generally for a white noise process and assuming and recall we said that we generally tend to assume that is a normally distributed white noise process. So, a white noise IID process with a normal distribution which has no autocorrelation structure if you simulate the



autocorrelations τ_s they are approximately that means in large numbers if they are large number of observations asymptotically, they are normally distributed. So, for a white noise normal IID process the autocorrelation themselves will be asymptotically that means with large number of observations are approximately normally distributed with a mean of 0 and a variance of 1 upon t that means if we plot it, it will appear something like a normal distribution of the shape curve with a mean 0 and a variance of 1 by t. Now using this property, we can set up a confidence interval for any such process where we want to establish whether there is some autocorrelation structure or not.

τ_s are ~ (approximately or asymptotically) normally distributed. $\hat{\tau_s} \sim N(0,1/T)$. For example, a 95% non-rejection region would be given by the following interval: $-1.96 * 1/\sqrt{T}$ to $1.96 * 1/\sqrt{T}$. If the sample autocorrelation coefficient, $\hat{\tau_s}$, falls outside this region, for a given lag s, then the null hypothesis that the true value is zero (at lag s) is rejected.

Autocorrelation Coefficient

- Let us come to the joint test. The null hypothesis is that all the first five coefficients are jointly zero. That is, H_0: $\tau_1 = 0; \tau_2 = 0;\ \tau_3 = 0;\ \tau_4 = 0, and\ \tau_5 = 0$. The relevant critical values are obtained from χ^2 distribution with five degrees of freedom. These are 11.1 at 5% level, 15.1 at 1% level.
- The test statistic for BP test is given below. $Q = T\sum_{k=1}^{m}\tau_k^2$
- The LB statistics is computed as follows. $T(T+2)\sum_{k=1}^{m}\frac{\tau_k^2}{T-k}$

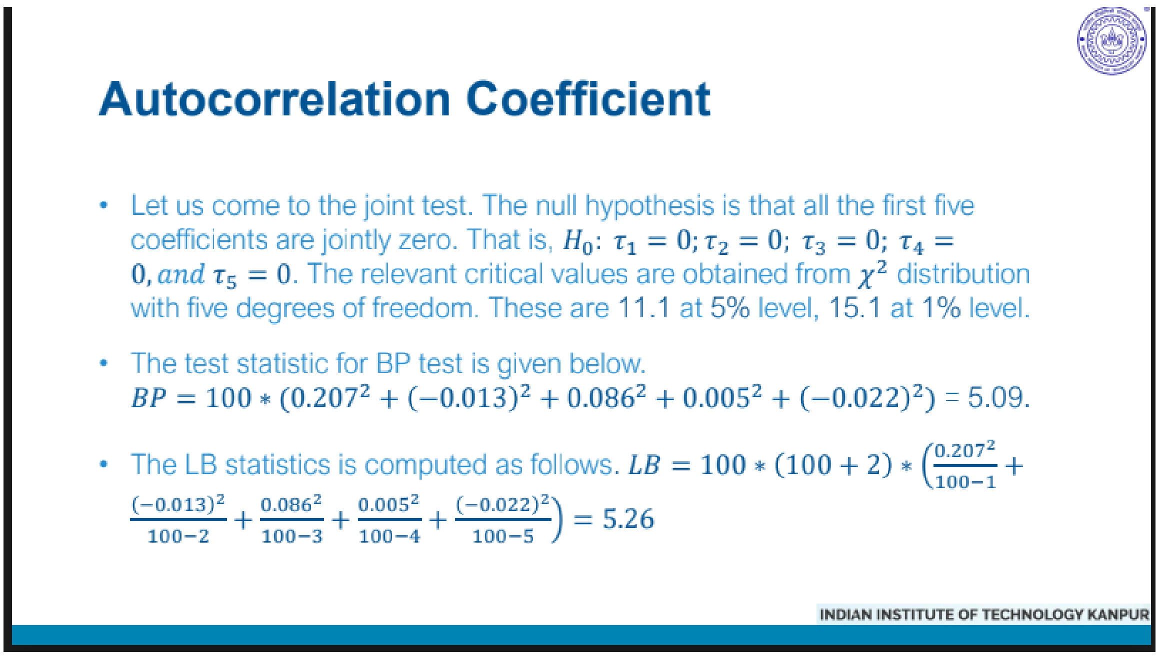
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Similarly, you can set up 99 percent or any other confidence interval using the corresponding z values. So, in summary we are saying that as long as we are setting this 95 percent confidence interval with the assumption that process is a white noise IID process that means there is no discernible structure you can reject the null if your sample autocorrelation τ_s lies outside this region for a given lag s and that would indicate the rejection of null hypothesis that means if the correlation is outside this region and you are able to reject the null which would suggest acceptance of alternate Ha in general we call alternate as Ha hypothesis that there is indeed a significant I repeat significant or what we call is non-zero coefficient at lag s now whether that coefficient is positive or negative that will depend whether you would get in this tail or the right tail. If it is rejection in right tail then probably it is positive non-zero positive if it is on the left tail then it is for non-zero negative. Now there are two advancements in this statistic one is called Box-Pierces test statistic which is $Q = T\sum_{k=1}^{m}\tau_k^2$ and LB test, Ljung–Box which is called $Q^* = T(T+2)\sum_{k=1}^{m}\frac{\tau_k^2}{T-k}$Now these are two are slightly different from the previous hypothesis or confidence interval estimation of autocorrelation coefficient.

In the previous case we were testing autocorrelation coefficient at different lags individually these are joint test of autocorrelation coefficient that means when I run this test for Q or Q^* essentially I am testing that up to the lag k up to this lag k all the coefficients are jointly equal to 0. I repeat the null hypothesis here is that all the coefficient up till lag k are 0. The rejection of null that is acceptance of Ha will indicate that at least one of them is significant. So, these are slightly more powerful test that statistics are asymptotically distributed as a chi-square statistic. So, you pick up the relevant values from the chi-square this chi-square table and compare them with the Q and Q^* and depending upon whether Q and Q^* are higher or lower than these statistics you reject or accept the null.

So, we will see some of the examples in next set of video. Notice if t is sufficiently large then both of these statistics they converge towards the same value that means Q^* and Q tends to become same they converge towards each other when t is very large and increases towards infinity. Now, to summarize what is the important what is the application or important role of this autocorrelation coefficient. The idea here is that the good idea here is that that when you model when you model any process any time series process through models such as arima model and extract the residual out of it μ_t is those residual you assume you tend to assume as a null hypothesis that these residuals are normally distributed they are like a white noise process with a normalized distributions that is independent and identically distributed with a normal distribution. However, then you apply these kind of null with either individual hypothesis test that we did in the previous discussion or now with Q and Q^* statistic and you are able to reject the null that there is no autocorrelation then you have to conclude that there is some autocorrelation structure and then probably your model modeling of process was not a very good fit and that is the whole point of it to find this autocorrelation structure in the process whether there is some reasonable structure and your error terms indeed are not white noise if they are not white noise then the model that you have fit to your model original process y_t because y_t was your original process which you fitted with some kind of arima or some model to obtain these error terms that original model parameters were not very appropriate or model itself was not very well specified and you need to relook at your modeling exercise that is the conclusion with this. So, to summarize this video we have introduced again autocorrelation coefficient and we also saw their application in modeling time series processes. Please note when you plot these autocorrelation coefficient on x axis with different labs s equal to 0, 1, 2 and so on this kind of diagram is called autocorrelation function or ACF or correlogram when you plot it visually across lags. Now that we have understood the theoretical underpinnings of autocorrelation coefficient we will try to explain its working and application with the help of a simple example. Consider a simple experiment or examination of a process y_t where you have 100 observations and at lags 1, 2, 3, 4, 5 you observe the following correlation coefficients first order coefficient $\tau_1 = 0.207; \tau_2 = -0.013;\ \tau_3 = 0.086;\ \tau_4 = 0.005,\ and\ \tau_5 = -0.022$. You want to test first individually whether these correlations are indeed statistically significant that means non-zero by that confidence interval approach that is the setting of this 95% confidence interval and checking each of the correlation coefficients individually whether they are non-zero and statistically significant and the second joint test or joint hypothesis approach with using Box-Pierce (BP) test and jointly Ljung-Box (LB) test. to conduct the joint hypothesis whether all of them together are significantly or non-significant or equal to 0. Let us start with the simple 95% confidence approach where we would like to set up 95% confidence interval: $= \pm 1.96 * \frac{1}{\sqrt{T}}$. here T=100. Thus, the confidence interval lies in between -0.196 to +0.196.

Now if we compare our coefficients individually one by one let us see so by that logic this coefficient 0.201 is significant this is not significant this is not significant this is not significant and this is not significant so that means only one autocorrelation coefficient at lag 1 is significant and non-zero and this is positive also so this is what we conclude from the first test. Let us look at the joint test. In the joint test our null hypothesis is that all of them $H_0\text{: } \tau_1 = 0; \tau_2 = 0;\ \tau_3 = 0;\ \tau_4 = 0,\ and\ \tau_5 = 0.$ all of them are jointly equal to 0 and now we would like to compute the statistics respectively Q and Q^* with these formulas t here is 100 all the τ_{k} are given τ_1 τ_2 and τ_3 τ_4 τ_5 are given to us and it is given to us that statistics corresponding to 5% significance or alternatively 95% confidence level is 11.1



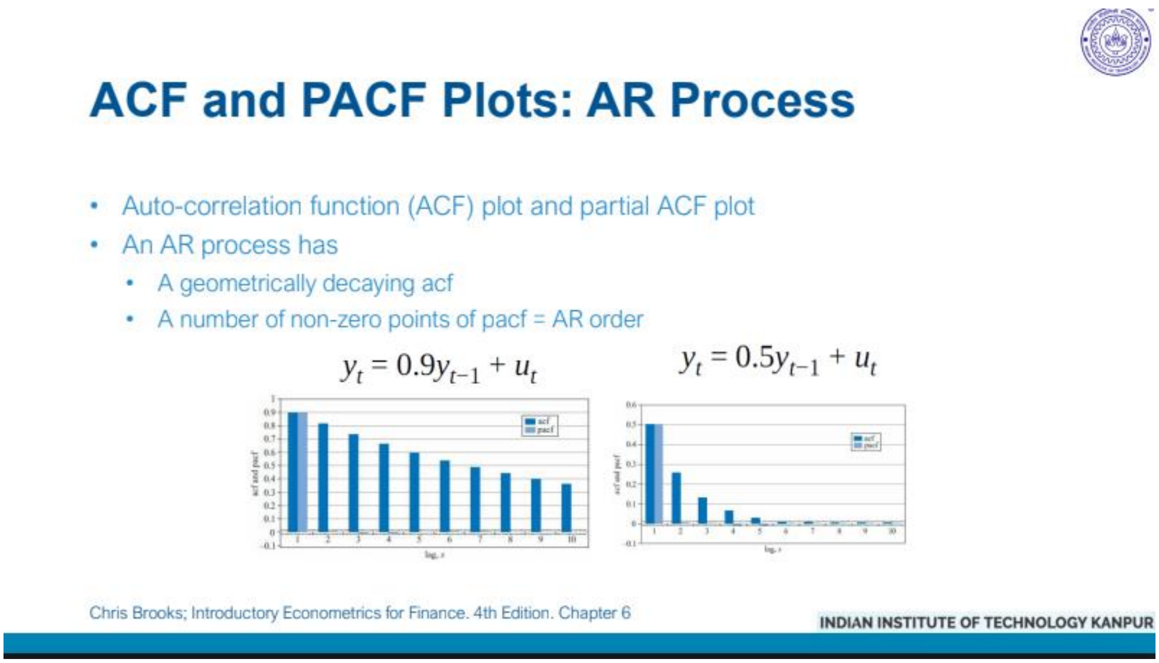
and at 1% significance that is 99% confidence is 15.1. The test statistic for BP test is given below. $Q = T \sum_{k=1}^{m} \tau_k^2$. The LB statistics is computed as follows. $\mathrm{T(T+2)} \sum_{\mathrm{k=1}}^{\mathrm{m}} \frac{\tau_{\mathrm{k}}^2}{\mathrm{T-k}}$. So let us work out the numbers so if I compute the box pair statistic $BP = 100 * (0.207^2 + (-0.013)^2 + 0.086^2 + 0.005^2 + (-0.022)^2)$ = 5.09. Similarly, LB statistics is also given to us which is $LB = 100 * (100 + 2) * \left(\frac{0.207^2}{100-1} + \frac{(-0.013)^2}{100-2} + \frac{0.086^2}{100-3} + \frac{0.005^2}{100-4} + \frac{(-0.022)^2}{100-5}\right) = 5.26$. Now if you recall our statistics at 5% and 1% level none of them are significant that means whether I compare with 5% or 1% both the statistics are insignificant.

What do we conclude here? In both cases with the help of joint null hypothesis that all of them are 0 we are fail we fail to get the null that means we do not accept the ultimate and we conclude that all the coefficients are equal to 0 cannot be rejected. So, we have to take them together as non-zero. However, this is in contrast with the first confidence interval approach where the first coefficient of 0.206 appears to be significant individually while others were non-significant. So, why this kind of contrasting result? So, to summarize or conclude here what is happening is that at lag 1 the coefficient is pretty large it is 0.206 which is significantly different from 0. However, when we conduct the joint hypothesis that means all of them whether all of them are equal to 0 given the small magnitude given the small magnitude of these tests that for example τ_2, τ_3 and so on these are so small that the overall power of joint test is decreased because of the smaller coefficient, and they go in favor of null. I repeat because of the small magnitude of these coefficients the overall power of test is pretty low and we have to conclude when we conduct the joint hypothesis, we have to go ahead with the null that all of them are together equal to 0. But when we examine them individually clearly the first lag auto correlation coefficient is pretty large and we have to conclude that this process has a significant first order auto correlation.

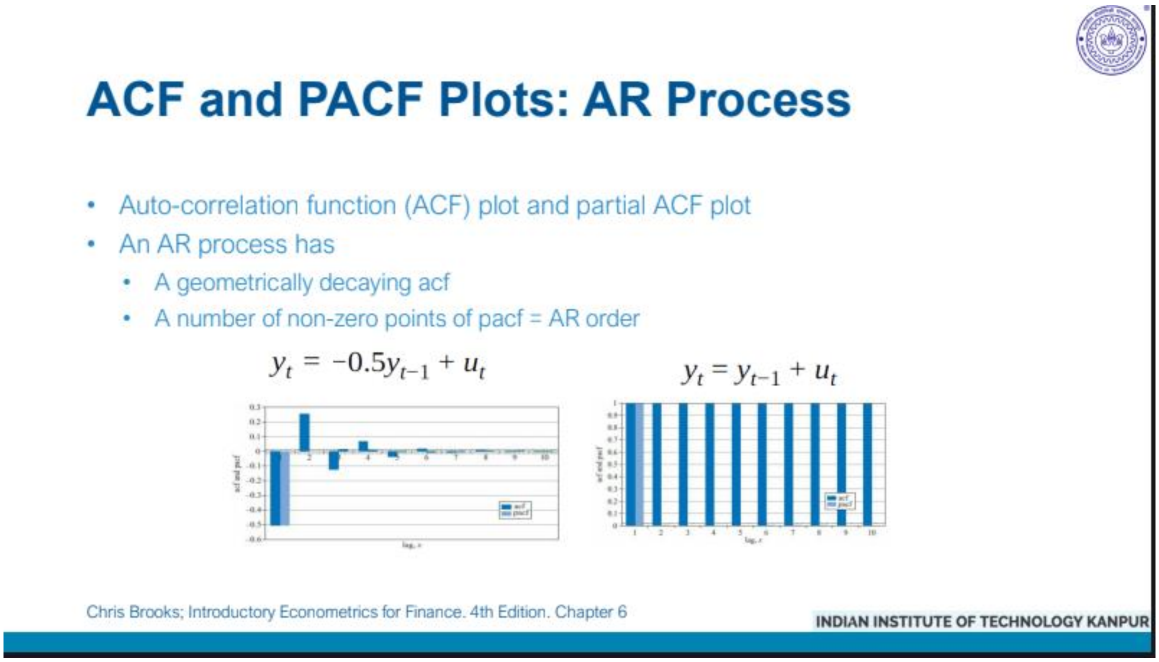
Now here μ is almost 0 we consider μ to be 0 and this kind of process will appear. This is error term which is generally a white noise process so you can consider this error term and the process will appear like this. In this video we will introduce auto correlation function and partial auto correlation function and understand them with the help of AR process and how to differentiate AR process with the help of ACF and PACF diagrams. To begin with we have already understood that auto correlation function or correlogram is a plot of these auto correlations across different lags starting from 0, 1, 2 and so on and that this auto correlation function at lag 0 takes a value 1 and then subsequent lags it takes different values. However, the problem is that because the process is correlated for example there is some kind of correlation between y_t and y_{t-1}. If even if there is a single order correlation that is some correlation φ between y_t and y_{t-1} and y_{t-1} and y_{t-2} there is a spurious correlation mathematically there is a spurious correlation that results between y_t and y_{t-2} which is smaller than φ somewhere around $\varphi * \varphi$ closer to φ^2 square driven by the fact that there is a correlation between y_t and y_{t-1} and $y_{t-1}$and y_{t-2}.

So closer to φ^2 kind of correlation appears between y_t and y_{t-2} and φ^3 between y_t and y_{t-3} and so on. And therefore, if I want to know the exact auto correlation structure, I need to go with something called PACF partial auto correlation function which accounts or sort of accounts for this spurious auto correlation. Now how does an AR process figure on these ACF and PACF grams? Conceptually when you have an auto correlation structure the AR process has a geometrically decaying ACF. I repeat AR process has a geometrically decaying ACF because of that spurious correlation. However, the PACF process accounts for this spurious correlation and only shows the exact auto correlation structure that is present and therefore on PACF diagram you have nonzero points only equal to AR order.

For example, if your actual auto correlation structure is of two lag that is with lag 1 and lag 2 then PACF will have two non-zero points at lag 1 and 2 and rest of all will have zero values. Let's look at how it appears. We have a simple AR1 process with first order auto correlation of 0.9. So, these dark blue bars are ACF and notice it takes a value of 0.9 at first lag but second lag it is closer to 0.81 and then 0.89 cube and so on so kind of exponentially declining structure with ACF diagram. As we discussed the rest of these the first co auto correlation is correct but rest of them are spurious.

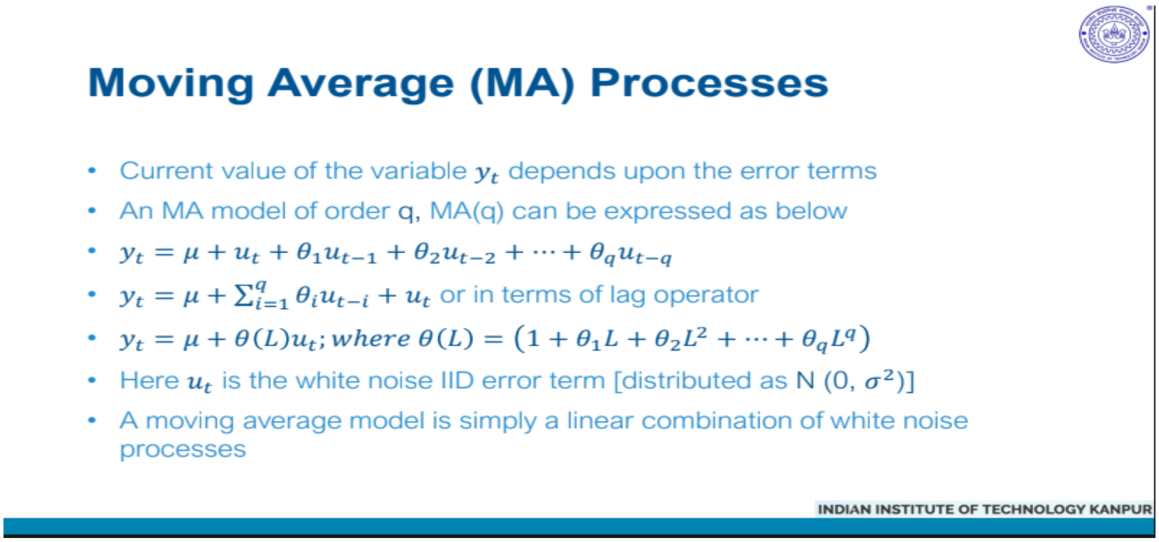


So how PACF accounts for it. Let's look at the lighter blue diagram lighter blue bar which is PACF. The first bar at first lag is exactly identical as the ACF diagram but then it is cut off there itself then there are no lags. So that means PACF diagram clearly helps us identify that this is a first order auto correlation structure. Similarly look at this AR1 process where first order auto correlation is 0.5. The first lag at lag 1 we have ACF and PACF values are non-zero and same as 0.5 but then ACF exponentially declines while PACF there are the values are straightaway cut off to 0. So in this fashion PACF helps us in identifying their process. Please note that if there were two AR terms they will be appearing on PACF process in the form of two bars and so on for 3, 4 and more auto correlation structure but on ACF they appear as exponentially declining bars. Let's look at another two diagrams.



Interestingly in this we have AR1 process with a negative auto correlation structure. So first at lag 1 we have value of -0.5 which is same for PACF and ACF both this PACF and ACF. However after that you will notice that it takes -0.5^2 which is close to 0.25 and therefore magnitude exponentially declining but sign is continuously changing with the power of -0.5^n depending upon n first 1, 2, 3 and so on this value the bar is positive or negative but on PACF there is only one bar which is first lag -0.5. Again PACF helps us in identifying the AR process its lag structure. If you look at a non-stationary process notice this is a non-stationary process because here the auto correlation coefficient is 1. So it is on the unit root circle and notice on ACF diagram because of this property that shocks are not declining it is a non-stationary process and shocks are not decaying on ACF diagram we have continuously all the bars at 1 including 1, 2, 3 and all the subsequent lags.

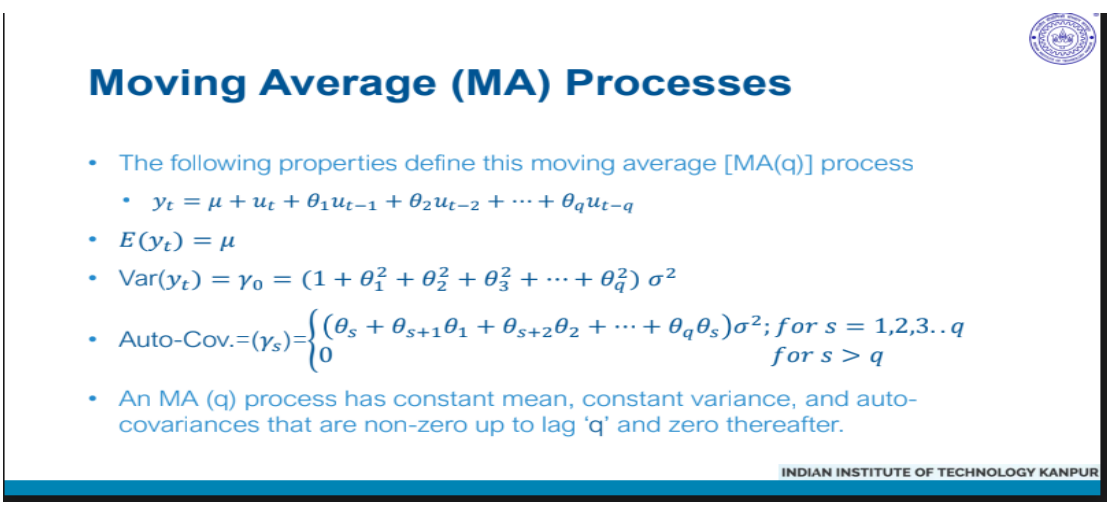
However again on PACF the bar is only non-zero and takes a value of 1 at first lag and then it is cut off. So this identifies the process non-stationary because now the ACF bars are non-declining they are constant which shows non-decay there is no decay of shock which is sort of indication of a non-stationary process but PACF again we can identify that non-stationarity appears at first lag with the order of 1. So to conclude and summarize in this video we saw how with the help of ACF and PACF process we can identify AR process. So when we look at AR process it is ACF is geometrically declining for a stationary process and PACF has exactly non-zero lags which are same as the order of the process AR process. In this video we will introduce another very important class of time series processes called moving average processes or MA processes.



This form means that in ARIMA model that MA component moving average process a very important component of ARIMA model. Let us introduce the model and concept behind the MA process. $y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$

$y_t = \mu + \sum_{i=1}^{q} \theta_i u_{t-i} + u_t$ or in terms of lag operator.

$y_t = \mu + \theta(L)u_t; where\ \theta(L) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)$. Here u_t is the white noise IID error term [distributed as N (0, σ^2)]. Conceptually these μ_t are like information shocks that have arrived in the process at time t. Think of day t let us think t in terms of days and any sentiment or information shock that has arrived on day is modeled as μ_t. Now a white noise process μ_t with a sort of zero mean that is expected value of μ_t as zero and a constant variance σ^2 the variance of μ_t as σ^2 are employed to model this μ_t term. Generally, it is believed that information that is arriving has no connection with its own self previously or historically for example the information that has arrived today probably may have no relation with information that has arrived tomorrow or yesterday.



μ here is the average some kind of average value or constant value for practical modeling purposes later on we will consider this μ_t as 0 so this constant term is 0. Let's take 30 seconds to understand what this constant mean is this constant mean for a process like security market return is average return average year average period by period return. Now if you are talking about financial market and a stock gives you a fixed abnormal extra average period by period returns that sort of violates the concept of market efficiency and therefore for all practical purposes this number should be very close insignificantly different from 0 and therefore we may take it as 0 and that will help us ease the mathematical modeling part of the process as well. So let's begin this assumption that this mu is very close to 0 and all we are interested in is modeling these current and historical shocks or white noise processes mu t mu t minus 1 and so on and theta 1 theta these thetas being the correlate sort of coefficients of these shocks. The following properties define this moving average [MA(q)] process

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$$

$$E(y_t) = \mu$$

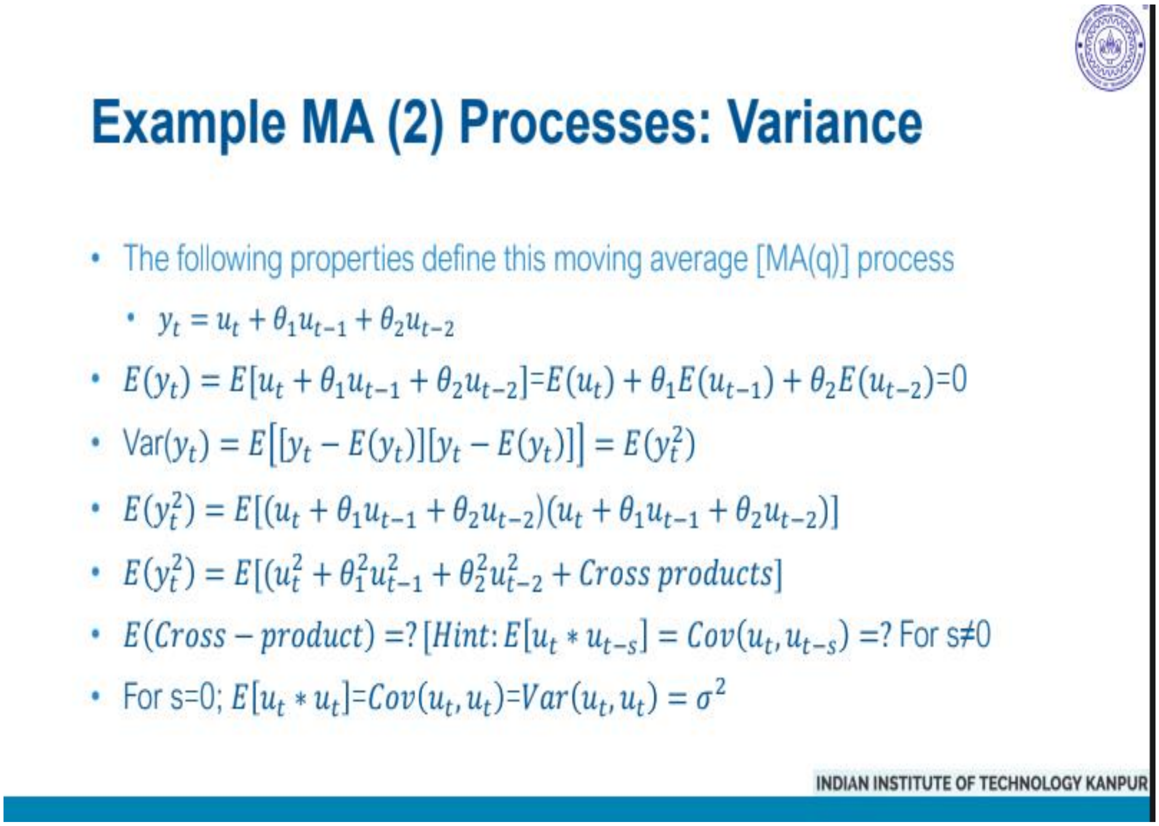
$$\text{Var}(y_t) = \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \theta_3^2 + \cdots + \theta_q^2)\ \sigma^2$$

$$\text{Auto-Cov.}=(\gamma_s)=\begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \cdots + \theta_q\theta_s)\sigma^2; & for\ s = 1,2,3..q \\ 0 & for\ s > q \end{cases}$$

An MA (q) process has constant mean, constant variance, and auto-covariances that are non-zero up to lag ‘q’ and zero thereafter.

Last another very important point while modeling this we have taken μ_t is to be white noise iid terms that are distributed with a zero mean and constant variance for ease of modeling we take the distribution at which they are distributed as normal so this iid or independently identically distributed μ_t are taken as normally distributed. The following properties define this moving average [MA(q)] process: $y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$

- $E(y_t) = E[u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}] = E(u_t) + \theta_1 E(u_{t-1}) + \theta_2 E(u_{t-2}) = 0$
- $\text{Var}(y_t) = E\big[[y_t - E(y_t)][y_t - E(y_t)]\big] = E(y_t^2)$
- $E(y_t^2) = E[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})]$
- $E(y_t^2) = E[(u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2 + Cross\ products]$
- $E(Cross - product) =?$ $[Hint{:}\ E[u_t * u_{t-s}] = Cov(u_t,\ u_{t-s}) =?$ For s≠0
- For s=0; $E[u_t * u_t] = Cov(u_t,\ u_t) = Var(u_t,\ u_t) = \sigma^2$



And to summarize a moving average model is simply a linear combination of these white noise mu t's processes of current at current time and historical times. Once we have finalized these assumptions about the MA process we can derive certain useful results so look at this μ_t process or time series MA q order process so these mu t's run up till time lag q this is considered as an MA process of order q. There are some desirable results the expected value or the mean or this process is mu again this will become zero for us while later on when we are modeling these results but still theoretically this value is μ but in financial markets this will be taken as close to zero. The variance of μ_t is this constant number which is a function of the variance of sigma square which is the variance of μ_t itself and summation of these theta one square theta two square which is nothing but the coefficient of these μ_t. The following properties define this moving average [MA(q)] process

- $y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$

- $E(y_t) = E[u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}]$=$E(u_t) + \theta_1 E(u_{t-1}) + \theta_2 E(u_{t-2})$=0
- Var$(y_t) = E\big[[y_t - E(y_t)][y_t - E(y_t)]\big]$
- $E(y_t^2) = E[(u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2] = (1 + \theta_1^2 + \theta_2^2)\ \sigma^2$

Example MA (2) Processes: Variance

- The following properties define this moving average [MA(q)] process
  - $y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$
- $E(y_t) = E[u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}] = E(u_t) + \theta_1 E(u_{t-1}) + \theta_2 E(u_{t-2}) = 0$
- $\text{Var}(y_t) = E\big[[y_t - E(y_t)][y_t - E(y_t)]\big]$
- $E(y_t^2) = E[(u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2] = (1 + \theta_1^2 + \theta_2^2)\ \sigma^2$

So because we have considered an MA process of order q these auto covariances structure will start from lag one and up till lag q. So the q, qr lag will be the last lag by which auto covariances will exist after any lag higher than lag q auto covariances will be zero. So to summarize these notation what we can call about this MA process how we can summarize this MA process as for MA q order process it has a constant mean mu a constant variance this one which also depends how many what is the order of the process theta q and auto covariances that are non-zero up to lag q and zero thereafter. So up till lag q they are there and zero after that. So now we have understood the concept and mathematical model behind the MA process in next set of videos we'll try to work out some of these MA process with the help of simple examples. The following properties define this moving average [MA(q)] process

- $u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$

- $\gamma_1 = E\big[[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]\big] = E[y_t * y_{t-1}]$
- $\gamma_1 = E\big[[u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}][u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3}]\big]$
- $\gamma_1 = E\big[[\theta_1 u_{t-1}^2 + \theta_1 \theta_2 u_{t-2}^2] + Cross\ Product\ terms\big]$
- $\gamma_1 = (\theta_1 + \theta_1 \theta_2)\ \sigma^2$
- $\gamma_2 = (\theta_2)\ \sigma^2$
- $\gamma_3 = 0$

Example MA (2) Processes: Auto-Covariance

- The following properties define this moving average [MA(q)] process
  - $u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$
- $\gamma_1 = E\big[[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]\big] = E[y_t * y_{t-1}]$
- $\gamma_1 = E\big[[u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}][u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3}]\big]$
- $\gamma_1 = E\left[[\theta_1 u_{t-1}^2 + \theta_1\theta_2 u_{t-2}^2] + Cross\ Product\ terms\right]$
- $\gamma_1 = (\theta_1 + \theta_1\theta_2)\, \sigma^2$
- $\gamma_2 = (\theta_2)\, \sigma^2$
- $\gamma_3 = 0$

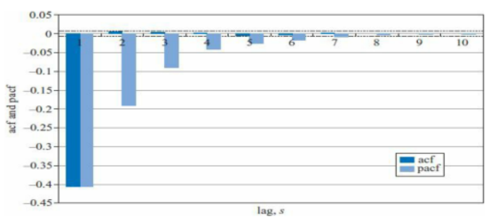
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

- $\tau_0 = \frac{\gamma_0}{\gamma_0} = 1$
- $\tau_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\theta_1+\theta_1\theta_2)\,\sigma^2}{(1+\theta_1^2+\theta_2^2)\,\sigma^2} = \frac{(\theta_1+\theta_1\theta_2)}{(1+\theta_1^2+\theta_2^2)}$
- $\tau_2 = \frac{\gamma_s}{\gamma_0} = \frac{(\theta_2)}{(1+\theta_1^2+\theta_2^2)}$

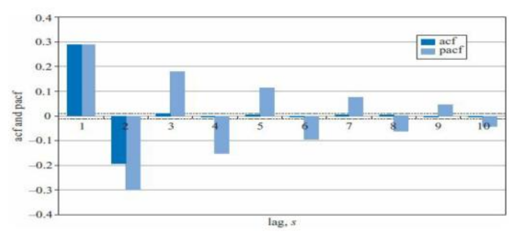
Auto-correlation function (ACF) plot and partial ACF plot. An MA process has

- Number of non-zero points of acf = MA order
- A geometrically decaying pacf.

$y_t = -0.5u_{t-1} + u_t$



$y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$



Example MA (2) Processes: Auto-Correlation

- $\tau_0 = \frac{\gamma_0}{\gamma_0} = 1$
- $\tau_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\theta_1+\theta_1\theta_2)\ \sigma^2}{(1+\theta_1^2+\theta_2^2)\ \sigma^2} = \frac{(\theta_1+\theta_1\theta_2)}{(1+\theta_1^2+\theta_2^2)}$
- $\tau_2 = \frac{\gamma_s}{\gamma_0} = \frac{(\theta_2)}{(1+\theta_1^2+\theta_2^2)}$

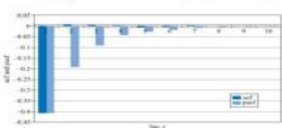
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

We also noted because this was MA2 process any higher order auto correlations and covariances would be zero because the process is restricted to second order lag of mu DC. In this video we will try to understand the auto correlation function and partial auto correlation function that is ACF-PACF for an MA process and differentiate it with AR process. To begin with recall in the previous video we said that MA order MA process has an auto correlation function or ACF which is exactly the number of auto correlations or the number of white noise terms the order of MA process which is same as order of MA process. So, the number of auto correlation terms τ wise will be significant which is equal to exactly the MA order.

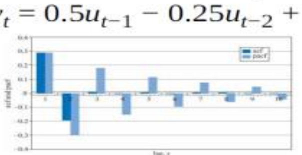
ACF and PACF Plots: MA Process

- Auto-correlation function (ACF) plot and partial ACF plot
- An MA process has
  - Number of non-zero points of acf = MA order
  - A geometrically decaying pacf.

$y_t = -0.5u_{t-1} + u_t$



$y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$



Chris Brooks; Introductory Econometrics for Finance. 4th Edition. Chapter 6

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

In the previous case it was second order process so we had auto correlation functions tau wise significant up to second order. So, that clearly explains that when we will examine the ACF it will be the number of ACF or auto correlations will be same as MA order but what about PACF? Without going into the mathematical proof theory suggests that an MA process an MA process can be transformed into infinite order AR process. I repeat an MA process can be transformed into an infinite order AR process and therefore when it is examined on the PACF plot when examined on the PACF plot there will be infinite number of correlations for an MA process. And therefore when we examine ACF for the MA process for example look at this MA process which is a simple first order MA process with the coefficient as -0.5 for the white noise μ_{t-1}. So, if you look at ACF which is in dark blue it is cutting off at the first level itself only there is one bar which indicates that this is a first order MA one process however if you look at PACF their magnitude is exponentially declining.

So, the magnitude of PACF is exponentially declining through a reasonable number of large terms so it is like exponentially declining. Similarly if you look at this MA2 process where one term is positive and other is -0.5 and -0.5 we look at the dark blue as two specific the ACF has specific two bars while again the magnitude of PACF is exponentially declining in terms of graph positive negative which reflects this kind of structure positive negative but the in terms of magnitude the magnitude is exponentially declining. So, this is how we interpret the ACF and PACF plots for MA process. To summarize if we contrast these PACF PACF with the AR process recall for AR process the ACF diagram was exponentially declining while PACF was exactly cutting off at the number of bars same as AR order of the process ARQ then order Q.

In contrast for MA process the number of ACF bars are exactly limited to MA order so if it is a MA P order process the number of bars in a ACF diagram will be P. However, if you look at the PACF plots they will be exponentially declining in terms of magnitude because like I said an MA process can be converted into infinite AR process and PACF will represent that characteristic that the number of lags will be exponentially declining that infinite AR kind of process. Obviously that infinite AR process has coefficients that are declining as we go ahead and pass so this declining nature will be captured in the PACF that infinite AR process will be captured with these PACF diagrams. So, this is how we differentiate and contrast between AR and MA process on ACF and PACF plots.

To summarize we noted that time series models offer an extremely simple and useful class of models. In particular we noted that time series processes are effectively modeled with ARMA ARMA processes. However, any time series modeling requires a stationary data that is a constant mean, constant variance and auto covariance structure that is not changing across time. We also noted that a non-stationary property has a very undesirable property that its estimation results in spurious relationships that do not exist. Then we introduced autoregressive or AR processes which are a very important class or component of ARMA models.

In the AR process a time series is expected to be dependent on one or more of its own pass values. An ARP process is said to have an order P that is the series is dependent on its own pass P values. The strength of this relationship is defined by the autocorrelation coefficients of the process. Moreover, if the series is stationary then these autocorrelation coefficients have a declining structure that is they tend towards zero as we go back in time. This is a desirable property as it indicates that the impact of distant terms is less in magnitude. A stationary AR process can be identified with the help of one exponentially declining ACF plot and two a PACF plot that has non-zero bars up till the order of AR process and no bars that is zero thereafter.

An MA process is a linear combination of current and past white noise error terms. An MAQ process is said to have an order P if it is dependent on its P historical values. An MA process is said to be invertible if the coefficients of white noise terms have a declining structure as we move further down in past. Also, then it can be converted to an infinite AR process as well. A white noise process has no discernible structure and MA process can be further identified with an exponentially declining PACF diagram and an ACF plot that has a non-zero bars up till the order of MA process and no bars that is zero thereafter. Thank you.