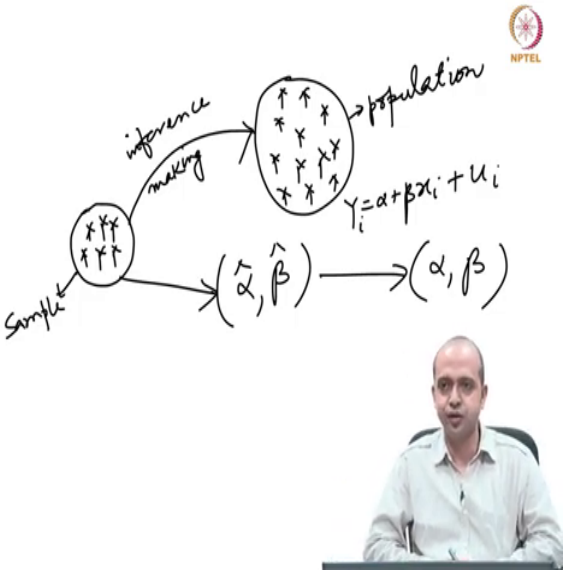
Introduction to Econometrics
Professor. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture No. 04
Different steps in econometric analysis Part - 2

(Refer Slide Time: 00:14)



So, let me explain the broader objective of econometric analysis which is drawing inferences about the population. Let us say that this is my entire population consisting of several individuals. So, I want to study the consumption behaviour of this particular population. Let us say that we are talking about the consumption behavior of the entire Tamil Nadu and what is the consumption behavior? The consumption behaviour what we have previously mentioned is that we will study how the people in Tamil Nadu behave when their income changes. As income changes what is the additional change that these individuals in the population of Tamil Nadu make in their consumption. But what we will do? We will not go to each and every individual of Tamil Nadu because there are so many individuals residing in Tamil Nadu and it is not actually possible to go to each and every individual and collect data because that is very time consuming and costly affair. Rather we will collect a sample from this population, let us say that I have collected a sample of 1000 individuals from this population. This is my sample. So, that means, if you think about observing this sample, we will collect this sample and then we will analyze and will apply economic tool on this sample and then we will try to infer something about the population parameter. This is about inference making.

So, we will have a population in this particular context and as an example, let us assume that the entire Tamil Nadu is our population. From that particular population we have collected a sample of let us say 1000 individuals. Then we will apply econometric tools on the sample and then we will try to infer something about the population. This is basically the first objective of econometric analysis.

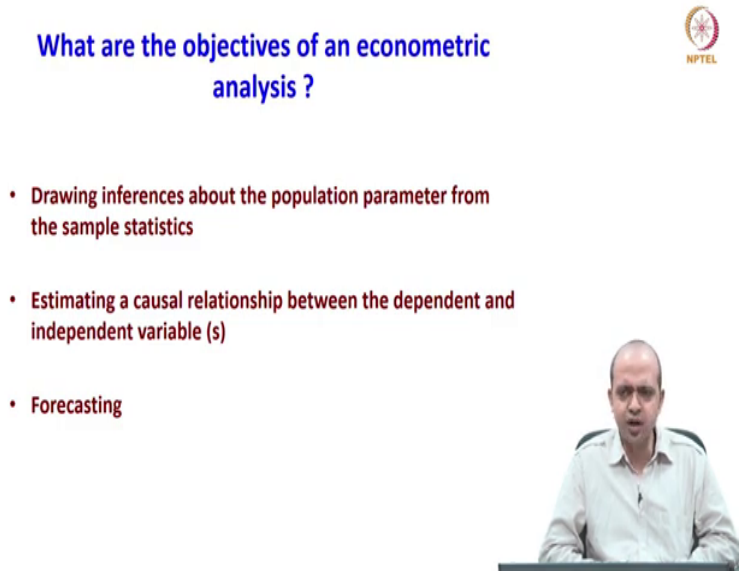
In terms of the population, what is the particular econometric model that we have specified? $Y_i = \alpha + \beta X_i + u_i$. Earlier when you were talking about mathematician's objective and mathematical models, we were talking about only $Y_i = \alpha + \beta X_i$. We never introduced the term u_i. Now, here I have introduced the error term as well, which is u_i. This error term basically captures the other important variables which are not included in my model but they are also having some kind of impact on your consumption. It is very easy to understand that apart from income there might be many other variables that might have some impact on your consumption-maybe your gender, whether you are a male or female, maybe your taste and preference, maybe your age, maybe your education so on and so forth. All these variables are not included in the model. Even though you include all those variables in the model, consumption is basically on human behaviour and human beings do not behave like a computer.

So, even if you know somebody's income, consumption, age, education etc, it is very difficult to predict exactly what will be the individual's consumption. That is why before conducting econometric analysis we convert the mathematical model into a statistical one by adding this stochastic error term. This error term u_i basically captures the influence of other variables that is other omitted variables on the consumption. So, we will specify this model and then collect the sample and from the sample we will estimate $\hat{\alpha}$ and $\hat{\beta}$. So, from the sample, we will get $\hat{\alpha}$ and $\hat{\beta}$, and using this $\hat{\alpha}$ and $\hat{\beta}$, what we will do? We will try to infer something about the true population parameter. So, our true population parameter is α and β because that basically explains the structure of the population. That means, they explain the behaviour in the population. They are actually unknown population parameters and the sample counterpart of those population parameters are known as sample statistic.

So, the objective of the entire econometric analysis is basically to draw inferences about the true population parameter α and β from that sample statistic- $\hat{\alpha}$ and $\hat{\beta}$. This is the first

objective of econometric analysis which is drawing inferences about the population. And how we will do? This particular figure explains how we actually do about it. We have a population, we collect a sample, and then we estimate the econometric model using econometric tools. How will you estimate? That I will explain in a later part of our discussion. Once you estimate, you will get $\hat{\alpha}$ and $\hat{\beta}$. They are basically the sample counterpart of the true population parameter. Using this $\hat{\alpha}$ and $\hat{\beta}$, we will draw some inferences about α and β, the unknown population parameters and our objective is to estimate $\hat{\alpha}$ and $\hat{\beta}$ in such a way that $\hat{\alpha}$ and $\hat{\beta}$ is as close as possible to the true population parameters α and β. So, this is the first objective of an econometric analysis. What is the second objective?

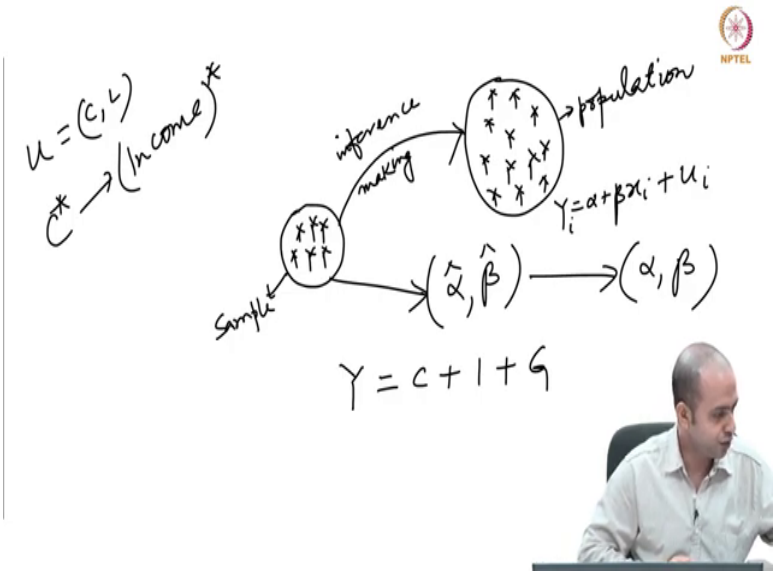
(Refer Slide Time: 08:13)



Second objective is estimating a causal relationship between the dependent and independent variables. Here I am using two new terms-Dependent and Independent variable. So, in our example, when we are talking about the consumption-income relationship, consumption is called the dependent variable because that is dependent on somebody's income. We are assuming that consumption is a dependent variable and income is the independent variable in this model. This is a very simplistic model where we have included only one variable. That is why it is called a simple two-variable model and then we are trying to establish a causal relationship between income and consumption.

Now, we have to keep one thing in mind. If you recall, initially we mentioned that econometricians' role is like an artist. So, here you assume econometric analysis per se will not tell you what should be the dependent variable and independent variables. That should come purely from your own understanding or from the existing theory. So, when you are trying to formulate the econometric model then you have to take help from the existing literature or you have to apply your own logic to decide among these two variables between income and consumption. What should be the dependent variable and what should be my independent variable? When you are thinking income as the independent variable, it is very easy to establish the link between income and consumption because generally, it is income that leads to somebody's consumption. But is it possible to think in another way? Can consumption also lead to income? Apparently, you may think no and it is not possible and it is only income that should lead to consumption. But it is also possible that consumption will also lead to income. For that probably what we should do?

(Refer Slide Time: 10:46)



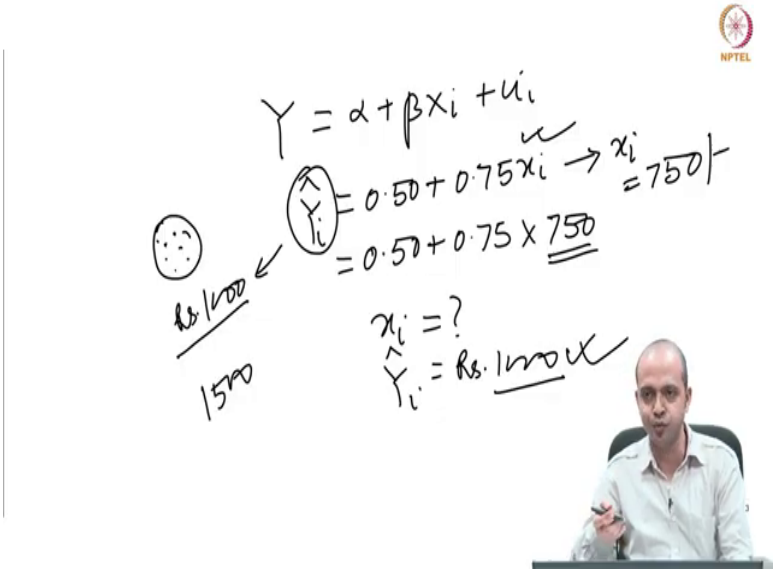
We should think about one macroeconomic identity. If you are aware of national income identity, where Y is basically equal to C plus I plus G. C is consumption expenditure, I is investment expenditure, G is the government expenditure. Here consumption is basically leading to income. So, that means, at the macro level consumption also leads to income. This is by expenditure method at the macro level. At the micro-level also, if you think, depending on my consumption level, I will try to get a job, I will try to participate in the labour market so that I can maximize my consumption. If you think that the individual is facing a particular utility function, where utility is let us say a function of consumption and the amount of

leisure that I will take, then more the consumption, more I will have the utility or satisfaction; and to get the optimum level of consumption, I should have the optimum level of income as well, which is let us say income star. So, depending on now this C^* may vary from person to person, you may have a higher level of consumption for which you might be willing to participate in the labour market, you might be willing to supply more labour than me, because my demand for this consumption that is my consumption is lower than yours.

So, depending on the individual's consumption they will participate in the labour market and will decide what type of job to take in the labor market, how much labour hours to supply in the labour market all of which have been already explained as economic data. So what I am trying to say is that consumption can also lead to income at both micro as well as the macro level. So, what type of econometric model you will specify? Whether you will specify $Y_i = \alpha + \beta X_i$ or $X_i = \alpha + \beta Y_i$ depends on your research objective. That is why the researcher or the econometrician himself or herself has to decide about the dependent and independent variable. Econometric analysis per say will not tell you what should be your dependent and independent variable.

So, once your dependent-independent model, then you can actually establish the causal relationship by looking at the values of the beta hat that you have estimated from your model. And then lastly, we can use econometric models for forecasting as well.

(Refer Slide Time: 14:31)



How will you use that? Suppose, this is your consumption function $Y_i = \alpha + \beta X_i + u_i$. This is your model and after estimation, from the sample you have got this one which is your estimated model. Now how could this be used for forecasting? For any particular year, let us say I want to use this in 2018 data, let us say this the macro consumption function that I have estimated, or let us say for a particular individual instead of 2018, let us say I would like to put the i individuals. Let us say X_i equals to individuals' income which is 750 rupees. So, if you put 750 here, then you will get Y_i as 0.75 multiplied by 750. So, you will get the predicted consumption for the i^{th} individual.

So, this is how I can use the estimated econometric model to predict or forecast. Actually, the term forecasting is used in a time series context and prediction is used in both time series and cross-sectional contexts. In this particular context when you are using cross-sectional data, we might be using the term prediction. So, when you have this $\hat{Y}_i$ equals to let us say $0.50+0.75X_i$, then for a particular value of X_i, you will get $\hat{Y}_i$ and that will be your predicted value. And that predicted consumption you can use for several other purposes.

Sometimes, what we want to know? Let us say I want to generate that level of income which will result in 750 rupees of consumption per week. I want to generate that level of income that will generate let us say 1000 rupees of consumption per week. Why I want consumption of 1000 rupees? Because sometimes policymakers might be interested to know the minimum

income that the individuals must-have so that the individual is able to pay 1000 rupees weekly on different consumption items like Rice, Dal, Oil, so on and so forth.

Because we believe that if the consumer is able to buy that consumption bundle that will give you the minimum calorie that satisfies individual's nutritional requirement, I want my individuals to consume 1000 rupees. So, what should be my level of income? So, that also you can easily estimate from this estimated equation. So, you can solve for X_i. What is the value of X_i? When let us say $\hat{Y}_i$ equals to rupees 1000 that also you can do. So, two things you can do here. You can plug in any value of X_i and that will result in the corresponding $\hat{Y}_i$.

If I put X_i equal to 750, then basically you will get a corresponding $\hat{Y}_i$ value and that is a prediction. Given any value of X_i, I can always predict the value of Y_i. When my objective is not to predict the value of Y_i given the value of X_i, what I would like to derive? Let us say, what should be my corresponding X_i value so that my predicted consumption is 1000 rupees because with 1000 rupees which minimum consumption, what should be the value of X_i.

So, that the consumer can spend minimum 1000 rupees because this minimum 1000 rupees is required to buy that consumption bundle that will give the minimum nutritional value. This is how sometimes the poverty line is defined by the policymakers. Policymakers know this consumption of 1000 rupees is required. Now, to sustain that 1000 rupees consumption what should be my level of income? Whether the existing level of income is enough or additional support has to be given. For example, let us say if the 1000 rupees consumption is required; let us say that income is 1500 rupees. So, then policymaker will decide how many individuals are actually having this income of 1500 rupees on a weekly basis. So, if they are not having this income of 1500 rupees, then government has to provide external support. So, this is one example of how to use this econometric model for prediction and forecasting.

Forecasting is a time series phenomenon as I said that. Let us say this is a time-series data collected from several years of GDP and let us say your consumption is a function of, let us say GDP and you plug in any year's value on X_i. Given the value of X_i, you will get your total consumption expenditure of the country or state. So, if you have time series data for any particular year, you put your X_i and then you will get your $\hat{Y}_i$. So, that means you can also use this estimated econometric model for forecasting.

With this, we have completed our discussion on the objectives of an econometric analysis. Firstly, drawing inferences about the population specifically drawing inferences about the population parameter α and β with the sample statistic $\hat{\alpha}$ and $\hat{\beta}$ and then estimating a causal relationship between the dependent and independent variable; and how will you get the relationship? Depending on whether the coefficient of the independent variable, here β is the coefficient of income, depending on whether the β is statistically significant or not, we will establish whether income actually is causing consumption or not.

How will you check the statistical significance? That we will discuss in a later part of our discussion. So, for the time being, just keep in mind that testing causal relationship basically comes from checking the significance of the estimates of true population parameter β which is $\hat{\beta}$ and then forecasting. Once you estimate the model, use any particular value of X and you will get a corresponding value of Y. If it is a cross-sectional data which generally is a prediction or if it is a time-series data at the macro level and you are trying to predict the consumption of a particular state or country it is called forecasting. These are the 3 objectives of an econometric analysis.