Advanced Financial Instruments for Sustainable Business and Decentralized Markets

Prof. Abhinava Tripathi

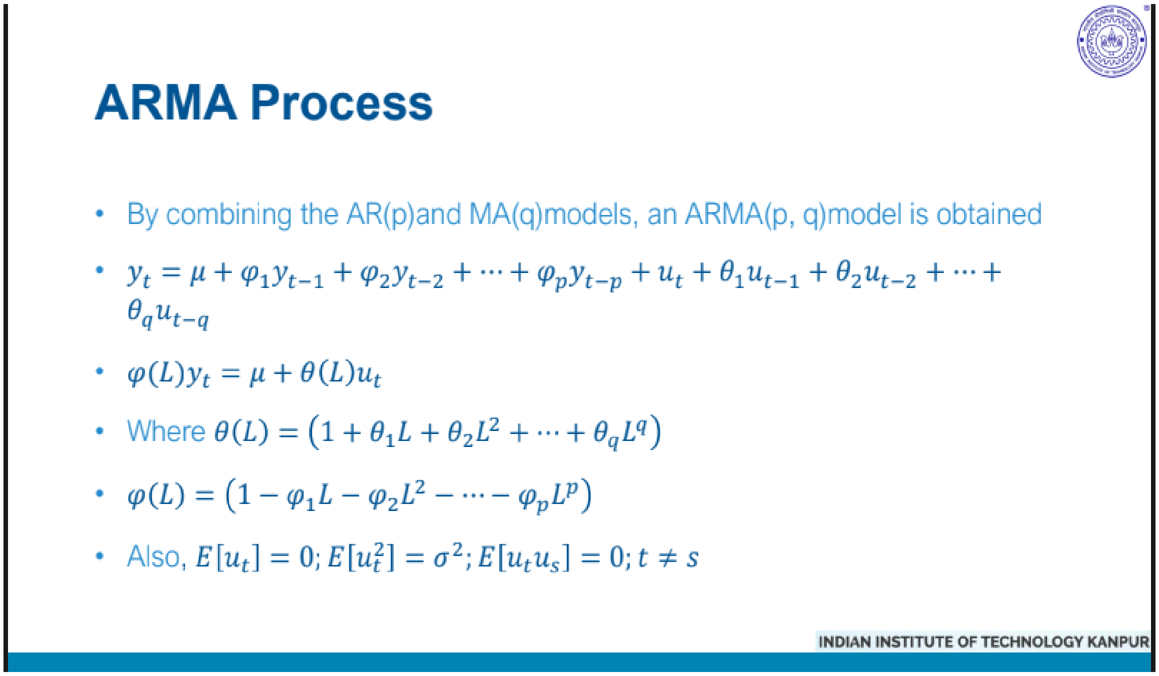
Department of Management Sciences

Indian Institute of Technology, Kanpur

Lecture 22

Week 7

In this video, we discuss ARMA process which is simply a combination of AR and MA processes. We also discuss how to identify the ARMA process with the help of ACF and PACF diagrams. Next we discuss in a step by step manner how to build an ARMA model for forecasting. We also discuss the prediction process with time series models and compare it with the prediction process with structural models.



ARMA Process

- By combining the AR(p)and MA(q)models, an ARMA(p, q)model is obtained
- $y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$
- $\varphi(L) y_t = \mu + \theta(L) u_t$
- Where $\theta(L) = \left(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\right)$
- $\varphi(L) = \left(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p\right)$
- Also, $E[u_t] = 0; E[u_t^2] = \sigma^2; E[u_t u_s] = 0; t \neq s$

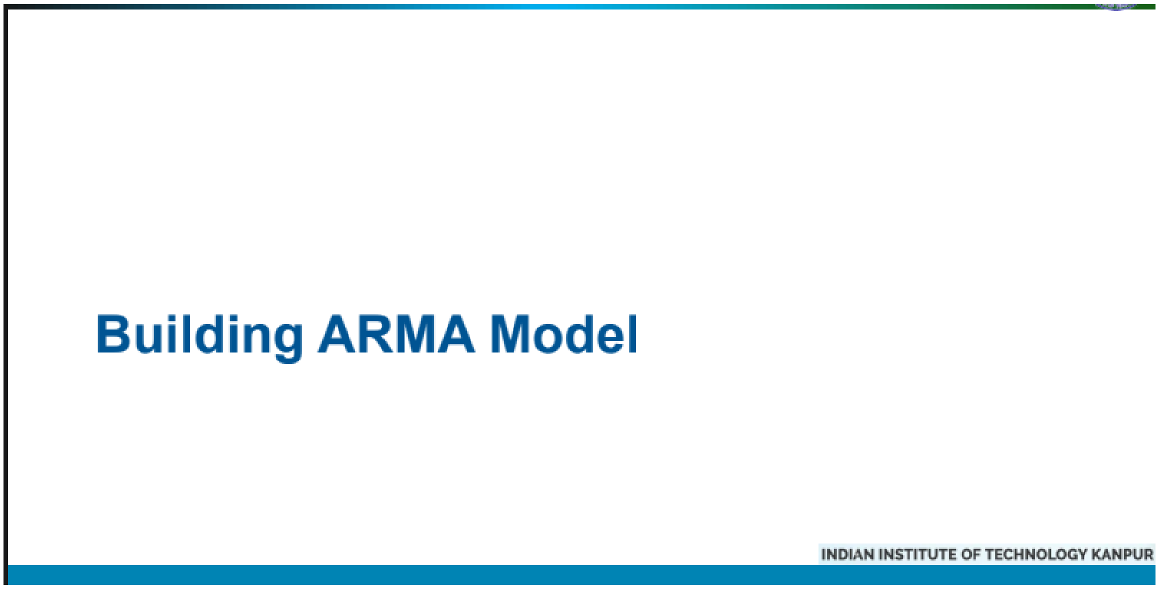
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

We also discuss how to determine the accuracy of out of sample forecasts. In this video, we will introduce ARMA which is autoregressive AR and moving average which is MA. So in this video, we will introduce ARMA process which is a very important component of ARMA models, ARMA class of models. Just to make an distinction here, I here is nothing but the integration of the process. So depending upon the levels of non-stationarity and we will define it in subsequent videos depending upon the level of non-stationarity, the order will be integrated of that order. And other than that, the most important part of that ARMA is AR and MA which we are going to discuss in this video. To put it succinctly, an MA process is nothing, but a combination of AR and MA process of different orders say (p, q). So, if AR process is of order p and MA process is of order q and we combine them, we get an ARMA process of order (p, q).

By combining the AR(p)and MA(q)models, an ARMA (p, q) model is obtained:

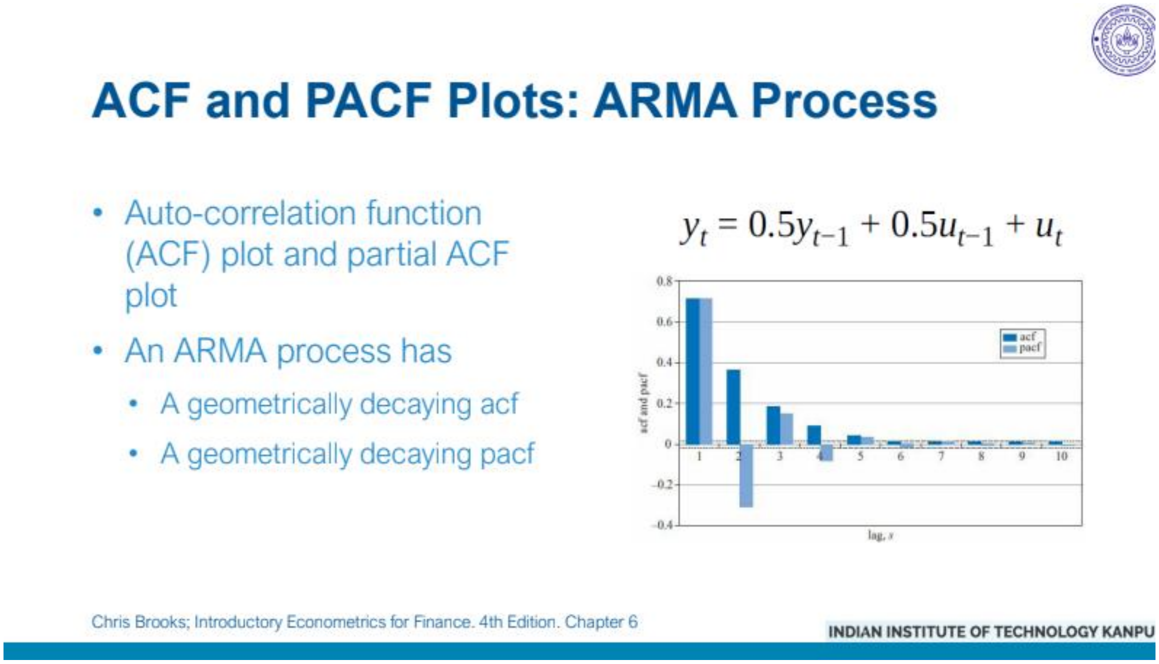
- $y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$
- $\varphi(L) y_t = \mu + \theta(L) u_t$
- Where $\theta(L) = \left(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\right)$
- $\varphi(L) = \left(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p\right)$
- Also, $E[u_t] = 0; E[u_t^2] = \sigma^2; E[u_t u_s] = 0; t \neq s$

So, this is our ARMA process. Now as we have already seen, we can put both these AR and MA terms in a more compact notation like this where $\varphi(L)$ is nothing but $1 - \varphi_1 L - \varphi_2 L^2$ and so on, the algebraic polynomial. Similarly, the theta L or the MA process can be written as $\theta(L) = (1 + \theta_1 L + \theta_2 L^2 + \cdots)$ and so on, which is the algebraic notation. Also, we said that u_t here are white noise processes and these u_t, the white noise processes that these u_t are can be defined with the expected value of 0, their variance which is expected value of u_t^2 as σ^2 and their auto covariance as 0 for all the times which are not same as s for all the t's that are not same as s. The auto covariance is 0 because white noise is a process with more discernible structure.



Building ARMA Model

INDIAN INSTITUTE OF TECHNOLOGY KANPUR



ACF and PACF Plots: ARMA Process

- Auto-correlation function (ACF) plot and partial ACF plot
- An ARMA process has
  - A geometrically decaying acf
  - A geometrically decaying pacf

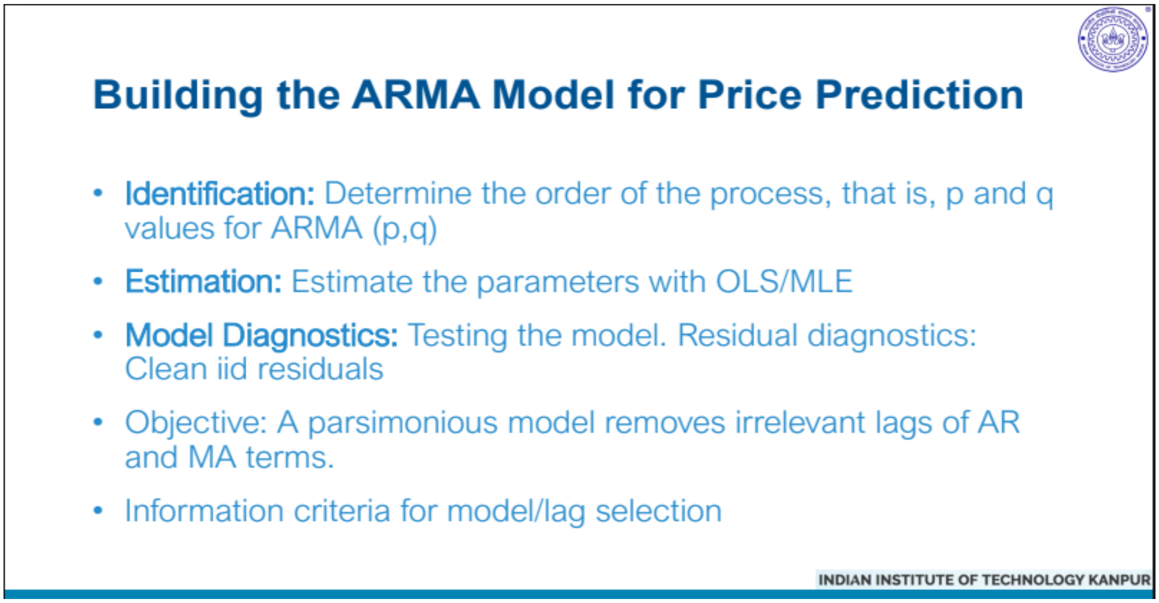
$$y_t = 0.5y_{t-1} + 0.5u_{t-1} + u_t$$

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

INDIAN INSTITUTE OF TECHNOLOGY KANPU

What about the ACF and PACF plots? Interestingly being combination of AR and MA process, the ACF and PACF plots are sort of combination of AR and MA process taken separately. For example, if we put them together, both the ACF and PACF plots are exponentially declining in magnitude irrespective of their order. For example, look at this ARMA process of order (1,1), ARMA (1,1) because here AR term is order of 1 and MA term is also order 1. So this ARMA (1,1) process and it has exponentially declining structure, which is sort of combination of AR and MA itself. If you recall, AR had exponentially declining ACF while single while cutoff of PACF same as the order of AR process. For MA process, it was exponentially declining PACF while ACF was cutting off at exactly the same as number of order of MA process. If we combine them both for AR and MA, if we combine these terms, we get both ACF and PACF which is exponentially declining for both ACF and PACF and that is how you can exactly identify and differentiate between AR, MA and PAC ARMA model on ACF PACF diagrams. To summarize, ARMA processes are nothing but a combination of AR and MA processes and part of a very wider class of ARIMA models. I in ARIMA would show the level of integration or sort of non-stationarity how, what is the characteristic of non-stationarity depending upon the order of non-stationarity whether it is level 1, level 2, the I term would be there. But in finance and economic series, mostly the I part is modeled separately and then the remaining part is modeled as ARIMA model.

So for example, if I process I2, first we will try to remove the non-stationarity of level 1 or 2 whatever it is and once you model out non-stationarity, you will focus on ARIMA part of it in modeling through an expression like this. And lastly, between AR, MA and ARIMA process, you can differentiate between them using the ACF and PACF plots visually. For ARMA process, both the ACF and PACF are geometrically declined which is different from ACF and PACF for AR and MA processes separately. In this video, we will introduce to building and estimating ARMA AR and MA class of models. Starting with ARMA model, the first step is identification of the ARMA process that means understanding what the order of AR and MA terms p and q is, order of TR process and order of q process.



Building the ARMA Model for Price Prediction

- **Identification:** Determine the order of the process, that is, p and q values for ARMA (p,q)
- **Estimation:** Estimate the parameters with OLS/MLE
- **Model Diagnostics:** Testing the model. Residual diagnostics: Clean iid residuals
- Objective: A parsimonious model removes irrelevant lags of AR and MA terms.
- Information criteria for model/lag selection

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Once you understand the structure and order of the model, then second step is estimation of these parameters. For example, if it is ARMA (p,q) model, you have parameters that is coefficients corresponding to AR and MA process that you will try to estimate with process such as OLS and MLE. OLS is ordinary least square estimation and MLE is maximum likelihood estimate depending upon the context and nature of model. For example, OLS is more suitable to linear modeling while MLE approach can be used for non-linear as well as linear. Then we come to as last step as model diagnostics that is you test the model.

Generally, you try to test the residual of the model that residual should be generally the assumption with residuals is that they are white noise term. So, they should have no autocorrelation structure that is one. Often the model is also tested for overfitting. For example, you start with an overfitted model and then you test the parameters whether they are significant or not. So, you keep on eliminating insignificant parameter to arrive at a very parsimonious model.

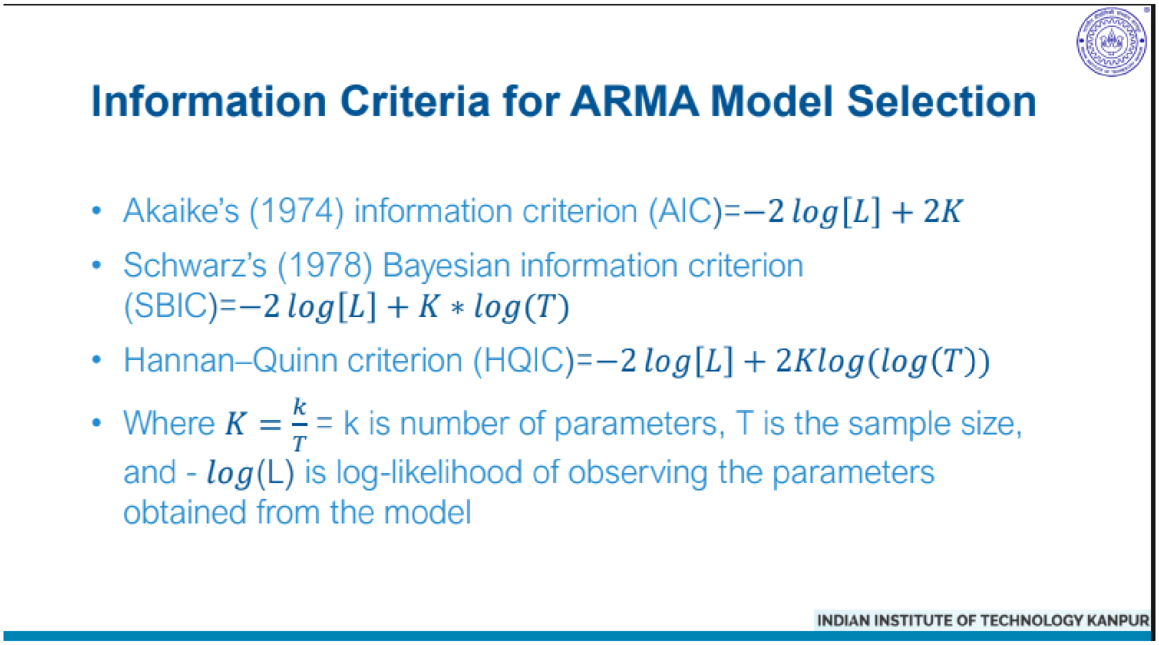
The idea behind using a parsimonious model is to choose as less number of AR and MA terms as possible. A parsimonious model has two very good properties. First and foremost, it consumes less degrees of freedom which means observations or data observed resulting in lower standard errors. So, if you have a parsimonious model relative to a bulky and more complex model, it will consume less degrees of freedom and result in lower standard of errors that means good and high power of test. The second step is second part is overfitting.

So, a model which is bulky and complex and not less not so much parsimonious may have a problem of overfitting that means it may fit two data specific properties which we often refer to as noise also in physics parsons. While a more parsimonious model has less tendency of overfitting and it may rather fit two important properties of data which is often referred to as signal in physics parlance and which means the implication is that a very bulky model may give very high and good fit with a given sample on which it was trained while in out of sample a new data it may perform poorly. In contrast, a parsimonious model may be relatively not as efficient in a given set of data in which it was model but its relative performance on out of sample data would be extremely good. How do we choose between two computing models? So there we use something called information criteria for selecting different lags or structure of ARMA process that is (p,q) order ARMA orders that is (p,q) terms. We have very three famous:

- Akaike's (1974) information criterion (AIC)=$-2\, log[L] + 2K.$ Schwarz's (1978) Bayesian information criterion (SBIC)=$-2\, log[L] + K * log(T)$
- Hannan–Quinn criterion (HQIC)=$-2\, log[L] + 2Klog(log(T)),$ Where $K = \frac{k}{T}$ = k is number of parameters, T is the sample size, and - log(L) is log-likelihood of observing the parameters obtained from the model.

These criteria are nothing but a combination of two terms. The formula is different but essentially they are combination of two terms sort of trade of or two competing parts. First part, this part which is $log[L]$ which is the log likelihood function sort of log likelihood function which is computed based on the residual errors from the model. As you keep on increasing the parameters there are two competing effects at play. This residual these residual errors will come down.

So as you keep on adding more and more parameters this first term will go down as residual errors will go down. The second term which is k increases with parameters, so number of parameters. So the two competing effects residual, this term represents residual for example here this term represents the decrease in decline in residuals with addition of new variables while the second term, this term captures increase in number of parameters. So you try to choose AIC value which is lower k here is$= \frac{k}{T}$ where k is number of parameters and t is the sample size. $log[L]$ like I said is the log likelihood of observing the parameters essentially it is computed from the residual squares.



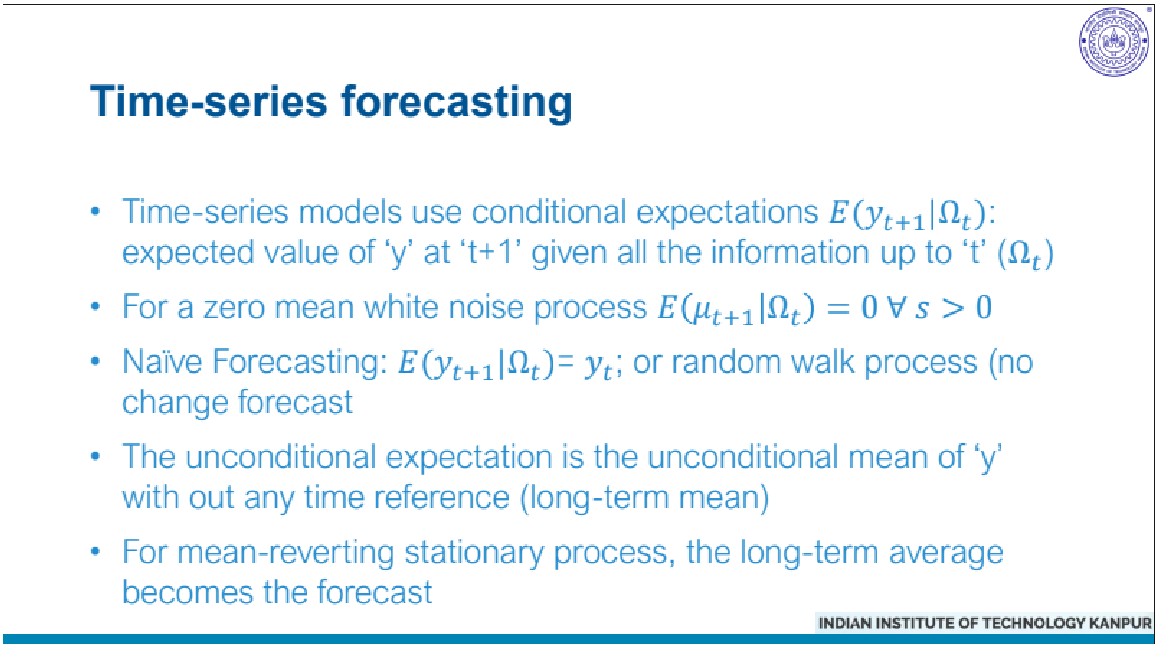
Information Criteria for ARMA Model Selection

- Akaike's (1974) information criterion (AIC)=$-2\,log[L] + 2K$
- Schwarz's (1978) Bayesian information criterion (SBIC)=$-2\,log[L] + K * log(T)$
- Hannan–Quinn criterion (HQIC)=$-2\,log[L] + 2Klog(log(T))$
- Where $K = \frac{k}{T}$ = k is number of parameters, T is the sample size, and - log(L) is log-likelihood of observing the parameters obtained from the model

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

So here these two competing effects are at play and they their net which is the AIC information criteria they are used to compare the models with different structure that is different order of the process, ARIMA-PQ process. So to summarize in this video we discussed how to build estimate and test an ARIMA class of model and while deciding the structure of model how do we select the lags of AR and ME terms using the information criteria AIC, BIC and HQIC criteria. In this video we will introduce forecasting with time series models. We will also discuss and compare them with structural models of forecasting and prediction. To begin with understanding the time series forecasting process let us understand this notation that is conditional expectations $E(y_{t+1}|\Omega_t)$: expected value of 'y' at 't+1' given all the information up to 't' (Ω_t)

So this notation $E(y_{t+1}|\Omega_t)$ (given information Ω_t is for information up till time t it is called conditional expectation of $E(y_{t+1})$ given information available Ω_t up till time t this is this notation. Similarly, we can write it for time t+2, t+3 and so on. Now let us think of a zero mean process like white noise process μ_t its conditional expectation at time t +1 given time t is said to be 0 for all s greater than 0. So for all future periods it is forecast the expected value of its forecast is 0. There are two very simple methods of forecasting one is called naive no change forecasting.



Time-series forecasting

- Time-series models use conditional expectations $E(y_{t+1}|\Omega_t)$: expected value of 'y' at 't+1' given all the information up to 't' (Ω_t)
- For a zero mean white noise process $E(\mu_{t+1}|\Omega_t) = 0 \; \forall \; s > 0$
- Naïve Forecasting: $E(y_{t+1}|\Omega_t)$= y_t; or random walk process (no change forecast
- The unconditional expectation is the unconditional mean of 'y' with out any time reference (long-term mean)
- For mean-reverting stationary process, the long-term average becomes the forecast

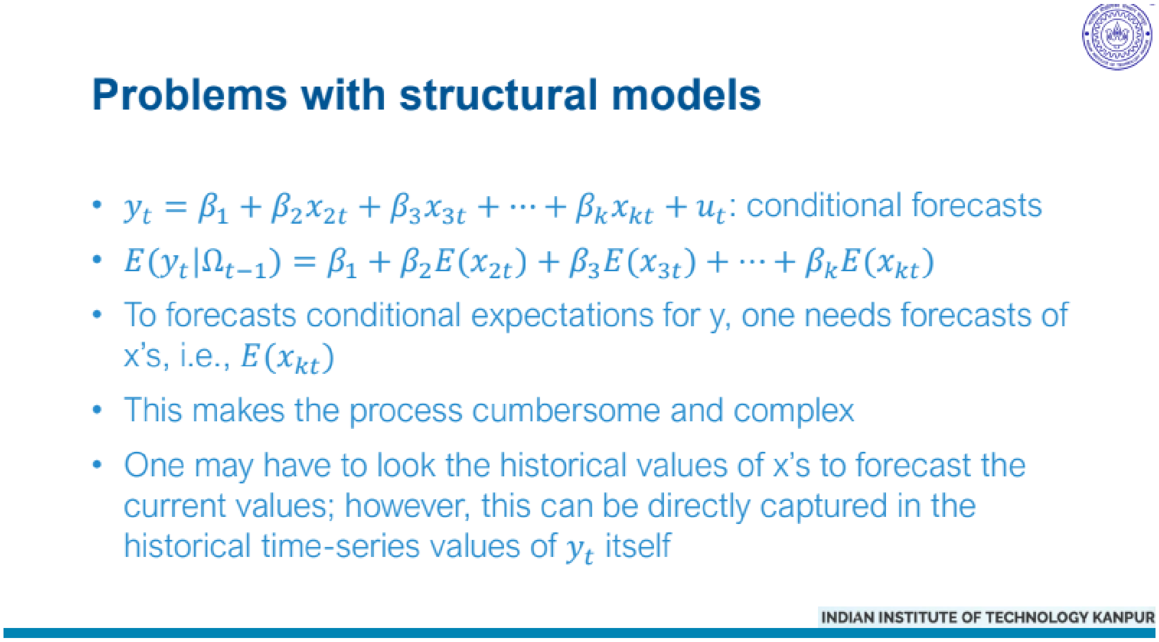
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

This naive forecasting method suggests that forecast of a series y at time t+1 given it information time t is nothing but its previous value y_t at time t. This is also more applicable in the random walk process where the process is purely random and therefore the best forecast of tomorrow is whatever the value today. So therefore the conditional expectation of the series y at time t +1 given that all the information and values known at time t is nothing but the value at time t itself which is y_t. This is also called no change forecast because we assume that future value the best estimate of future value is today value itself. There is another set of processes which are mean reverting which are expected to revert to their long term mean and therefore in that case the unconditional expectation is taken and that unconditional expectation of mean of y is taken as forecast sort of long term mean.

So in this case for a process which is mean reverting which is known to revert towards some kind of long term mean the unconditional long term mean. What is unconditional that we have taken the mean over a long horizon not specific to a particular information but over a long horizon that unconditional mean is taken. It is also in other words this is some kind of long term average that is used to forecast at any time t. Now this long term forecast remain valid for whether you are looking at t +1 or t+2 or so on and therefore this is unconditional. This is unconditional at what time you are looking is simply the unconditional forecast.

So we simply write expected value of y_{t+1} is equal to that long term mean or expected value of y_{t+2} is equal to that long term mean. Now notice there is no conditional this conditional omega t term is not there which means that this is unconditional. Now let us compare these time series models and forecasting with time series models with structural models. A structural model would look something like this $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} +$

$\cdots + \beta_k x_{kt} + u_t$: conditional forecasts and then there is this error term this you can think of as a white noise error. When you are forecasting or predicting with this kind of structural model which is obviously driven by some kind of theoretical underpinning the conditional expectation of y_t at time t with given information at t-1 would appear something like this.

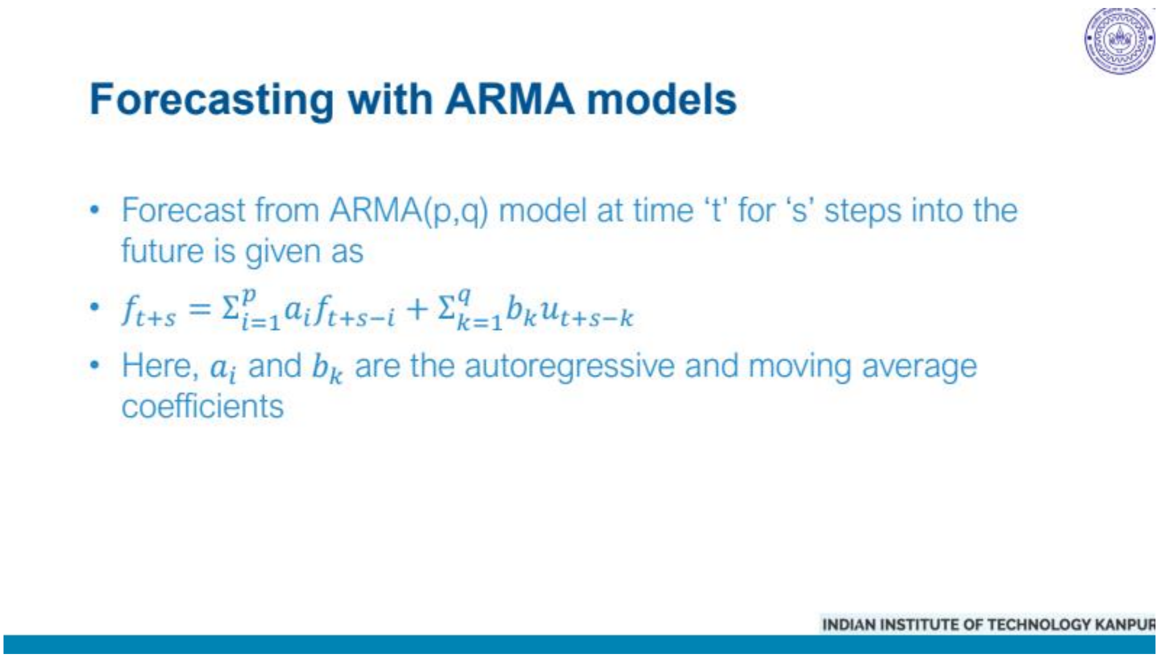


Problems with structural models

- $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t$: conditional forecasts
- $E(y_t|\Omega_{t-1}) = \beta_1 + \beta_2 E(x_{2t}) + \beta_3 E(x_{3t}) + \cdots + \beta_k E(x_{kt})$
- To forecasts conditional expectations for y, one needs forecasts of x's, i.e., $E(x_{kt})$
- This makes the process cumbersome and complex
- One may have to look the historical values of x's to forecast the current values; however, this can be directly captured in the historical time-series values of y_t itself

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

However this kind of forecasting and because these betas are constant they are sort of population parameters so they are taken out. So inside expectations operator you have taken expectations on both sides we have $x_{2t}, x_{3t},\ x_{4t}$ these are the values in the expectation. Now in order to conduct the forecasting or prediction or what we call as conditional expectation of y_t we need forecast of $x_{2t}, x_{3t},\ x_{4t}$ which are time series processes themselves and that makes this prediction with the help of structural model extremely complex process because now you have to rely on all those k variables and do the modeling and prediction for the variables themselves. This is a difficult part and therefore one may have to examine the historical values of all these x_s maybe find their some kind of mean or expected values first you have to find the expected values and then in turn you have to forecast the expected value of y_t . However rather than going through this cumbersome exercise if you are relying on simple time series model of y_t by modeling through its own values like y_{t-1}, y_{t-2}, and the white noise information shocks μ_t that have arrived currently and previously you can very easily model y_t without considering these values the idea here is that this structural relationship or any other kind of structural relationship will remain anyway valid and therefore this relationship will remain also valid historically that means for y_{t-1}, y_{t-2}, and therefore when we are modeling y_t with y_{t-1}, y_{t-2}, and so on essentially we are factoring these relationship that have appeared or captured their information historically.



Forecasting with ARMA models

- Forecast from ARMA(p,q) model at time ‘t’ for ‘s’ steps into the future is given as
- $f_{t+s} = \Sigma_{i=1}^{p} a_i f_{t+s-i} + \Sigma_{k=1}^{q} b_k u_{t+s-k}$
- Here, a_i and b_k are the autoregressive and moving average coefficients

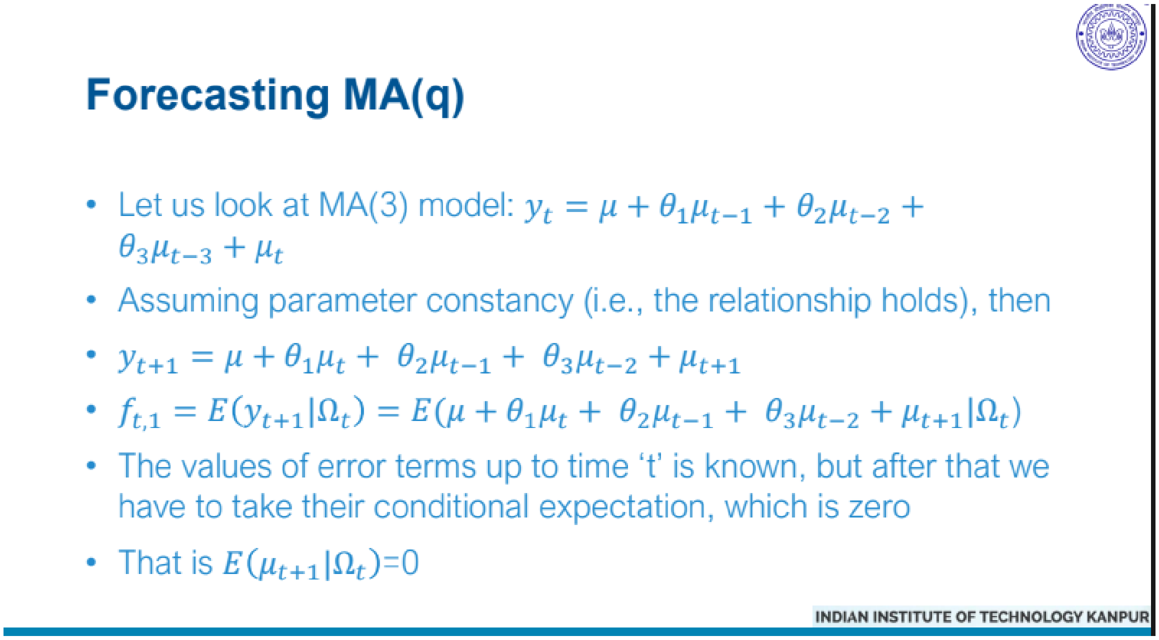
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

So, this relationship the information pertaining to such structural relationship is in historical terms is already captured in y_{t-1}, y_{t-2}, and therefore when I am modeling y_t with its own historical values and white noise error terms such as this through some kind of ARMA process I am essentially without even going into this structural model I am factoring this. And therefore this kind of modeling is extremely powerful that it saves lot of time and effort in modeling the structural model and able to capture the information content of this structural process itself. So, to summarize this video we have understood what is time series forecasting, what are the elements of time series forecasting and how it is a better way when we are doing some kind of future forecast or prediction as compared to structural models it while it takes the information content of those structural models still it saves or avoids lot of problems in terms of information collection and get information gathering and building cumbersome models as done with the structural models. In a series of next few videos we will try to understand forecasting process with ARMA class of models. In this particular video we will understand forecasting with MA model.

- Forecast from ARMA(p,q) model at time ‘t’ for ‘s’ steps into the future is given as
- $f_{t+s} = \Sigma_{i=1}^{p} a_i f_{t+s-i} + \Sigma_{k=1}^{q} b_k u_{t+s-k}$
- Here, a_i and b_k are the autoregressive and moving average coefficients.

Let us examine this simple ARMA (p,q) model at time t with s steps into future. The forecast would appear something like this notation is f_{t+s} representing at time t s steps at forecast. Now it will have two components first AR component that is its own previous terms if the actual value is not available then forecast f will be employed which is the previous values of the series itself and its coefficient $a_i{}'$s which are autoregressive coefficients and then you have MA terms white noise process mu t s previous values and current values of mu t with MA coefficients. So, let us start with forecasting of MA process. Let us take a simple example of MA (3) order process y_t equal to μ this is the constant term which we assume some kind of long term constant value which is known to us from historical experience and then three terms μ_{t-1}, μ_{t-2}, μ_{t-3} and their coefficient θ_1 ,θ_2 θ_3 and the current information or innovation in shock μ_t.

- Let us look at MA(3) model: $y_t = \mu + \theta_1\mu_{t-1} + \theta_2\mu_{t-2} + \theta_3\mu_{t-3} + \mu_t$
- Assuming parameter constancy (i.e., the relationship holds), then
- $y_{t+1} = \mu + \theta_1\mu_t + \ \theta_2\mu_{t-1} + \ \theta_3\mu_{t-2} + \mu_{t+1}$
- $f_{t,1} = E(y_{t+1}|\Omega_t) = E(\mu + \theta_1\mu_t + \ \theta_2\mu_{t-1} + \ \theta_3\mu_{t-2} + \mu_{t+1}|\Omega_t)$
- The values of error terms up to time ‘t’ is known, but after that we have to take their conditional expectation, which is zero
- That is $E(\mu_{t+1}|\Omega_t)$=0



Forecasting MA(q)

- Let us look at MA(3) model: $y_t = \mu + \theta_1\mu_{t-1} + \theta_2\mu_{t-2} + \theta_3\mu_{t-3} + \mu_t$
- Assuming parameter constancy (i.e., the relationship holds), then
- $y_{t+1} = \mu + \theta_1\mu_t + \ \theta_2\mu_{t-1} + \ \theta_3\mu_{t-2} + \mu_{t+1}$
- $f_{t,1} = E(y_{t+1}|\Omega_t) = E(\mu + \theta_1\mu_t + \ \theta_2\mu_{t-1} + \ \theta_3\mu_{t-2} + \mu_{t+1}|\Omega_t)$
- The values of error terms up to time ‘t’ is known, but after that we have to take their conditional expectation, which is zero
- That is $E(\mu_{t+1}|\Omega_t)$=0

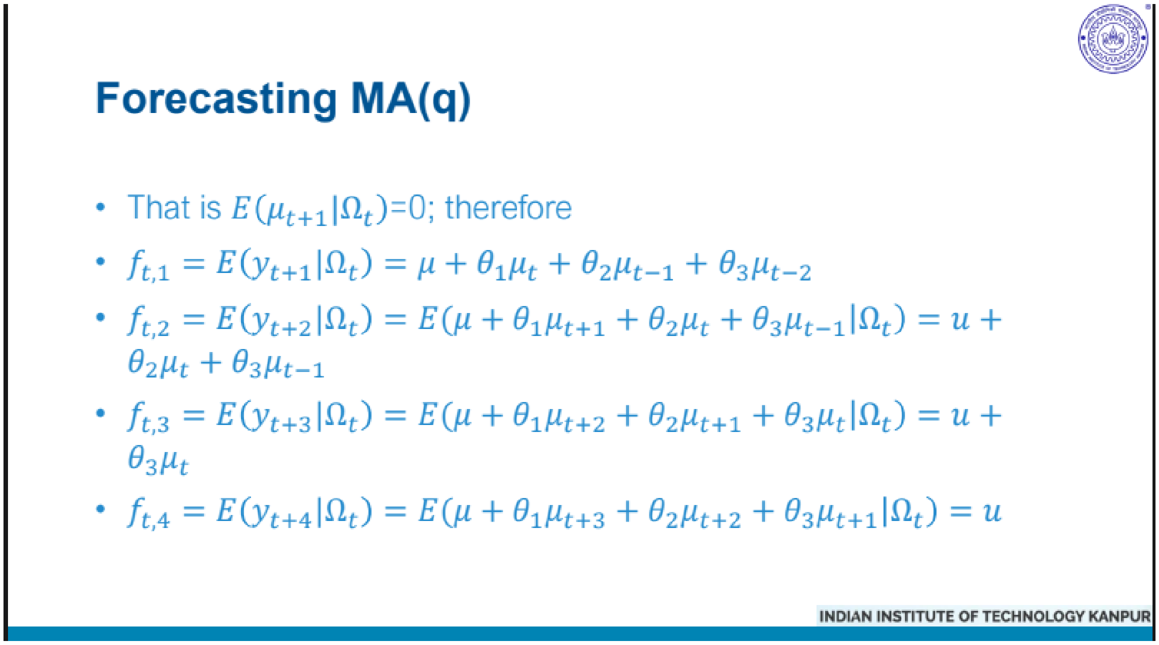
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

We will assume parameter constancy that this θ_1, θ_2 and $\theta_t$will remain unchanged as we move ahead into future. So, if this relationship holds the one step ahead forecast y_{t+1} will be obtained by simply just adding 1 in the subscript so the model become $\mu + \theta_1\mu_t + \ \theta_2\mu_{t-1} + \ \theta_3\mu_{t-2} + \mu_{t+1}$. In this model, we can obtain the forecast by simply taking expectation which is the forecast essentially nothing but conditional expectation of

y_{t+1} using information available up till t that is conditional up till σ_t. So, we will apply the expectation operator and the constant terms such as μ and θ_1, θ_2 θ_3 will remain unchanged and will be taken out. However, one important point till now we have been saying that the conditional expectation of error term as zero because these are white noise processes.

However, when we are standing ahead in today, all the information or these information shocks what we are if you recall we said these white noise processes are nothing but being white noise processes are used to model the information shock that has arrived in the process. So, whatever information shocks have arrived till date they are already known. So, their conditional expectation is not zero will not take it as zero, but we use the actual values. So, for example, up till information t up till time t we have all the information available. So, for that we will use the actual observed values of $\mu_t + \mu_{t-1} + \mu_{t-2}$.

However, for μ_{t+1} because this is not known its expectation obviously will be zero. So, expectation of this μ_{t+1} will be zero, but the conditional expectation of these terms $\mu_t + \mu_{t-1} + \mu_{t-2}$ will not be considered their actual values will be taken. So, in this spirit, when we take expectations this μ_{t+1} will become zero but for all the other terms we use the $\mu + \theta_1\mu_t + \theta_2\mu_{t-1} + \theta_3\mu_{t-2}$ with the caveat that these μ_t μ_{t-1} μ_{t-2} are actual values that are already observed standing at time t. So, with this being the case we already have the forecast of $f_{t,1}$ one step ahead in future which is this. Now let us do the forecasting for two step ahead in future which is $f_{t,2}$ which is nothing but conditional expectation of y_{t+2}again using the information of till time t.



Forecasting MA(q)

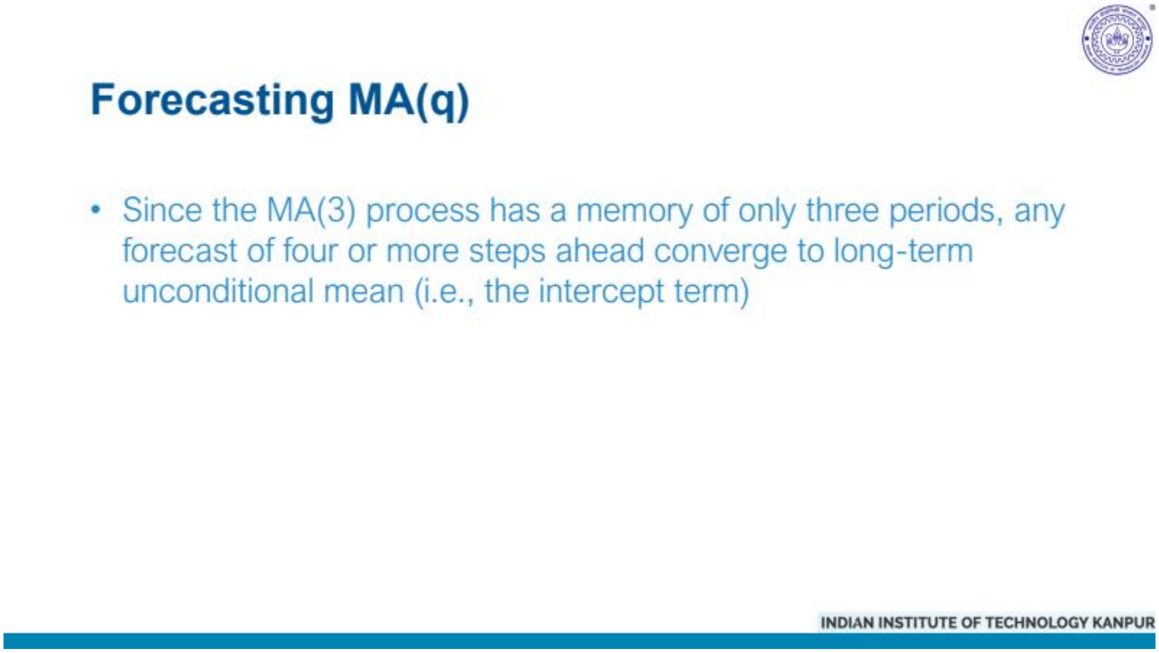
- That is $E(\mu_{t+1}|\Omega_t)$=0; therefore
- $f_{t,1} = E(y_{t+1}|\Omega_t) = \mu + \theta_1\mu_t + \theta_2\mu_{t-1} + \theta_3\mu_{t-2}$
- $f_{t,2} = E(y_{t+2}|\Omega_t) = E(\mu + \theta_1\mu_{t+1} + \theta_2\mu_t + \theta_3\mu_{t-1}|\Omega_t) = u + \theta_2\mu_t + \theta_3\mu_{t-1}$
- $f_{t,3} = E(y_{t+3}|\Omega_t) = E(\mu + \theta_1\mu_{t+2} + \theta_2\mu_{t+1} + \theta_3\mu_t|\Omega_t) = u + \theta_3\mu_t$
- $f_{t,4} = E(y_{t+4}|\Omega_t) = E(\mu + \theta_1\mu_{t+3} + \theta_2\mu_{t+2} + \theta_3\mu_{t+1}|\Omega_t) = u$

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Now, in this case we need to add the subscript here +1. So, for each we will add 1 to subscript and then we recall that when the expectation operator is applied t +1 will be zero but for μ_t μ_{t-1} actual values will be employed and therefore this term will get eliminated.

- That is $E(\mu_{t+1}|\Omega_t)$=0; therefore
- $f_{t,1} = E(y_{t+1}|\Omega_t) = \mu + \theta_1\mu_t + \theta_2\mu_{t-1} + \theta_3\mu_{t-2}$
- $f_{t,2} = E(y_{t+2}|\Omega_t) = E(\mu + \theta_1\mu_{t+1} + \theta_2\mu_t + \theta_3\mu_{t-1}|\Omega_t) = u + \theta_2\mu_t + \theta_3\mu_{t-1}$
- $f_{t,3} = E(y_{t+3}|\Omega_t) = E(\mu + \theta_1\mu_{t+2} + \theta_2\mu_{t+1} + \theta_3\mu_t|\Omega_t) = u + \theta_3\mu_t$
- $f_{t,4} = E(y_{t+4}|\Omega_t) = E(\mu + \theta_1\mu_{t+3} + \theta_2\mu_{t+2} + \theta_3\mu_{t+1}|\Omega_t) = u$

So, this will be zero this will be zero because they are in future only for μ_t the actual value will be employed and therefore we are left with $u + \theta_3\mu_t$. So, this is our forecast for three. Now please recall this is a MA (3) order process order of three. So, only information that is available up till three step ahead will be used after that there are no μ_t's that are actually available.



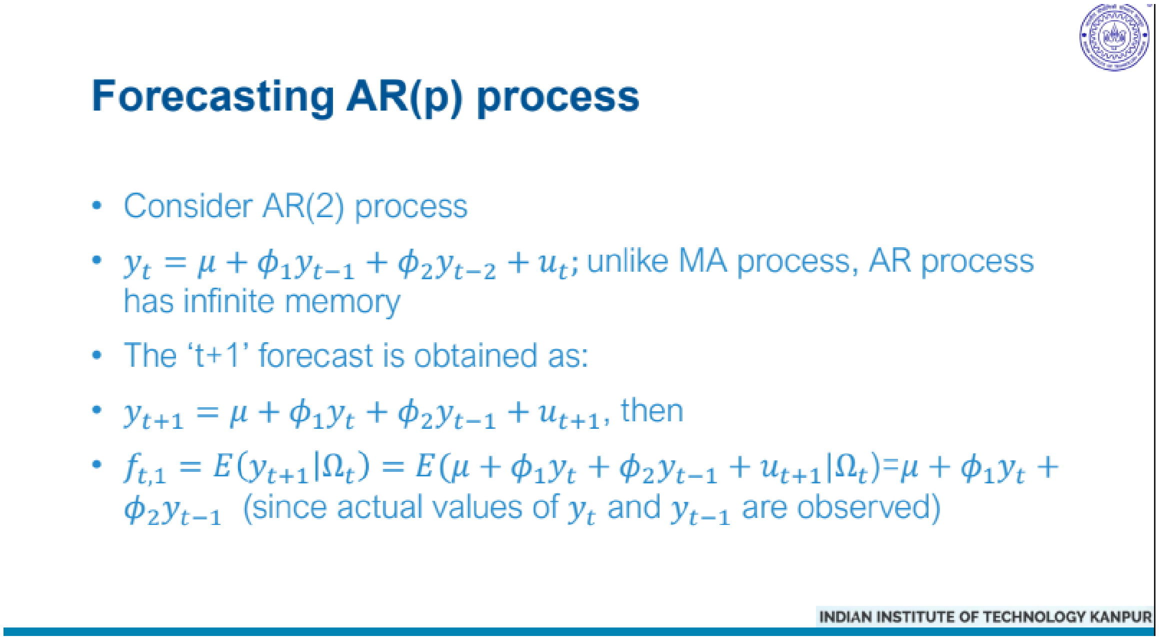
Forecasting MA(q)

- Since the MA(3) process has a memory of only three periods, any forecast of four or more steps ahead converge to long-term unconditional mean (i.e., the intercept term)

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Now in this case, as soon as we add one step ahead, we have t+3, t +2 and t+1 all these white noise μ_t are in future. So their conditional their expectations conditional expectations given Ω_t which is the information available till t will be zero. So, all these terms will be zero and we are only left with μ and I hope you can see where we are going with this all future forecast t,5 t,6 and so on will become μ only which is the constant or some kind of long term mean of this process. As we said earlier for practical implications, we will have this μ as zero. So in that case, this if this μ is taken as zero all these values will become zero and only those values up till three forecast and by only three because this is an MA(3) order process.

Since the MA(3) order process has a memory of only three periods, any forecast of four or more periods will converge to that long term and conditional mean mu or the intercept term. If this is being taken as zero, then this part will become zero only and only up till three period forecast will survive. To summarize in this video, we have discussed how to forecast with MA process. We noted that MA forecast become or converge to some kind of long term conditional mean for any term higher than the order of process. So if it is an MAQ process, any forecast which is more than Q steps, maybe Q plus one or Q plus two will converge to the long term and conditional mean which is the intercept term if it is zero, then that forecast will be zero and only those forecast that are up till Q steps ahead in future will survive for an MAQ order process.



Forecasting AR(p) process

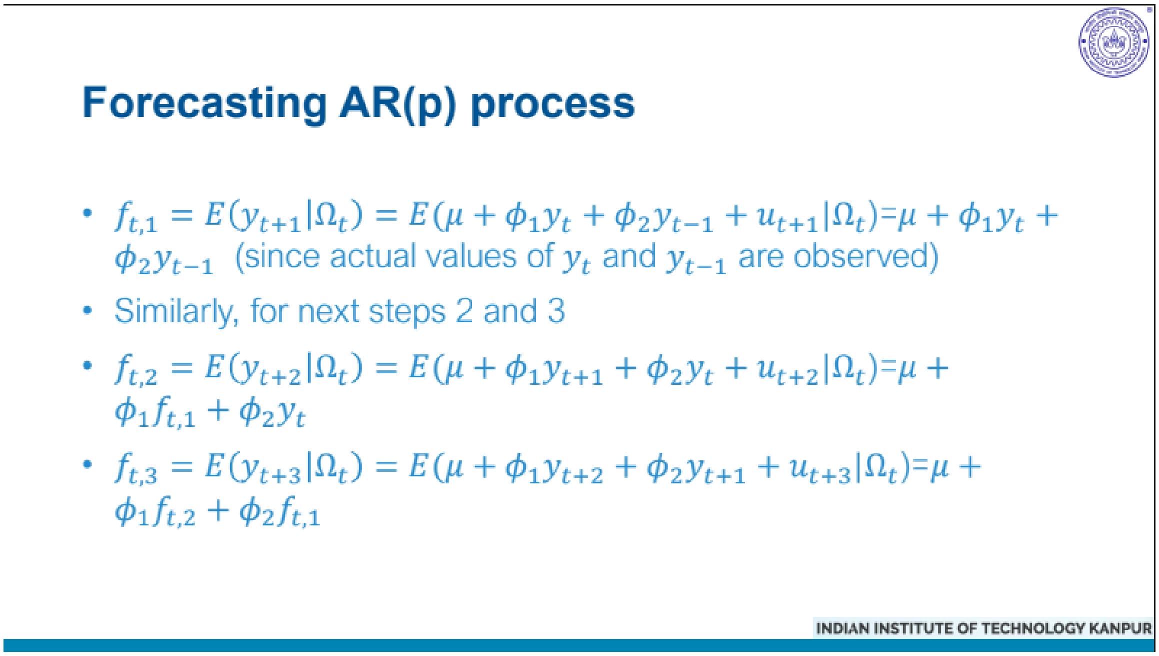
- Consider AR(2) process
- $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$; unlike MA process, AR process has infinite memory
- The 't+1' forecast is obtained as:
- $y_{t+1} = \mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}$, then
- $f_{t,1} = E(y_{t+1}|\Omega_t) = E(\mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}|\Omega_t)$=$\mu + \phi_1 y_t + \phi_2 y_{t-1}$ (since actual values of y_t and y_{t-1} are observed)

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

In this video, we will try to understand forecasting for AR processes. Let us consider a simple AR(2) process. Its structure would be $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$; unlike MA process, AR process has infinite memory. Here, μ which is again the constant term may be taken as zero also, $\phi_1 y_{t-1} + \phi_2 y_{t-2}$ which are the previous values along with the coefficients, autoregressive coefficients ϕ_1 ϕ_2 and the white noise order term. Now, as we will see shortly, very shortly that unlike MA process which was restricted to the number of lags of white noise terms, AR process has an infinite memory.

We will see that. So let us understand what will be the 't+1' forecast here. So forecast at 't+1'would be the constant term + ϕ_1 and we will add for each of these subscripts we will add one. So here we have $\mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}$ which is the white noise error. In this case, when we are computing the forecast for one step ahead and by taking the conditional expectation of y_{t+1} at an information available till time t, please recall that up till time t we have y_t and y_{t-1} already observed. So we will not be taking their expectation but actual values because they can be observed.

- $y_{t+1} = \mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}$, then
- $f_{t,1} = E(y_{t+1}|\Omega_t) = E(\mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}|\Omega_t \ \ ) \ = \ \mu + \phi_1 y_t + \phi_2 y_{t-1}$ (since actual values of y_t and y_{t-1} are observed).



Forecasting AR(p) process

- $f_{t,1} = E(y_{t+1}|\Omega_t) = E(\mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}|\Omega_t)$=$\mu + \phi_1 y_t + \phi_2 y_{t-1}$ (since actual values of y_t and y_{t-1} are observed)
- Similarly, for next steps 2 and 3
- $f_{t,2} = E(y_{t+2}|\Omega_t) = E(\mu + \phi_1 y_{t+1} + \phi_2 y_t + u_{t+2}|\Omega_t)$=$\mu + \phi_1 f_{t,1} + \phi_2 y_t$
- $f_{t,3} = E(y_{t+3}|\Omega_t) = E(\mu + \phi_1 y_{t+2} + \phi_2 y_{t+1} + u_{t+3}|\Omega_t)$=$\mu + \phi_1 f_{t,2} + \phi_2 f_{t,1}$

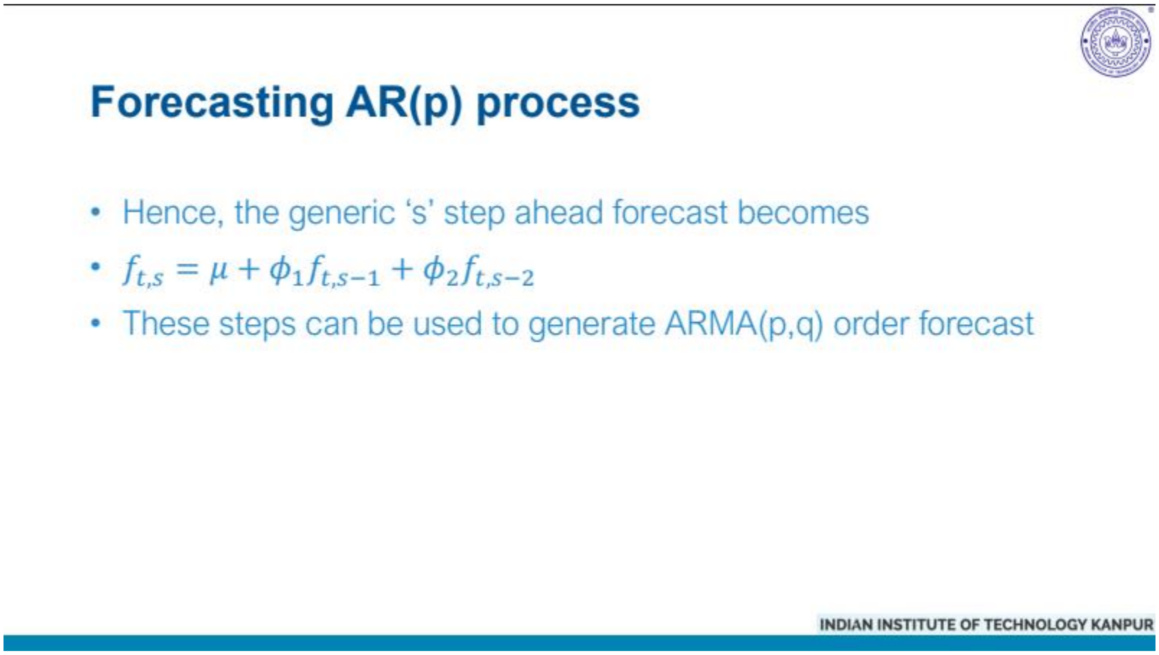
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

- $f_{t,1} = E(y_{t+1}|\Omega_t) = E(\mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}|\Omega_t \ \ ) \ = \ \mu + \phi_1 y_t + \phi_2 y_{t-1}$ (since actual values of y_t and y_{t-1} are observed). Similarly, for next steps 2 and 3
- $f_{t,2} = E(y_{t+2}|\Omega_t) = E(\mu + \phi_1 y_{t+1} + \phi_2 y_t + u_{t+2}|\Omega_t)$=$\mu + \phi_1 f_{t,1} + \phi_2 y_t$
- $f_{t,3} = E(y_{t+3}|\Omega_t) = E(\mu + \phi_1 y_{t+2} + \phi_2 y_{t+1} + u_{t+3}|\Omega_t)$=$\mu + \phi_1 f_{t,2} + \phi_2 f_{t,1}$

So the generic forecast will appear like this which is nothing but $f_{t,s}$, s steps had forecast with a constant term, immediate previous term s-1 and next previous term s-2. Why these two terms precisely because this was an AR two process. Hence, the generic ‘s’ step ahead forecast becomes:

- $f_{t,s} = \mu + \phi_1 f_{t,s-1} + \phi_2 f_{t,s-2}$. These steps can be used to generate ARMA(p,q) order forecast

So that's why we are having these two terms but with these we can go ahead infinitely in future and we'll always get a forecast which is not same as mu or zero as was in the case of MA which means for every n steps in future we can obtain some kind of conditional forecast using previous two values which do not become zero unlike the case was MA. In MA as soon as we are ahead of the MA order of the process which was MQ the forecast will converging to mu which is the conditional mean or zero if in case it was zero. But here every time you observe some kind of forecast depending upon its immediate previous values and that is why we said this AR process has infinite memory. More importantly now that we have forecast for this AR(p) order kind of process and MA(q) as we seen earlier we can combine them and obtain the complete forecast for ARIMA(p,q) process. To summarize this video we have understood how to forecast for AR process.



Forecasting AR(p) process

- Hence, the generic 's' step ahead forecast becomes
- $f_{t,s} = \mu + \phi_1 f_{t,s-1} + \phi_2 f_{t,s-2}$
- These steps can be used to generate ARMA(p,q) order forecast

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

We noted that unlike MA process which were restricted to the order of the process the forecast for MA process, the forecast for AR process have infinite memory that they can be made infinitely long horizons. But in conclusion if you want to forecast for ARIMA (p,q) model you can separately forecast for AR and MA processes and then add them together to get the complete full ARIMA (p,q) order forecast. In this video we will be introduced to some basic terminology related to forecasting models. First we will start with in-sample versus out-of-sample forecasting. For example, let us say we have data available from January 1990 up till December 99.

One approach to develop a model is to view the complete data from January 90 to December 1999 and test its fitness or try to see how fitted values will appear. But this is not a very good approach because generally it is expected that when we are training the data using this data training the model, the model would generally do well on this data itself. So, more in evolved way is to take a certain set of the sample maybe January 1990 to December 1998 call it training data set. Part of the sample on which the model is trained and parameter estimated and a certain set of sample which is left out or hold out sample from January 99 to December 99. These set of observations are not used in as a part of training data set and these observations are used for testing the model.

So, we use certain set and this is up to the searcher how many generally a good ratio is 70, 30 or 80, 20. So, on 70 to 80 percent of the observations the model parameters are trained and once the model is trained we estimate we try to predict on these set of observations and then test the model accuracy which is called out-of-sample forecasting. So, model this was the in-sample estimation and this is out-of-sample forecasting or evaluation of the model period. The next is what is the forecasting requirement? For example, I may want to forecast for one step ahead, somebody may want to forecast two step ahead or three step ahead. Generally, it is well accepted that if you go further and further if you go s steps in future where s is pretty large your forecasting accuracy decreases.

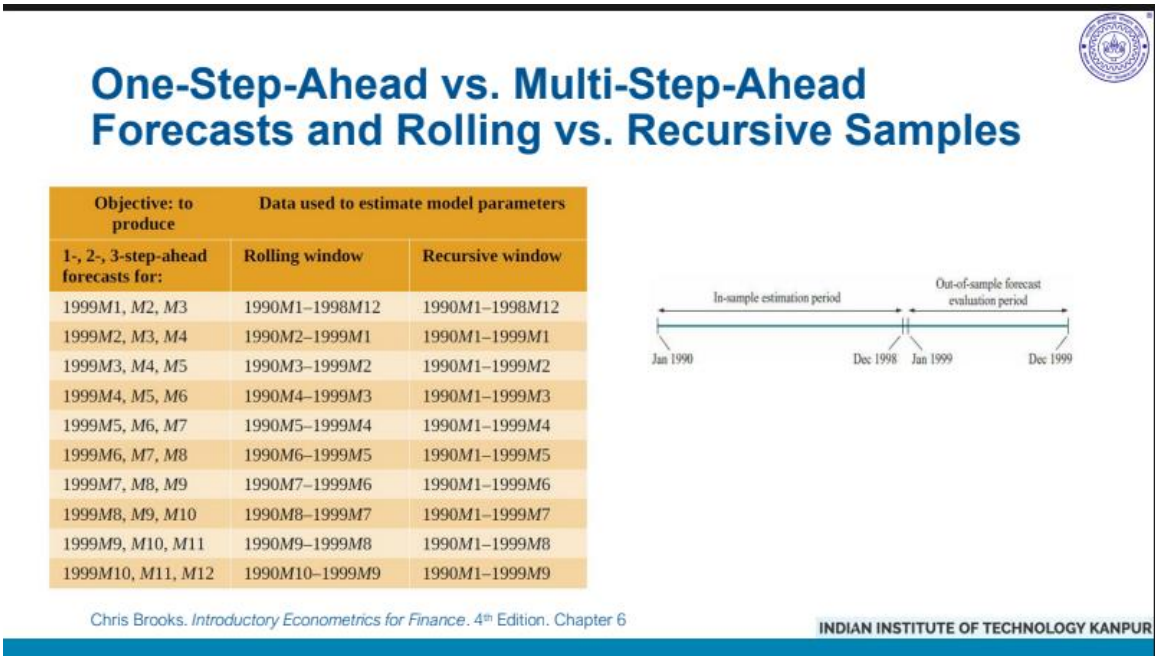
So, as a researcher I should have some idea, some experience about my data that whether one step or two step or three steps in future forecast are more efficient rather than a very lengthy forecast. Third is rolling versus recursive forecasting. Let us understand it. So, let us say I am planning for a one, two or some kind of s step and forecast. A rolling window approach would start with a certain initial set maybe January 1990 to 1998 December and then forecast maybe one, two, three steps in future.

Now, when as a researcher maybe I have found that s steps maybe three steps in future, three steps at forecast is good enough for me. So, when I am going to the next three steps, there are two ways either I can keep my initial window fixed and add the next three steps one t +1, t +2 and t +3 in my current set of observations and then make use of original sample s observations plus these three newly added observations to make the next set of forecast. Another approach would be to rather than keeping the initial, this is called the previous approach was called recursive window because the initial period was kept fixed and as we move ahead in future we keep on adding each observation one by one we keep on adding to our original sample. So, our sample size increases. The second approach is to have a rolling window approach where as we keep adding future we add one observation but at the same time we exclude the last observation.

So, this January 1990 will be excluded when January 99 will be added, then February 99, February will be excluded and February 1999 will be added and as we keep on moving ahead our sample estimation period will remain fixed, that is approximately this length is 180 month that remains fixed as we keep on moving ahead. So, we exclude one period when we add one period, this is called rolling window approach. Both these approaches have a trade-off. For example, with recursive window the benefit is that your sample estimation period in sample period becomes larger but at the same time the previous values that are still not taken away from the sample they may not be in sync with the current market conditions they may be probably more stale and therefore some of the very old values may not be up to the date and may not give me the right picture. While with rolling window we limit our sample size to in this case 180 or a certain number but at the same time as we move ahead we keep on excluding historical values so that our sample on which we are training the model is more in sync with the current market conditions.

So to summarize this video we learned some terminologies related to forecasting. This included one step versus multi-step at forecast. We have in-sample versus out-of-sample prediction and rolling versus recursive method of forecasting. In this video we will try to understand how to determine the accuracy of forecasting. As we noted earlier that often it is important to have model trained on in-sample on a given sample data while use the holdout sample or out-of-sample the second cut-off sample for out-of-sample prediction and then find the accuracy of forecasting on this out-of-sample prediction set.

Let us see and proceed how to do that. Let's say you have built some kind of Arma Arima model and you have as using the method that we have already seen you have forecasted certain values for one two three four and five steps ahead. These are your forecasted values on out-of-sample because this is out-of-sample you already have the sample values available actual sample values to test. The approach to do is compute some kind of error measure. In this particular example we will use mean square error, mean absolute error and root mean square error. In the first step we will compute the difference between forecasted and actual values which is the error but generally there are two ways either you take the square or the absolute value.



One-Step-Ahead vs. Multi-Step-Ahead Forecasts and Rolling vs. Recursive Samples

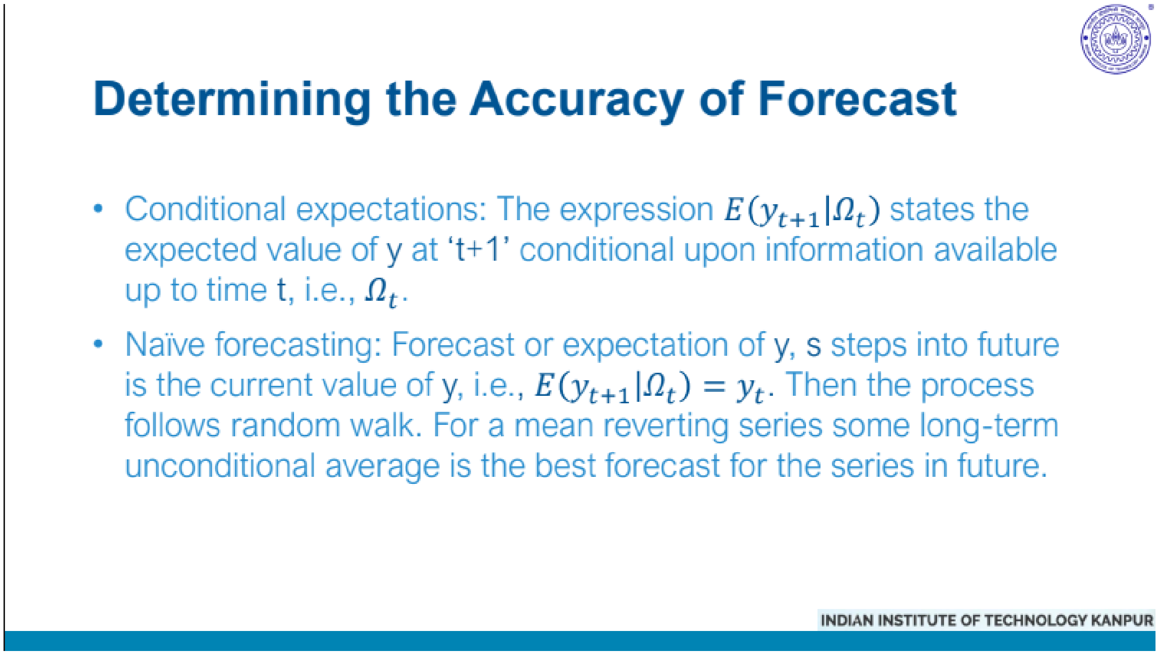
Objective: to produce	Data used to estimate model parameters	
1-, 2-, 3-step-ahead forecasts for:	Rolling window	Recursive window
1999*M*1, *M*2, *M*3	1990*M*1–1998*M*12	1990*M*1–1998*M*12
1999*M*2, *M*3, *M*4	1990*M*2–1999*M*1	1990*M*1–1999*M*1
1999*M*3, *M*4, *M*5	1990*M*3–1999*M*2	1990*M*1–1999*M*2
1999*M*4, *M*5, *M*6	1990*M*4–1999*M*3	1990*M*1–1999*M*3
1999*M*5, *M*6, *M*7	1990*M*5–1999*M*4	1990*M*1–1999*M*4
1999*M*6, *M*7, *M*8	1990*M*6–1999*M*5	1990*M*1–1999*M*5
1999*M*7, *M*8, *M*9	1990*M*7–1999*M*6	1990*M*1–1999*M*6
1999*M*8, *M*9, *M*10	1990*M*8–1999*M*7	1990*M*1–1999*M*7
1999*M*9, *M*10, *M*11	1990*M*9–1999*M*8	1990*M*1–1999*M*8
1999*M*10, *M*11, *M*12	1990*M*10–1999*M*9	1990*M*1–1999*M*9

In-sample estimation period
Out-of-sample forecast evaluation period
Jan 1990
Dec 1998
Jan 1999
Dec 1999

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

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Why we are doing this because if you are using some kind of summative measure summation for positive negative values they will cancel each other out. So large positive negative errors will cancel each other out and my measure will not be good. So generally a good measure is like mean square mean absolute they tend to take either square or absolute values of errors. So once you have this square of errors you take the difference which is your error so your error will be and then you square it you will get those values and then you can sum them up and take the mean. How to do that? So for example these are our squared values these are squared errors this is the squared number once we have the summation of this square I can divide it by five to get what we call as mean square error MSE squares and then divide to compute mean and this is mean square error.



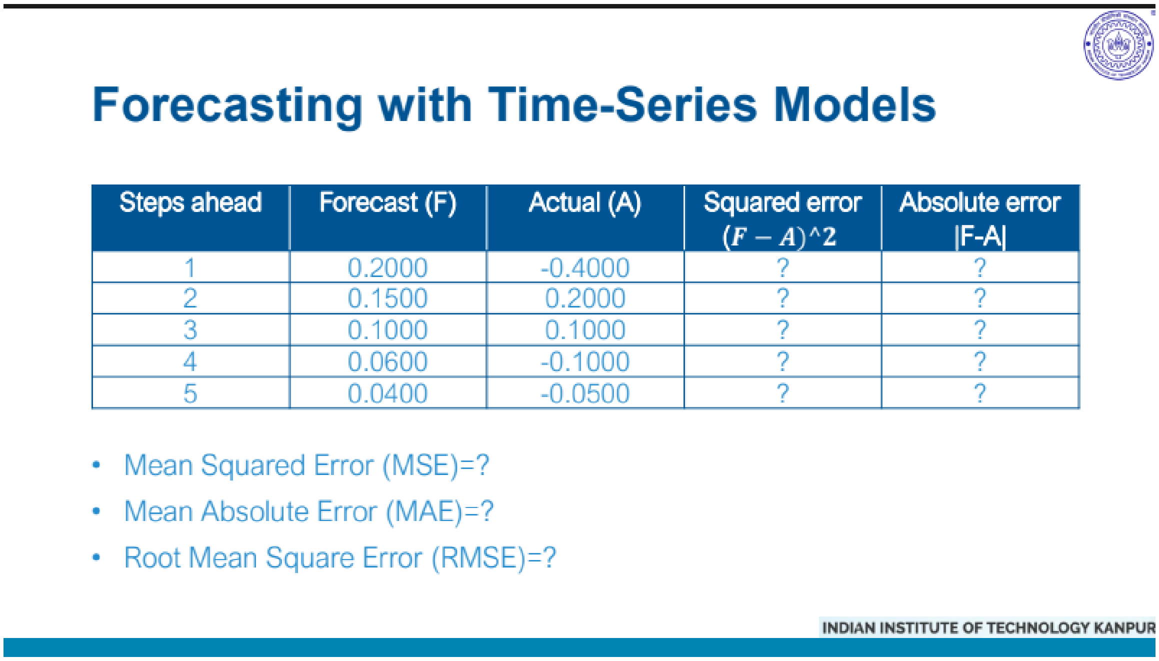
Determining the Accuracy of Forecast

- Conditional expectations: The expression $E(y_{t+1}|\Omega_t)$ states the expected value of y at 't+1' conditional upon information available up to time t, i.e., Ω_t.
- Naïve forecasting: Forecast or expectation of y, s steps into future is the current value of y, i.e., $E(y_{t+1}|\Omega_t) = y_t$. Then the process follows random walk. For a mean reverting series some long-term unconditional average is the best forecast for the series in future.

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

- MSE=(0.3600+0.0025+0.0000+0.0256+0.0081)/5=0.08
- MAE=(0.6000+0.0500+0.0000+0.1600+0.0900)/5=0.16
- RMSE=SQRT(MSE)=sqrt(0.08)=0.28

Also a more advanced version of this mean square error is root mean square which is nothing but the square root of this MSE in this case 0.08 I can take the square root to get the 0.28. The second way is to take the absolute value so again I can compute the error F minus A forecast minus actual take its mod or absolute value and then I can sum up all these values and divide by number of observations for example in this case I have these absolute errors I can sum them up all five and divide by five to get what we call as mean absolute error. To summarize in this video we have understood how to compare across different models by using some kind of error measures these error measures like MSE, MAE, RMSE we have seen there are more possible error measures but please note they do not have any meaning in absolute sense on it on their own they do not have much meaning per se when they are more useful in comparing across competing models and why they are being used because we want to compare different models on their capacity of out of sample prediction so using the actual and forecasted values from different models we compute these error measures on the out of sample or hold out sample part of the data and then



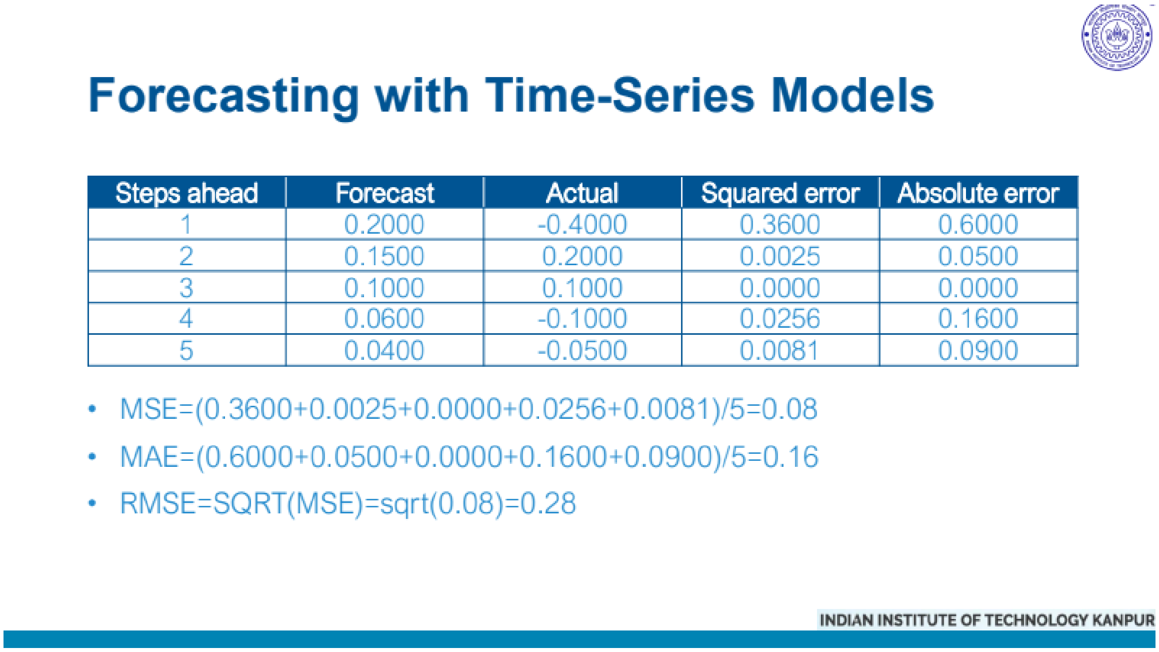
Forecasting with Time-Series Models

Steps ahead	Forecast (F)	Actual (A)	Squared error $(F-A)$^2	Absolute error |F-A|
1	0.2000	-0.4000	?	?
2	0.1500	0.2000	?	?
3	0.1000	0.1000	?	?
4	0.0600	-0.1000	?	?
5	0.0400	-0.0500	?	?

- Mean Squared Error (MSE)=?
- Mean Absolute Error (MAE)=?
- Root Mean Square Error (RMSE)=?

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

compare across different competing models to compare their error values on its own these error values do not make much sense we do not have a way to say that this is bad or good only when we have competing models and their respective errors we can compare them across different error measures also what kind of error measures are being used that is up to researcher in different cases different measure are used for example when you are working with financial market returns probably the positive errors are not so much as bad as negative errors so an investor would rather like to have those models where or evaluate a model based on its negative error how much extremely negative returns or losses he is witnessing maybe he is not very worried about positive returns or gains he probably is more interested on the negative side negative tail of errors so what measures error measures we are going to use that is up to the researcher at the context but these errors will be used on out of sample forecasting or prediction in comparing across different competing models to summarize an ARMA process is simply a combination of AR and MA processes it can be easily identified on ACF and PACF plots as exponentially declining process on both the plots to build an ARMA model first we identify the order of the process next we estimate the parameters through some kind of estimation procedure such as OLS or MLE procedure lastly we diagnose the model with the help of residual diagnostics and out of sample prediction accuracy of the model for lag selection process we use information criteria one of the important objective here is to find extremely parsimonious model which fits the data well in a strict theoretical sense for an ideal model the residuals extracted from the model should not have any discernible structure and should act as a white noise process forecasting with ARMA model requires conditional forecast for next 1 to S steps ahead for those terms where information is available the actual values are employed where the actual values are not available conditional expectations are employed often models appear to be extremely efficient when examined with the sample on which it was trained this often leads to overfitting process that is a very bulky and complex model which does not do very well in new data that is why different computing models are tested on out of sample or hold out sample which was not used while training the data the two key terminologies in forecasting are first rolling sample forecast which employs fixed window of sample which is moving in a step by step manner second is recursive window forecast where initial point is fixed and new data points are added and sample increases with each step in the forecasting process the choice between the two methods is a trade-off between a sample which is a large larger in size versus that which is in sync with the current market conditions lastly across different competing models we compare their performance on out of sample forecasting accuracy through error measures such as MSE which is mean square error MA mean absolute error or RMSE for example which is root mean square error and then select the best model. Thank you.



Forecasting with Time-Series Models

Steps ahead	Forecast	Actual	Squared error	Absolute error
1	0.2000	-0.4000	0.3600	0.6000
2	0.1500	0.2000	0.0025	0.0500
3	0.1000	0.1000	0.0000	0.0000
4	0.0600	-0.1000	0.0256	0.1600
5	0.0400	-0.0500	0.0081	0.0900

- MSE=(0.3600+0.0025+0.0000+0.0256+0.0081)/5=0.08
- MAE=(0.6000+0.0500+0.0000+0.1600+0.0900)/5=0.16
- RMSE=SQRT(MSE)=sqrt(0.08)=0.28

INDIAN INSTITUTE OF TECHNOLOGY KANPUR