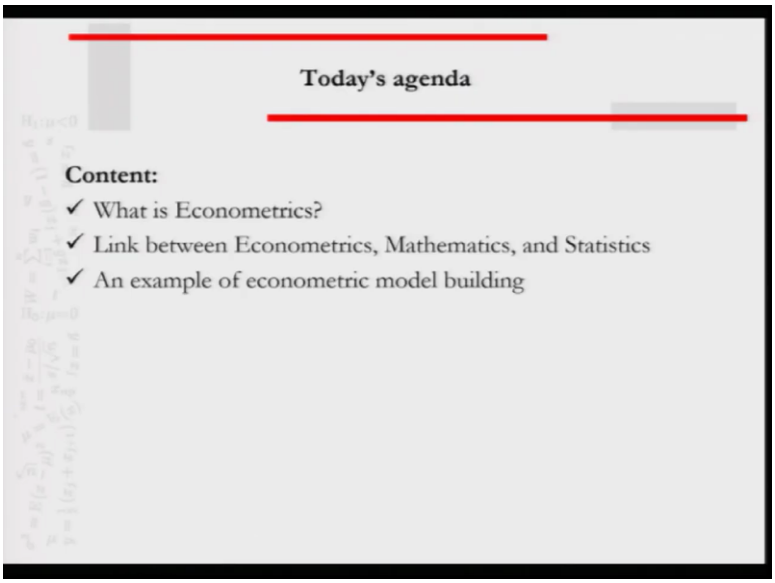
Applied Statistics and Econometrics
Professor Deep Mukherjee
Department of Economic Sciences
Indian Institute of Technology Kanpur
Lecture 02
Introduction to Econometrics

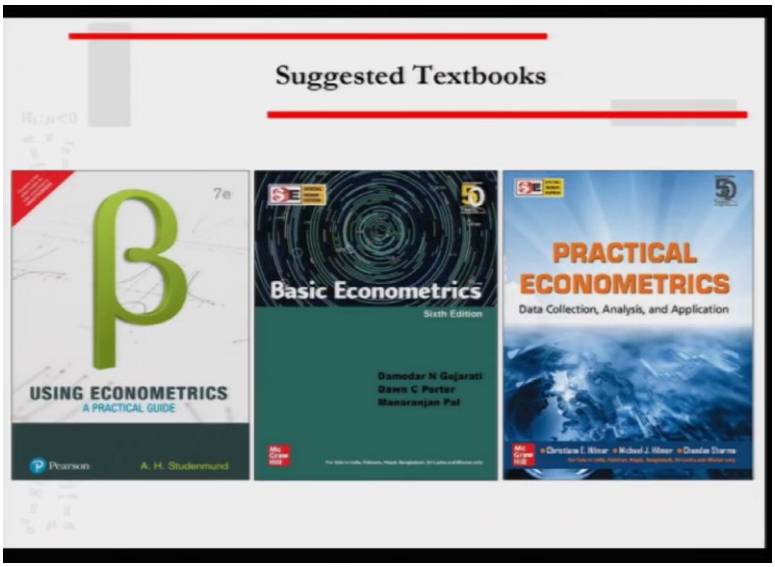
Hello friends. Welcome back to the course, Applied Statistics and Econometrics. This is lecture number 2. And today I am going to take you through the other fields that we are going to discuss in this course, namely econometrics. So, before we jump to different concepts that are there, we will set the agenda items first for today's lecture.

(Refer Slide Time: 00:41)



So, we are going to start with a basic description of the field, Econometrics. And then we are going to provide the link between 3 fields: Econometrics, Mathematics and Statistics. And then we are going to end this lecture with an example of econometric model building. So, the second part of the course is developed based on 3 textbooks and I have borrowed some materials from these excellent textbooks. Let me show you these books.

(Refer Slide Time: 01:11)



So here I show these 3 textbooks. So, let me start with the one titled Using Econometrics written by Studenmund. So, this is the simplest possible textbook teaching you Econometrics. So, those who are not from economics background or doing econometrics for the first time, even if you are coming from a Social Science or Management Science background, I recommend this book. it is written in very simple language. So, you will love to read it.

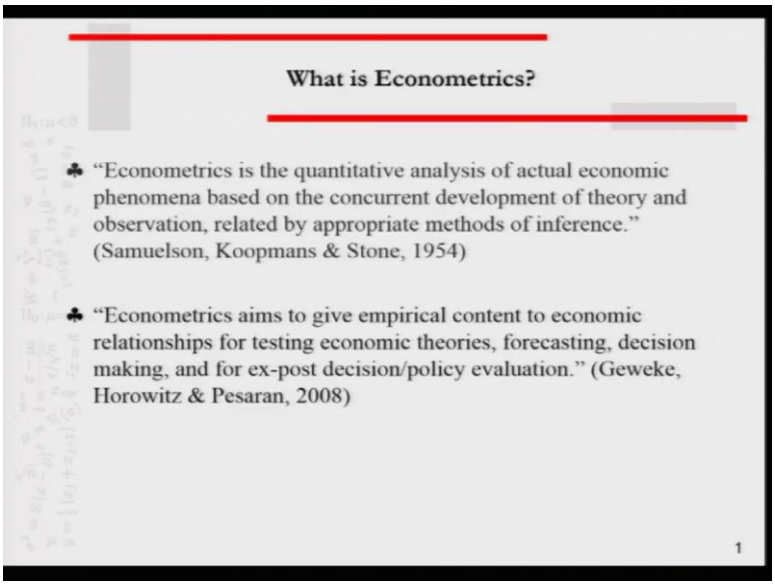
For the second part of the course, I have also borrowed concepts from the second book which is titled Basic Econometrics. Now, this book by Gujarati, Porter and Pal has this speciality that it covers a lot of topics that you may not find in the previous one by Studenmund. So, in terms of coverage, the book has comparitive edge over the other. And here, the authors are showing you econometric computations using Stata which is state of the R software for today's applied econometric research. And the best part is that for both books, the Indian edition is available.

Now, the third book that I would like to recommend and this book is by Hilmer, Hilmer and Sharma and the title of the book is Practical Econometrics. Now, this book is also a very good book because it shows you how to make use of Microsoft Excel for econometric calculations and estimations. So that way it adds value to our learning because the other two books that I have recommended are going to show you the computations using different software but not Excel. So, this third book actually shows you how to do econometric computations using Excel.

And again, Indian edition is available for this book. But needless to say, that there are many other books in the market which you can follow. I would also like to recommend another

book by Neeraj Hatekar and that book actually gives you the principles of econometric theory and it also shows you how to conduct econometric computations using R software. So, in a nutshell, you are free to choose any of these books and I would recommend that in addition to listening to the lectures, you should also get a copy of any of these textbooks and read it.

(Refer Slide Time: 04:05)



So, let us begin by looking at what wise men say. So, here I am going to show you 2 definitions provided by a well-known economist, some of them are Nobel laureates. So, let me first start with the definition provided by Samuelson, Koopmans and Stone. And they have put forward this definition way back in 1954. It is interesting to note that Paul Samuelson was the second economist to obtain Nobel Prize in Economics and T.C Koopmans also another Nobel laureate in economics. So, what did they say?

So, they say that "Econometrics is the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference." So, note that in the last lecture, I have already talked about what is theory and what are observations and I have also spoken about what is inference. So, based on this clearly defined concepts from the last lecture, I hope you can understand what econometrics is all about.

So, let me give an example. So, here let us take an economic phenomenon say demand as I said in the last lecture. So, the demand function is on either blackboard in a classroom or in the pages of a textbook. And we all know that economic theory says that it is the demand function is a downward sloping curve. So, that is basically an economic phenomenon that if price of a commodity increases, quantity demanded shall decrease.

Now, this is development of theory. Now, you collect data either by conducting sample survey or collecting some data from other sources. And then you conduct statistical inference analysis to find out whether this relationship actually holds in reality. So, that is basically the essence of their definition. Now, let us look at another definition which is more recent. So, this one is coming from noted econometricians of our times Geweke, Horowitz and Hashem Pesaran and this was put forward in 2008. Now, what do they say?

They say that econometrics is a field which aims to give empirical content to economic relationships for testing economic theories, forecasting, decision making and for ex-post decision or policy evaluation. Now, hearing this definition note that the noted econometricians of our times are emphasizing 3 different things. One is basically forecasting. The other one is decision making. And the other one is ex-post policy evaluation.

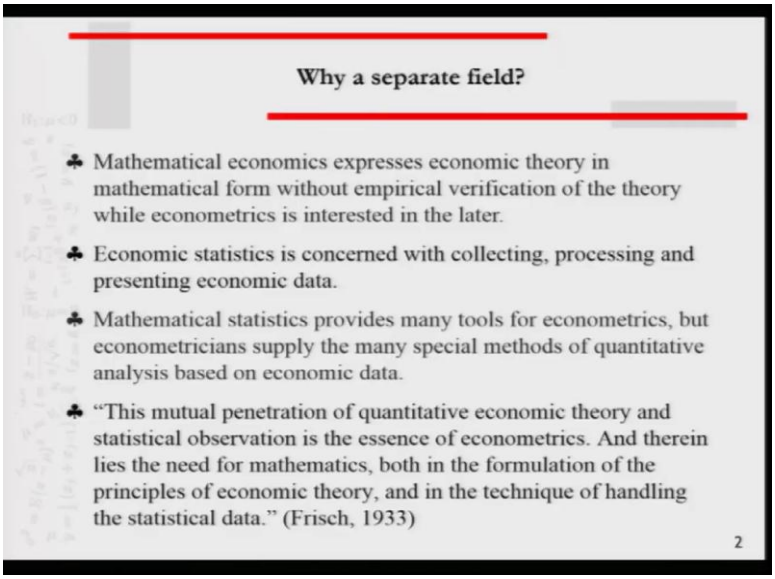
So, let me briefly discuss what they are. So, if you talk about forecasting, what does it mean? It means that okay you are observing some economic phenomena, you collected data for a particular time period. Now, you want to see whether this if this phenomenon is true. It is empirically evident that this phenomenon is going to continue in the future, then what would be the values of your dependent variable in the future.

You remember that we have spoken about dependent variable and independent variable in the previous lecture? So, from now, link this definition with the previous lecture. So, if someone asks you that, okay, I have an econometric model or I have a statistical model and I want to know how in future my dependent variable y is going to behave, if there is a change in the independent variable x. So, that is basically the topic of forecasting.

And of course, in financial econometrics, you see huge application of this forecasting tool. And not only financial econometrics. it is also useful for planning exercises. Now, the other thing that they are talking about is basically decision making and ex-post policy evaluation. Now, these days it has become very important for the funders to know whether after conducting a costly project whether there is an impact in reality.

Suppose there is some development project for the poor in a locality. And you want to know that after spending so much of money and effort for 2-3 years whether there is a change in real life, whether there is a change in the welfare of the people who were subjected to your program. Econometricians will help you to actually draw evidence-based conclusion whether a particular program was successful in real life or not.

(Refer Slide Time: 09:09)



Now, one may thing that okay we have fields like Mathematics, then Economics and Statistics. Why do we need another field with a different name called Econometrics? So, here we are going to discuss why we need a separate field and why econometrics is somewhat different from these 3 fields that I have already mentioned. So, note that there is a sub-branch of economics called Mathematical Economics. It expresses the economic theory in mathematical form the empirical verification of the theory. And econometrics is actually is interested in the empirical verification of the theory.

Now, note that the use of mathematics here. Because when we are talking about economic theory as we found in literature, they are mostly qualitative in nature. They qualitatively explain certain economic phenomena: how things behave; how things move. I mean, of course, I am talking about economic variables here. But then to empirically validate that first you have to have mathematical form of this theory.

Mathematics provides that mathematical form. And then once you have some mathematical equations which are coming from the economic theory, then you can take it to real life and you can empirically validate that theory using data analysis tools. Now, there is also a separate branch called Economic Statistics. What is it? It is concerned with collecting, processing and presenting economic data. Note that this is not econometrics. The economic statistics actually is basically collecting and sharing economic data in an efficient and meaningful manner.

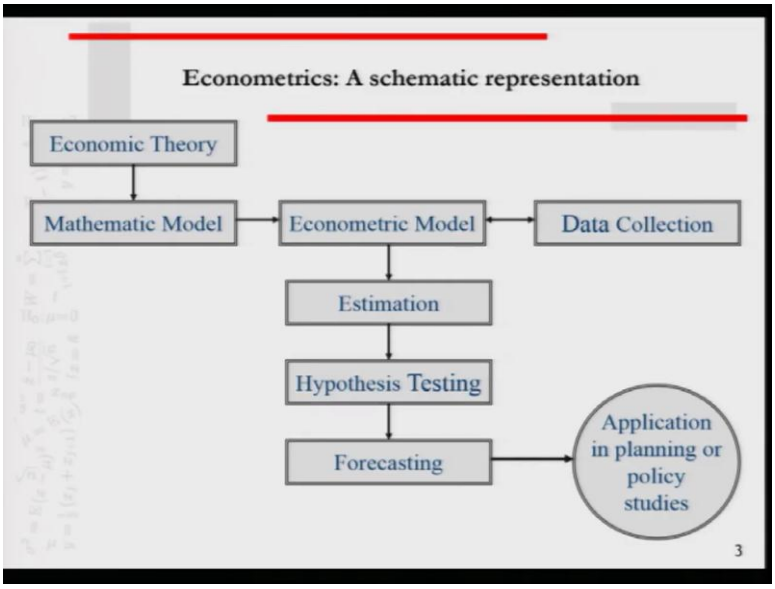
Now, here comes the other field statistics or more precisely mathematical statistics. This field provides many tools for econometrics. Actually, econometrics cannot stand alone. Most of

the tools that we are going to study in this course, you will see that they are actually borrowed from Mathematical Statistics. But also note that economic variables are special and that is why sometimes econometricians have also developed theory and estimators.

So, this inter linkage between 3 fields: Economics, Mathematics and Statistics can be best understood by a quotation coming from another Nobel laureate economist, Ragnar Frisch. In fact, it is interesting to note that it was him, Professor Ragnar Frisch who coined the term econometrics. He actually introduced the term econometrics in 1926. So, what did Professor Frisch said?

So, he said that, "This mutual penetration of quantitative economic theory and statistical observation is the essence of econometrics. And therein lies the need for mathematics, both in the formulation of the principles of economic theory, and in the technique of handling statistical data."

(Refer Slide Time: 12:21)



So, with this, let me look at a schematic presentation of econometrics. It will make things a lot of easier. That is at least I hope. So, you note that the journey of an applied researcher begins from economic theory. And then you derive certain mathematical models in form of mathematical equations, functions, etc. Once functional forms are given to economic qualitative economic theory, then comes the role of econometrics. Then you start developing an econometric model.

And once you have a statistical equation, we will see soon what are these things or how do they look like; once you have certain equation then you collect the data and then the data will

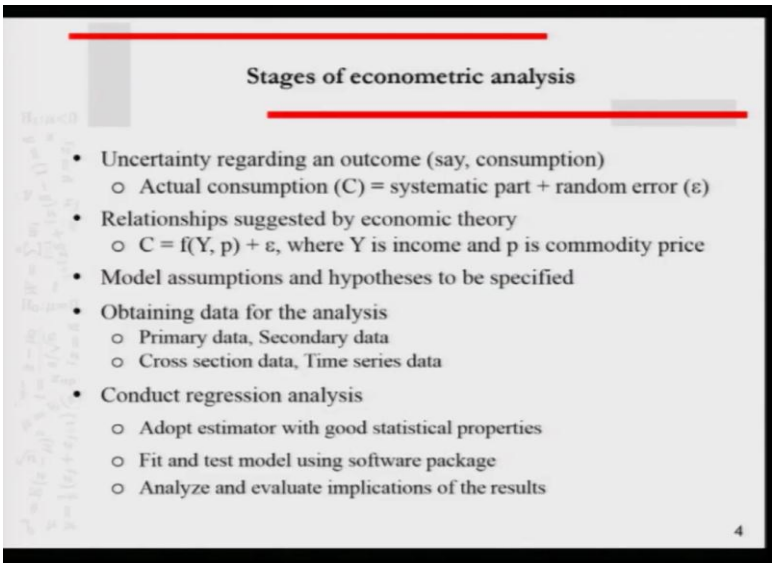
contribute to the estimation of your econometric model. What do I mean by estimation? So, if you remember from the previous lecture, I talked about the difference between population parameter and sample statistic and simple random sampling.

So, basically, that there is a universe, there is a population over which an economic theory may or may not hold. But you do not have time and money to collect data from the entire population. So, you draw a sample. And then from the sample you try to draw statistical evidence whether your economic theory is valid or not. So, that is basically the problem of statistical inference. And statistical inference actually has 2 parts. One is estimation. The other one is hypothesis testing.

So, basically, you estimate the statistical model. By that I mean that you actually get a proxy measure of the population parameters. You will never know the actual values of the population parameter. You can only calculate a proxy. Now, how to calculate in the most efficient way? How to get very good proxies that is basically the subject matter of estimation. And once the proxies are obtained, then actually you try to draw some inference or conclusion from the estimated equation and that is basically the subject matter of hypothesis testing.

So, once a statistical equation or an econometric equation is established from the data, then you take it to forecasting. I already have spoken about forecasting. And then once the forecasting is done, then you can actually apply the obtain results or the derive conclusions in planning or policy studies. So, this is econometrics is all about in a nutshell. Now, we are going to talk about stages of economic analysis through a simple example. Last time we have seen the stages of the empirical social research. Now, we are going to talk about the stages of econometric research.

(Refer Slide Time: 15:28)



So, let us start with a simple economic theory and this is known as the theory of consumption function. So, we write actual consumption denoted by capital C as a sum of two components. One is a systematic part and the other one is random error, epsilon. Now, our relationship that is suggested by economic theory can be written in mathematical form like the following: capital C equals to f of Y, capital Y comma p plus epsilon. So, here Y is income and p is the commodity price.

Now, of course, as this relationship is emerging from economic theory, you have certain theoretical or economic theoretical modelling assumptions and then you specify your, form your hypothesis regarding that. And then you obtain data for the analysis. So, now, here the note that data could be of similar types that we use for statistical or econometric analysis. It could be primary data. So, that is basically the data that you yourself collected by drawing random sample from the population or it could be secondary data. It is the data could come from another source.

So, somebody else have collected data already and you are borrowing the data from others and you are making use of it. So, that becomes secondary data. Now, there is also a distinction between cross section data and time series data. What are they? So, a cross section data is basically drawing a random sample from the population at a given point of time. So, the best example could be the population census.

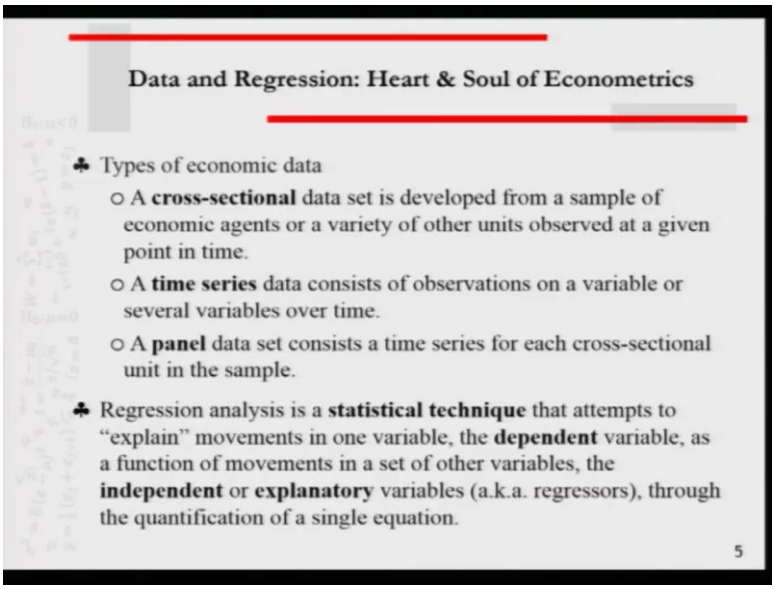
Now, time series data is another type which is very different from cross section data. Under the time series data actually you observe a particular variable over a period of time. So, the best example could be GDP of any economy or per capita income of Indian economy. So,

you observe the numbers over a period of time say 50 years or even less 10 years. So, that basically gives you time series data.

Then once the data is in your hand then you conduct regression analysis. Now, what is regression? Regression is basically the bread and butter of statisticians and econometricians. So, we will soon see what it is of course in layman's terms. So, now, we are going to look at different stages of regression analysis in layman's term. So, we have to adopt an estimator with good statistical properties. It may seem very Greek to you. Wait. as we move on with the course this thing is going to be clear to you. I will make it clear. I will devote separate lectures on this topic.

Now, the second issue is basically the issue of fitting. So, you fit the model and then test the model using a statistical software package. We are going to see the use of R and Excel to do statistical or econometric data analysis or regression analysis. And then finally, we are going to analyse and evaluate the implications of the result. Now, in the last slide only you have seen the different types of data I have spoken about. Now, in addition to the type of data that I have spoken in the last slide, I am going to add one more here and that is basically the panel data. So, what is a panel data?

(Refer Slide Time: 19:02)



A panel data set consists a time series for each cross-sectional unit in the sample. So, this is basically a bookish definition. Now, if I want to explain you in simple terms, how could I find example from real life? It is very easy, suppose, you have a sample of 10 companies and you want to study their profitability for 10 or 15 years. So, you gather a data set like that 100

observations. if you are studying or monitoring 10 companies over 10 years, you actually collect 100 data points and then that data set is called a panel data.

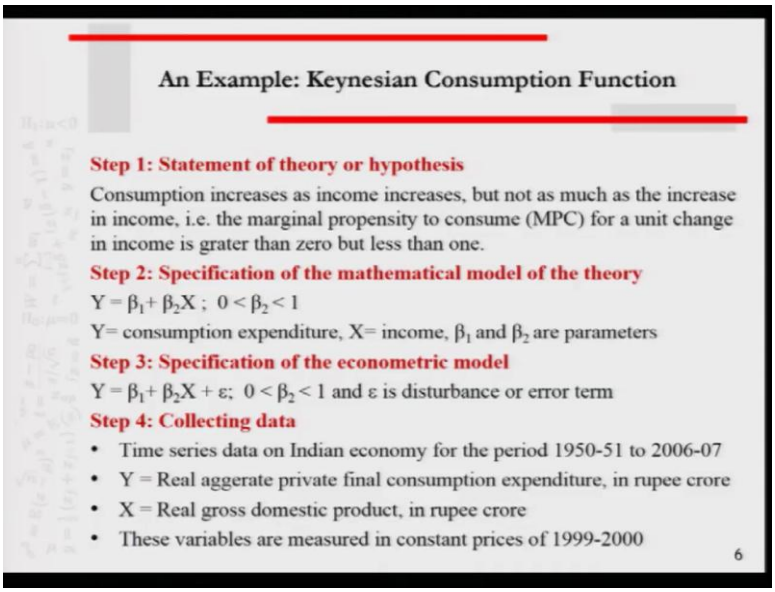
So, now, we are going to briefly visit the concept of regression. Because we have spoken about regression in this lecture. What is it? So, the idea of regression is not complicated but the details of regression analysis, we will cover in the later part of the course. So, till then you have to be patient. Right now, you please understand what is the philosophy of regression analysis through this simple definition.

So, regression analysis is a statistical technique that attempts to "explain" the movements in one variable, which is a dependent variable as a function of movements in a set of other variables, which are independent or explanatory variables. Sometimes they are also called regressors through the quantification of a single equation. So, now, if I want to link this bookish definition to real life, then through an example we can do it.

So, we have spoken about the consumption function. So, you can say that household consumption expenditure could be your dependent variable Y. You want to actually explain the variation in that dependent variable Y. And then actually what could be the excess? So, from the previous slide where we have discussed that example, your household income could be one explanatory variable, the price index could be one another explanatory variable. There could be other explanatory variables like tests and preferences you know; you have to find proxy. How to control for it?

There could be household size as an explanatory variable. So, if you have that kind of model then you write the mathematical equation, a linear equation and then you estimate it. So, that is basically regression analysis in a nutshell. So, the equation that involves the explanatory variables X actually will help you to explain the observed variation in the Y your, the dependent variable, which is the household expenditure here.

(Refer Slide Time: 21:53)



So, let us now discuss the idea of regression analysis through an example again from economics. So, the students from economics discipline are all aware of the Keynesian consumption function. This is the first thing that, or this is a concept that they go through in the first semester of their lessons in economics honors program. But those who are from the other disciplines, do not be afraid. it is a very simple concept. Let me explain.

So, what could be the step 1? So, step 1 is basically you have to state the theory and you have to form the hypothesis. So, this itself will explain what is the Keynesian consumption function. So, here consumption increases in a year. consumption could be in physical terms; consumption could be in monetary terms like consumption expenses. Now, this increases as income increases. But not as much as the increase in income.

So, it implies that if there is an increase of income of 1 rupee, then consumption expenses will increase but it will not increase by 1 rupee. It will be less than that. So, it could be 80 paisa. It could be 90 paisa. We do not know appropriately. But, from this theory, this concept is emerging and that is called marginal propensity to consume. So, that is basically the quantity of change in consumption expenditure from a unit change in income. And that is generally expected to be greater than 0 but it is hypothesized to be less than 1.

So, now, we move on to step 2. So, here we have to specify a mathematical model and we have to write some statistical or mathematical equation from the theory. So, that we can take it to real life data for empirical validity. So, here we are write: so, capital Y equal to beta 1 plus beta2 X. It is an equation of a straight line. Note that I am putting some range for my

beta2 coefficient because that beta2 coefficient is basically the marginal propensity to consume that I have already spoken.

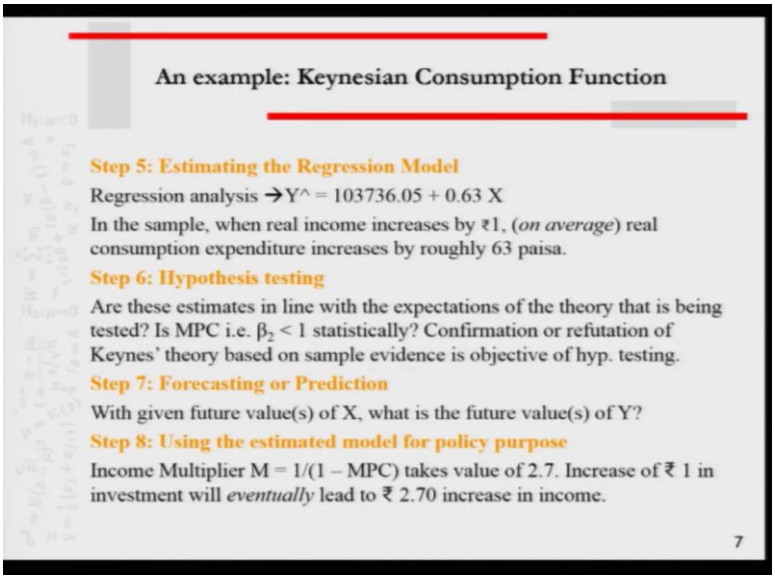
Now, step 3 is basically specification of the econometric model. So, note that I have already mentioned that when we do econometrics, we are not in a deterministic world. So, we are in an uncertain world. So, that uncertainty is contributed by a stochastic disturbance or a random disturbance term which is popularly abbreviated or measured through epsilon symbol.

And so, we write an econometric equation which is capital Y which is basically consumption expenditure equals to beta1. So, that is the intercept parameter plus beta2 which is the slope parameter times X which is basically the income plus epsilon. So, note that these beta1 and beta2 are the population parameters for which the values are unknown to me.

So, the step 4 would be collecting data. So, if we now want to estimate this equation, one can estimate into 2 ways. either you collected data from random sample and that is basically by estimating the equation from a cross sectional data. But suppose, I take the other way around. I collect time series data on Indian economy for the time periods in 1950-51 to 2006-07. And these are my variables.

So, Y is my dependent variable that is basically real aggregate private final consumption expenditure in rupees crores. And X is basically my explanatory or independent variable and that is real gross domestic product in rupees crore. So, that is basically proxying the income. Now, these variables are all measured in constant prices of 1999-2000. So, they are basically representing these monetary variables in the constant price. we will learn the art of index numbers in this course later on.

(Refer Slide Time: 26:13)



Now, with this equation and data all set, we move on to step 5 where we estimate the regression model. So, we conduct a regression analysis and we get a fitted line Y hat? Now, at this moment I request you to be patient. Do not ask me now, how come we are getting these numbers in place of this beta1 and beta2. You will soon realize and learn how we are going to get these numbers in place of beta1 and beta2. But as of now, you just take it as it is.

So, you see the population parameter beta1 and beta2 are actually proxied by the estimates beta1 hat and beta2 hat. And beta hat is a large number that is the estimated value of the intercept parameter. And then 0.63 is basically beta2 hat that is basically the estimate from the sample for the slope parameter of the straight line and that can be interpreted as the marginal propensity to consume.

So, as this slope parameter value has this interpretation. In words, we can say that when real income increases by up 1 on average, real consumption expenditure increases by roughly 63 paisa. Now, this is basically the estimation part of the statistical inference. Now, the second part of the statistical inference procedure is conducting hypothesis testing. Now, one may ask this question: "Are these estimates in line with the expectations of the theory that is being tested?"

We have already told you that we expect that MPC is bounded by 0 and 1. So, here we see a number which is 0.63 and is far away from 0 but it is well not that far away from 1. So, one can statistically test whether beta2 is indeed less than 1 or not. Because our theory says that as income increases, the consumption expense increases but not by the same amount. So, it has to be less.

So, we statistically test that and the confirmation or refutation of Keynesian's theory based on the sample evidence is the objective of the hypothesis testing procedure. So, in step 7, we conduct forecasting or prediction. Suppose, we have conducted hypothesis testing already and we found that well, there is empirical evidence, there is confirmation for this Keynesian consumption function estimate that we obtain from the data, then we take it to the forecasting or prediction step.

Now, the question by a policymaker or a researcher is of this sort that with given future values of X's, what is the future values of Y? So, one may retail the question in the layman's language that okay, if there is economic growth, if the country's GDP is increasing, then how the country's consumption expenditure is increasing with time? So, now, we come to the last step. So, that is basically you use the estimated model for policy purpose.

Now, note that we are going to use our result from the macroeconomics that is the concept of income multiplier, capital M. And that is basically defined as 1 over 1 minus MPC. So, that takes value 2.7. Now, what does it mean? So, it means that an increase of rupees 1 in investment will eventually lead to rupees 2.70 paisa increase in the income. So, those who are from economics background easily appreciate this example. And of course, the scene is very, very important for policymaking purpose. So, with this let me conclude today's lecture and please come back for the next lecture. Thank you.