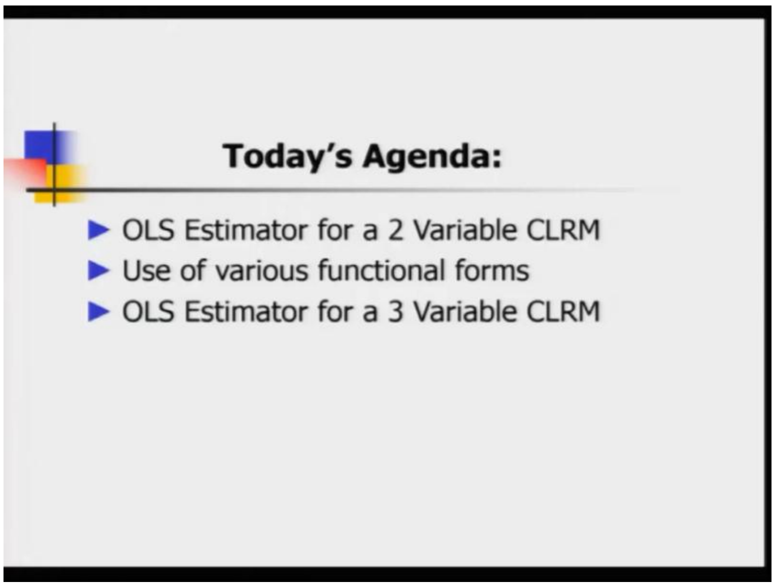
Applied Statistics and Econometrics
Professor Deep Mukherjee
Department of Economic Sciences
Indian Institute of Technology Kanpur
Lecture 26
Classical Linear Regression Model (Part-II)

Hello friends. Welcome back to the lecture series on Applied Statistics and Econometrics. So, today we are going to continue our discussion on Classical Linear Regression Model that we have started only in the last lecture.

(Refer Slide Time: 00:27)

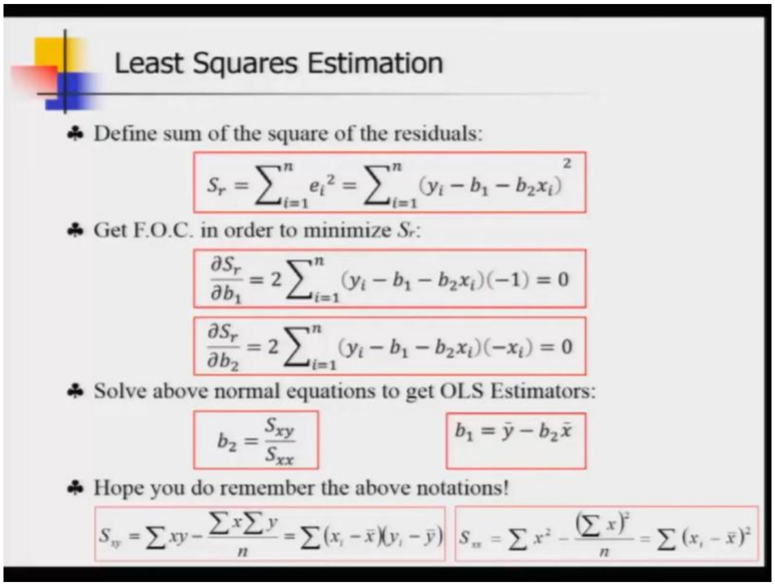


So, let us have a look at today's agenda item. So, if you remember in the last lecture we have discussed the case of least squares principle through a graph. So, here in this lecture I will now discuss about the mathematical derivation of the OLS estimator for a simple 2 variable classical linear regression model.

Then I am going to talk about various functional forms that you can assume for your regression function or equation. And then finally I am going to generalize the model a bit more. And now I am going to talk about a 3 variable classical linear regression model where you have 2 explanatory variables.

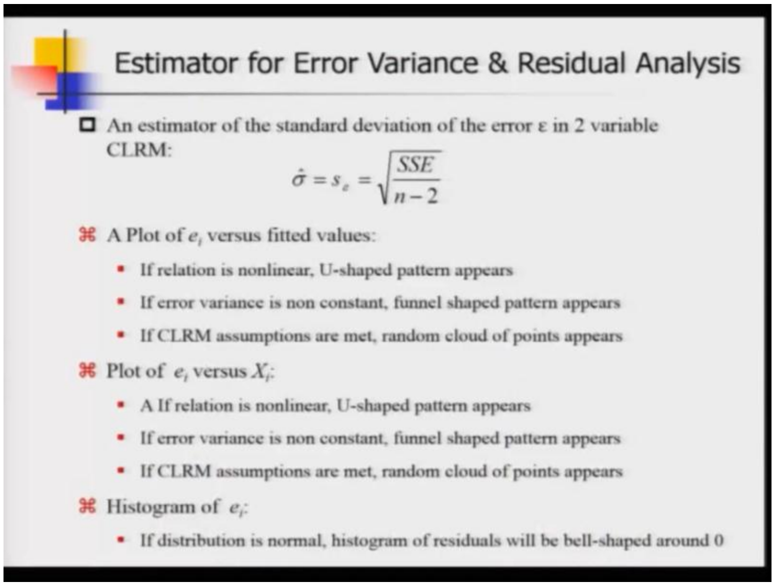
So, hopefully you remember from the previous lectures that we have this problem of finding the good proxies for population regression coefficients beta 1 and beta 2 which are unknown, So, how to get the best possible proxies? So, Gauss invented this method of least squares principle and then we are going to follow that principle to get the best proxies for these 2 unknown quantities.

(Refer Slide Time: 01:38)



So, this slide is going to give you a flavor of the least squares estimation technique step by step. So, we have got the estimates for the regression coefficients from a simple classical linear regression model but that is not the end of the story. Do not forget that there is another unknown population parameter that is embedded in the classical linear regression model and that is the unknown population error variance, I am talking about sigma square. So, we also have to get an estimate for that unknown quantity and for that we need an estimator formula. So, let us have a look at that.

(Refer Slide Time: 02:18)



So, here I am proposing an estimator of the standard deviation of the error term epsilon. So, that is basically in the 2 variable classical linear regression model case and that is given by sigma hat or it can be alternatively abbreviated as s because s stands for standard deviation in the sample context. And that can be given by square root of sum of square errors divided by n minus 2. Now we have already seen the formula for sum of square errors that is basically the sum of squares error residuals which you are minimizing to get the OLS estimates.

And you need to divide that by the degrees of freedom n minus 2, why? Because you are estimating the slope and the intercept coefficient of the linear regression equation before you can even attempt to find an estimator for the unknown population error variance. So, that is why the degrees of freedom is n minus 2. So, here if you take the square of this formula both sides then you get sigma square hat as the estimator of the unknown population error variance and that is given by sum of square errors divided by n minus 2.

So, do you think that we are done here? No, not. So, once you get your coefficient estimates and estimate for the unknown population error variance then we have to see whether the residuals that are generated by the OLS technique are obeying the assumptions of the classical linear regression model because if the residuals are not following certain patterns then we cannot be sure that the assumptions with which we have started our journey are actually validated by the sample data that we have.

So, it is important to do some residual analysis. Now residual analysis could be done in a very advanced manner but for this particular lecture we are going to look at very simple graphical tools which can be used for doing some residual analysis to check whether the residuals are showing us some signal whether the data obeys the classical linear regression model assumptions or not. So, let us have some graphical measures that can be used for residual analysis.

So, here I am going to talk about 3 plots. So, if in the scatter plot you find some kind of an U shape or inverted U shape kind of pattern then you can assume that well the relationship between the X and Y may be non-linear. And if the scatter plot shows you some kind of a funnel shaped pattern then you can guess that the error variance is not constant.

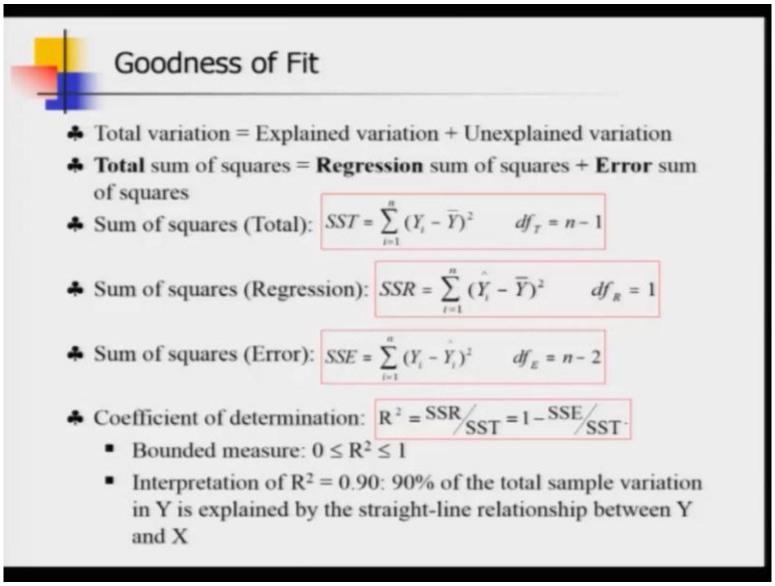
So, the homoscedasticity assumption is not fulfilled and finally if you see a random cloud of the points in the scatter plot then it is good for you then you can assume that my classical linear regression model assumptions are all fulfilled. And finally we are going to check for the normality of the fitted residuals. Well we do not have to do it mandatorily because note that nowhere we have assumed that my error follows a normal distribution.

So, we really do not have to figure out whether residuals follow a normal distribution or not but later on you will see that I will make assumption of normal distribution for the error term or the residual term and in that context you have to check whether the fitted residual values actually are showing you a pattern for normal distribution or not. So, for that we have to draw a histogram.

So, we make a plot of histogram of the fitted residual values and if the histogram of residuals is kind of bell shaped around 0 then we can guess that okay the distribution is very close to normal. So, the normality assumption that I made for my error term in classical linear regression model is also fulfilled. So, now we are going to talk about the goodness of fit measure for our regression model. Once you have got a data set and you fit a straight line or linear regression to the data then the next question arises in mind that how good or bad your fit is to the overall scatter plot of X and Y.

So, if you remember we had a discussion about the coefficient of determination couple of classes back. So, we are going to bring that concept here again and I am going to just briefly talk about the coefficient of determination and I am going to show you some extra results which you have not seen before.

(Refer Slide Time: 07:49)



So, note that total variation in the dependent variable Y can be broken down into 2 parts and they are explained variation and the unexplained variation. So, what is my explained variation? So, explained variation is basically the variation in Y which is explained by the fitted regression model. So, one way you can also call it a regression sum of squares and what is unexplained variation? Of course the name explains itself.

So, whatever variation in Y cannot be explained by the regression model or the set of explanatory variables that is basically called the unexplained variation. So, in the other words you can call it error sum of squares. So, the total sum of squares can be defined as SST and I am showing you the formula here. So, that is basically you first have to deduct the sample mean of Y from each and every individual observations on Y then you need to square it and then you need to sum over all individual observations. So, that is the way you get your sum of squares total.

So, note that there is a degrees of freedom associated with this formula because you have to first calculate the sample mean. So, the degrees of freedom is n minus 1. So, note that SST can be broken in SSR and SSE. So, here that is coming from the second bullet point in the slide. So, now I am going to talk about 2 components SSR and SSE. First we will talk about the sum of squares that emerge from the regression model and that is given in the 2nd red box here. So, SSR is defined as the sum of difference squared difference of Y hat i minus Y bar.

So, what you have to do? You have to basically first calculate the coefficients b1 and b2. Then you have to get the fitted values which are Y hat i and you already had calculated the sample mean for Y. So, now you have to subtract that sample mean of Y from this fitted values, square them and then you sum over all individuals and that is the way you get your SSR. So, now we are going to briefly talk about coefficient of determination.

Now note that last time we have talked about small r square. Here I am talking about capital R square. They are basically the same thing but capital R square is generally used for multiple regression purposes and that last one small r square that we discussed couple of lectures back that was done for two variables only. So, I mean it is not a big difference and the concepts are same. So, here this capital R square it is defined as the ratio of SSR and SST.

So, now note that SSR can be written as SST minus SSE and that comes from the identity total sum of squares equal to regression sum of squares plus error sum of squares. So, now if you have that identity in mind then you can actually rewrite your coefficient of determination formula as 1 minus SSE divided by SST. So, note that this is a bounded measure and it is going to lie between 0 and 1. Why? Because note SSE can take the lowest possible value as 0.

So, that means that your sum of squares from the errors is 0 that means that the fitted value is exactly equal to the observed value of Y for all individuals. So, that means you have got a perfect fit. So, in that case actually you can observe R square value to be equal to 1. Now the other extreme could be that SSE and SST will take the same value.

So, that means that you have thrown a couple of regressors or explanatory variables in your regression model. But unfortunately they fail to explain any bit of variation in Y and then of course your sum of square residuals will be exactly equal to the total sum of squares. Hence these 2 terms will cancel out to give you 1, 1 minus 1 you will get 0. So, the another bound is 0 but note that this extreme cases like you observing R square equal to 1 and R square equal to 0 is not a very realistic case.

So, most of the times when you are dealing with real life datasets you are going to observe R square as a fractional number. Now how do I interpret R square? Suppose I have got a R square value of 0.9. So, what does it mean? So, it means that 90 percent of variation in the dependent variable Y is captured by the fitted straight line.

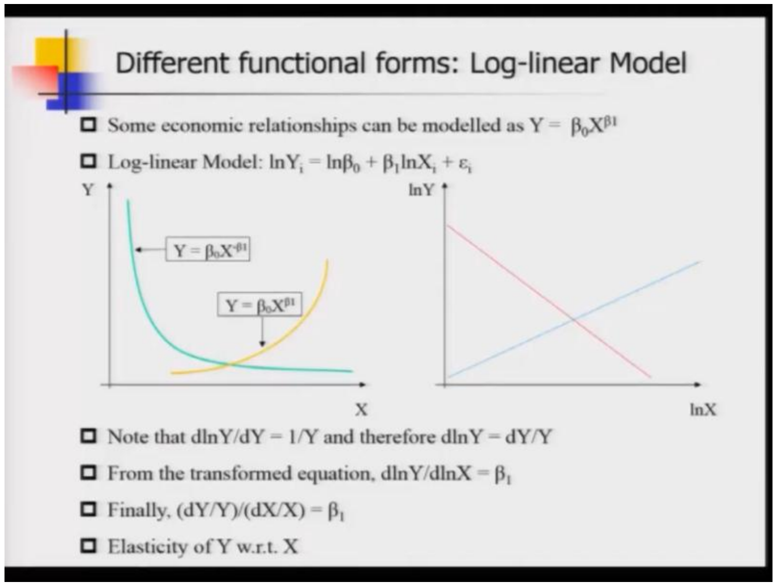
So, now I am going to talk about different issue. So, far we have got this specific functional form for the regression equation which is basically the equation of a straight line. It can be an upward sloping straight line it can be a downward sloping straight line but it is a straight line. Now you may be under an impression that well does it mean that then linear regression analysis is applicable to quantify linear relationships only? The answer is no, not at all.

The linear regression analysis is called linear because it is linear in parameters. The variables can be non-linear. So, linear regression analysis is actually a very general concept which can also capture non-linearities in data. So, how do you then introduce non-linearity in a linear regression equation? So, that is basically through the transformation of the variables. And there are many possibilities exist in real life because the non-linear relationship between 2 variables can be defined in many ways.

So, we have different functional forms in econometrics which will help you to capture if you suspect some bit of non-linearity between 2 or more variables. So, in economics there are many such relations where you see that for larger or smaller values of X, the explanatory variable the values of dependent variable Y tends to converge to some value.

So, basically there is some kind of an asymptotic relationship between X and Y and it can be concave to X axis kind of curve where the curve you see is approaching a constant value of Y say Y tilde for very large values of X. There could be also a case like the Y value is falling and falling with increase in X values and it is converging to some value say Y dash and again that is an asymptotic relationship between Y and X. So, in this type of cases one can try this inverse functional form.

(Refer Slide Time: 15:36)



So, now in this slide we are going to talk about a particular type of non-linear functional form and that is called log linear model. This is probably the most popular functional form used for economic analysis. So, there we start with assuming some economic relationship which is non-linear in nature and that is given by capital Y equals to beta naught times X to the power beta 1. So, this is basically a non-linear relationship and I am showing you what sort of curve can be generated if I assume this kind of non-linear relationship between 2 variables.

So, now concentrate on the diagram in this slide and in the 1st diagram I am showing you the non-linear relationship between the two variables X and Y. So, here there could be two different cases. So, 1 is basically the case of beta 1 being negative and if that is the case then you get some kind of a rectangular hyperbola kind of a shape and that is in green in the diagram. But if beta 1 takes positive value then you can get an upward sloping curve.

And that kind of relationship is given by this orange colored or yellow colored or golden colored curve that you see here for beta 1 positive values. Now once you have this kind of non-linear relationship between X and Y of course your linear regression equation may not look a good fit to model this kind of non-linearity. But wait a minute there is a solution. So, what is the remedy?

I will show you that by taking simple transformation, a logarithmic transformation I can convert the non-linear model into a linear model. So, in the 2nd bullet point I am showing you the steps. So, suppose I start with the 1st in equation from the 1st bullet I take logarithm in both sides and it could be natural logarithm, it could be in log with 10 base does not matter. So, suppose here I am taking natural logarithm.

So, my equation will be now log of Yi equals to log of beta naught plus beta 1 times log of Xi plus epsilon i. So, here basically this is a log linear regression model as it is a regression model that I derived from the mathematical model in the 1st bullet I have added the stochastic term epsilon i here, so that the random fluctuations across data points can be modeled.

And here you see that log of beta naught can be further assumed to be new parameter which is like beta naught prime or something like that. So, now you have an equation of a straight line but now note that the variables are not in level values, they are the values after you take the logarithm. So, in the diagram you have to now measure log X along the horizontal axis and log Y along the vertical axis.

And if you do So, then actually you get straight lines. So, this red straight line which is downward sloping corresponds to the green curve that you see in diagram 1 and the sky blue colored upward sloping straight line corresponds to the golden color upward sloping curve that you figured out or found out in the 1st diagram. So, after that what to do? Of course we got a linear equation, so it is a model linear in parameters and linear in variables.

So, no problem we can go ahead with our OLS and after that the most important thing is know how to interpret. How can I make use of this model? And here comes the beauty of this model because we can have very nice interpretation from this log linear models. So, to derive or arrive at the interpretation of the slope coefficient, I told you previously that as an economist we are mostly interested in the interpretation of the slope coefficient seldom we are interested in the value of the intercept parameter.

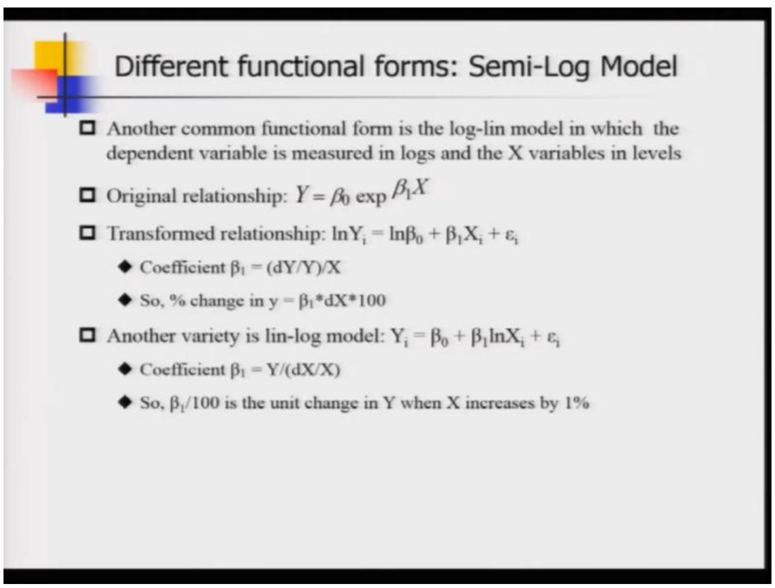
So, I am going to talk about the interpretation of the slope coefficient here. So, now let us start with some derivative. So, if you now concentrate on the 3rd bullet point I am showing you here that if I now have some derivative expression like, d of log Y dY then the result from the calculus formula and that is 1 over Y. So, I can rewrite this expression as d of log Y equals to dY divided by Y. So, what does that mean?

So, that gives me percentage change in Y. Now note that I can do the similar exercise for the variable X also or log X also. So, I can derive d of log X equals to d of X divided by X from calculus and then by that I get the percentage change in xs. Now you focus on the 2nd diagram or the log linear model regression equation that I have.

So, if I now take a derivative like d of log Y d of log X, So, basically I am trying to measure the slope of the log linear regression model equation then I get beta 1 and note that this beta 1 expression actually is the value for the derivative expression that is d log Y d log X. Now from the previous exercise that we know we did only couple of minutes before this derivative expression can be replaced by an alternative derivative expression which reads as dY divided by Y divided by dX divided by X.

So, basically now my beta 1 coefficient is giving me a measure of elasticity, So, why it is called elasticity? Because you know that in the numerator we are talking about percentage change in Y and in the denominator we are talking about percentage change in X. So, basically in total we can say beta 1 measures the percentage change in Y if there is a 1 percent change in my explanatory variable X.

(Refer Slide Time: 22:46)



Now we move on to another class of non-linear models and they are called semi-log model. So, in this category we know I am going to discuss two types of models. First I am going to discuss the case of log lean model. So, in the log lean model what we do. So, there the dependent variable Y is measured in logs and the X variable we keep that in the level values.

So, it emerges from an original mathematical relationship in a functional form I am showing here that reads as Y equals to beta naught times exponential of beta 1 times X. So, now if I take natural logarithm of both sides I get a transformed relationship and if I add epsilon value to take care of the random fluctuations in data then we have these regression equations to be estimated. And in this context I can get the coefficient beta 1 as d of Y divided by Y whole divided by X.

So, what does beta 1 gives me? If I want to derive some kind of interpretation for it from this derivative expression, So, it gives me the percentage change in my dependent variable Y if there is 1 unit change in my explanatory variable X. So, basically the percentage change in Y any particular value can be given by this expression beta 1 multiplied by dx the change of units or unit change in the explanatory variable X and that should be multiplied finally by 100.

Now you may be interested to know where can I apply this log lean model. So, suppose we want to find out the growth rate of some economic variable and data is given for different time periods. So, let Yt be the value of that economic variable at time t and Y naught is basically the initial value of that economic variable in the base period. So, we have done time series analysis, so, you know and you remember hopefully the term base period.

So, now we can assume that there is a compound interest type formula that is actually the relationship between Yt and Y naught. So, we know from our macroeconomic theory that if there is growth in some economic variable then we can have an equation like Yt equals to Y naught times 1 plus R to the power t where R is the compound over time rate of growth of the economic variable Y.

And in that case if I now take natural logarithm of this macroeconomic functional relationship then we can write log of Yt equals to log of Y naught plus t times log of 1 plus R. And now if I denote log of Y naught as beta naught and if I denote log of 1 plus R as beta 1 then you see I get a perfect simple linear regression equation which is a lean log model and the model can be written as log of Yt equals to beta naught plus beta 1 times t.

So, t is basically the trend variable here and of course I can add the stochastic term epsilon t to make it a perfect linear regression equation to be estimated. And once I get the estimates for beta naught and beta 1 in this model then how do I interpret beta 1 in this particular case? So, here the slope coefficient beta 1 measures the constant proportional or relative change in the Y variable for a given absolute change in the value of the regressor.

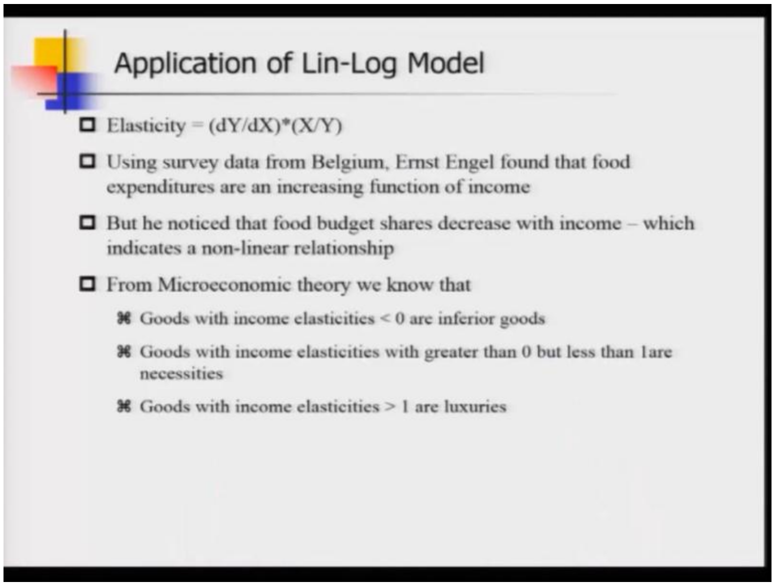
So, that is basically my time variable t. So, basically if there is 1 unit change in the time then what would be the relative change in my variable Y. So, now if I multiply this relative change in Y by 100 then basically we get what is called the percentage change or the growth rate of the economic variable Y. And that is why in the literature this beta 1 is known as semi elasticity of Y with respect to the explanatory variable. There could be another alternative form of the semi log models and that is called lean log model.

And in the lean log model we have the regression equation to start with as Yi equals to beta naught plus beta 1 times log of Xi plus epsilon i. So, here the coefficient beta 1 takes a different interpretation. So, here you see beta 1 is derived as Y divided by dx divided by X from calculus formula.

So, it says that if there is some 1 percent change in my explanatory variable X then by what units my dependent variable Y is going to change. So, in this context beta 1 has to be divided by 100 and then that is basically the unit change in Y when X increases by 1 percent. So, in the lean log model context I can say that this beta 1 coefficient can be seen as a ratio of 2 changes. So, in the numerator we will have the change in Y that is the change in levels of the Y.

And in the denominator we will have the relative change in X. So, now we are going to discuss a case or an application of semi log models. And specifically in particular we are going to look at an application of lean log models and that application I will bring from the economics literature and that is famous or popularly known as the Engel curve.

(Refer Slide Time: 29:36)



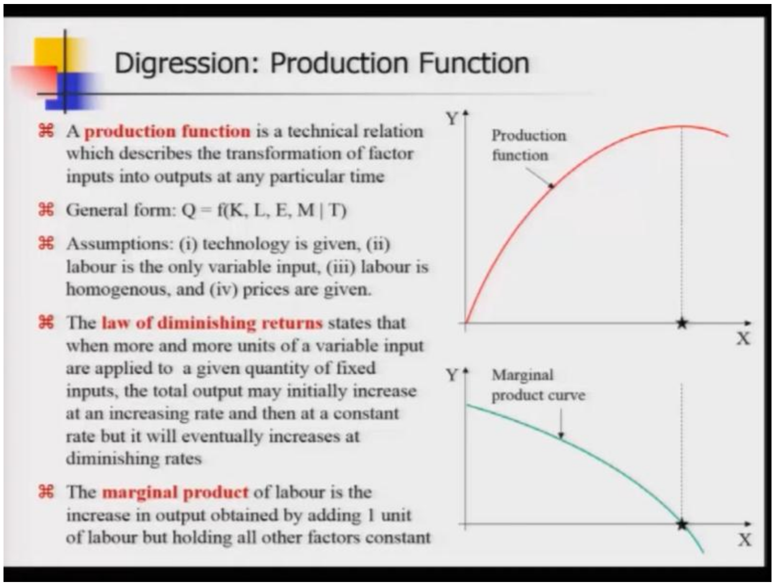
So, I will start by reminding you the elasticity formula and that is given in the 1st bullet point. Now we know we are talking about an economist of 19th century Ernst Engel. So, he actually first who got some survey data from households in Belgium and then he was interested to study the behavior of food expenditure as there is increase in income.

So, he noticed that these food expenditures are an increasing function of income but he also noticed that the food budget share decreases with income. So, that indicates probably a non-linear relationship between these 2 variables. So, how do I classify these goods and services that we see around us in these 3 buckets or categories called inferior, necessary and luxury.

So, microeconomic theory tells us that if I calculate the income elasticity and if I find that the income elasticity is negative then that good is called inferior. And if I find the income elasticity lies between 0 and 1 then we can say that the good is necessary good. And if I figure out that income elasticity is greater than 1 then we can say that the good in question is a luxury good.

Now I would like to mention that there are several other functional forms which are available there to help you to modern non-linearity between 2 variables Y and X. And of course in this lecture we do not have time to explore all sort of functional forms but I am going to mention about another functional form which is very useful in applied economic research and that is basically the form of polynomial.

(Refer Slide Time: 31:43)



In microeconomics the theory of firm says that there is a concept called production function. So, a production function is a technical relation which describes the transformation of factor inputs into outputs at any particular time. So, you can assume that there is a general mathematical function which is given by Q the quantity produced of a particular good Y and then that is a mathematical function of K that is basically the capital input, L that is my labor input, E that is my energy input, M is basically my raw material input.

And there note that I have also assumed that time or the status of technology is given. So, that is why I have use this given T expression within the parenthesis. So, note that if we want to talk about short run production function then we have to make certain assumptions. And these are very simple assumptions and most likely you are not going to contest these assumptions that much.

So, the first assumption says that the technology is given for a farm or a production unit and the second assumption says that labor is the only variable input that means that we cannot vary the levels of other inputs like K, E and M that we talked about in the previous expression. And we are also going to assume that labor is homogeneous. So, there is no quality difference in the units of labor and finally we assume that prices are also given. So, if these assumptions are all fulfilled then we have something which is known as the law of diminishing marginal returns or law of diminishing marginal productivity.

So, the law of diminishing returns states that when more and more units of a variable input are applied to a given quantity of fixed inputs, the total output may initially increase at an increasing rate and then at a constant rate but it will eventually increases at a decreasing rate. So, the marginal product of the variable input here in this case it is labor is the increase in output obtained by adding 1 unit of labor by holding all other factors constant.

So, these concepts of production function and marginal product of the variable input I am showing here through two simple graphs. So, look at the graph that you see at the north east corner of the slide the 1st diagram. So, there in the horizontal axis I am measuring the units of the variable input X and I am measuring the units of output being produced that is denoted by Y and that I measure along the vertical axis.

So, in this diagram I am going to show the stage 2 of the production function which is coming from the law of diminishing returns as proposed by professor Alfred Marshall. Why stage 2 because that is basically in fact not stage 2, I am also going to show you the stage 3. So, what is stage 1? So, stage 1 actually says that as the marginal product of the variable input is increasing at an increasing rate, so, the Y will increase at an increasing rate.

So, basically in stage 2 you see as X increases Y increases also and then Y reaches some highest value for some level of input as I denote this by asterisks here on the X axis and then as you keep on adding more and more units of X, actually Y does not increase actually it reduces. So, Y falls. So, the production function bends backward.

So, this is basically the typical neoclassical production function in microeconomic theory of firm. So, for that production function how do I get the marginal product function? Note that marginal product function can be mathematically derived by taking the 1st order derivative of the total product function. So, let us now go back to the diagram again it will be much clearer to you.

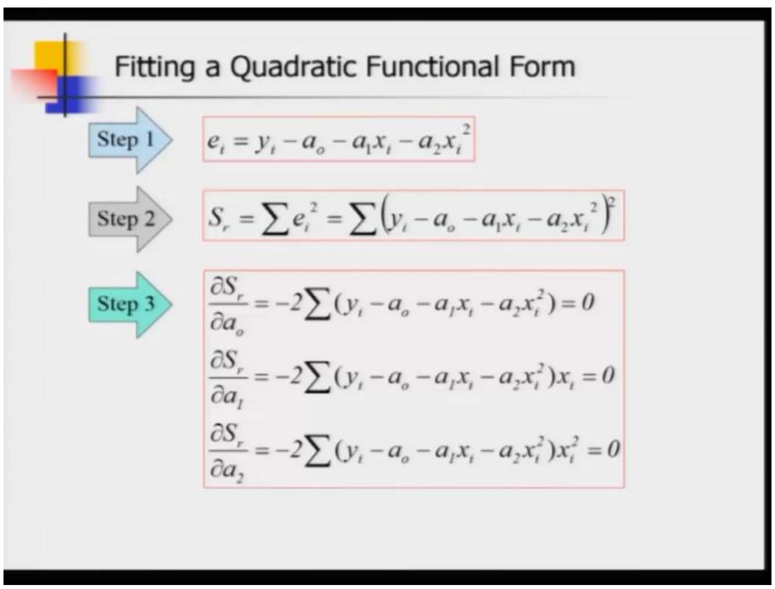
So, here we look at the production function the total product curve which is given in red color and now you see the if I change my variable input by 1 unit, the resulting change in Y is given by the slope of the total product curve. And slope of the total product curve or total production function curve means that we are talking about derivative.

So, now if I plot these derivative values for different levels of X, what I get? I get the marginal product curve and you see here that for that level of X where the total product curve reaches the highest level of production for that level the marginal product curve actually cuts the X axis. So, it takes value 0. So, if you keep on adding the units of input X the variable input then actually marginal product becomes negative and hence the total production function bends backward.

Now I am going to assume that I am going to model this kind of non-linear functional concept say production function by a quadratic functional form and let us see how we are going to make use of the quadratic functional form to model this kind of situation. So, simple quadratic functional form can be seen as a function Y equals to let us use some different notation because in different textbooks you are going to see different notations for slope and intercept coefficients. So, do not be surprised.

So, let me assume here that the slope coefficient is alpha naught and then I have alpha 1 times X plus alpha 2 times X square. So, that is basically the full quadratic functional form that I am going to use to model my production function. And then I am going to talk about how first the quadratic functional specification is estimated and then we are going to discuss an illustration. So, let us have a look at them.

(Refer Slide Time: 38:26)



Now note that in the previous slide we have represented the multiple regression models which involves more than 1 explanatory variables. So, far we have studied the case of least squares principle involving only 1 explanatory variable and 1 intercept term of course, but what do you do if you have got 2 explanatory variables and then you also have got the intercept term. So, basically you are talking about 3 unknown regression coefficients to be estimated. Can we still apply least squares principles? Yes, of course why not.

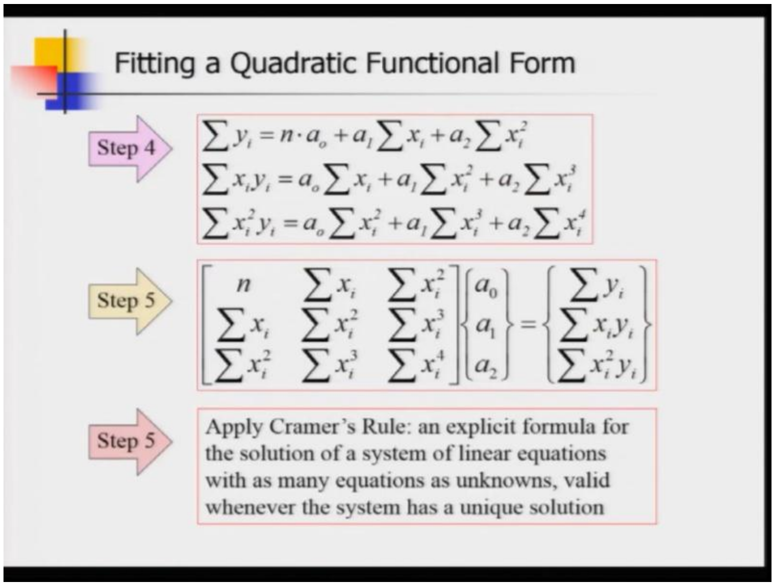
The least squares principle is very general in that way. It can be applied to k number of variables where k is greater than equal to 2 and it can also be useful when you have non-linearity in data because if you just write a linear in parameter regression equation that is good enough. And you can apply the least squares principle.

So, now in the next slide I am going to take you through the least squares principles again but this time for the quadratic functional form case. And I am also going to talk about 1 mathematical method called Cramer’s rule which is very helpful when you are dealing with more than 2 explanatory variables because then it will ease your mathematical calculations quite a bit.

So, as usual in the 1st step we define our residual. So, when we have a quadratic functional form the fitted value is, so, you have to take the difference between the original or the observed value of Y for the ith individual and the fitted value that comes from the regression. So, in step 2 then you need to find out the sum of square residuals.

So, that is also not new to you. In step 3 again you have to take the 1st order derivatives So, that you can write the 1st order conditions, for this minimization problem, you have to minimize the error sum of squares residuals and here note that in contrast to the previous case now we have 3 unknown variables a naught, a1 and a2 and that is why you need three 1st order conditions you set them equal to 0, all three of them.

(Refer Slide Time: 40:57)



So, now after step 3 you have to rewrite these 1st order conditions in 3 normal equations such that you keep all known quantities in one side and you keep all unknown quantities on the other side. So, here if you look at the red box corresponding to step 4 you will see that I have kept all the known quantities, known means that the values from the sample and all these known quantities are kept in the left hand side and in the right hand side I have kept all the expressions for which I do not know the values.

So, here I do not know the values of a naught, a1 and a2. So, that is how you see I have kept all a naught, a1, a2 involving expressions in one side of the equation. So, now you have 3 equations in 3 unknowns. So, you can solve these 3 normal equations to find the values a naught, a1 and a2. So, when you have more than 2 explanatory variables one can make use of method from matrix algebra linear algebra and that is called the method of Cramer's rule to find the solutions of unknown quantities from the linear equation of systems.

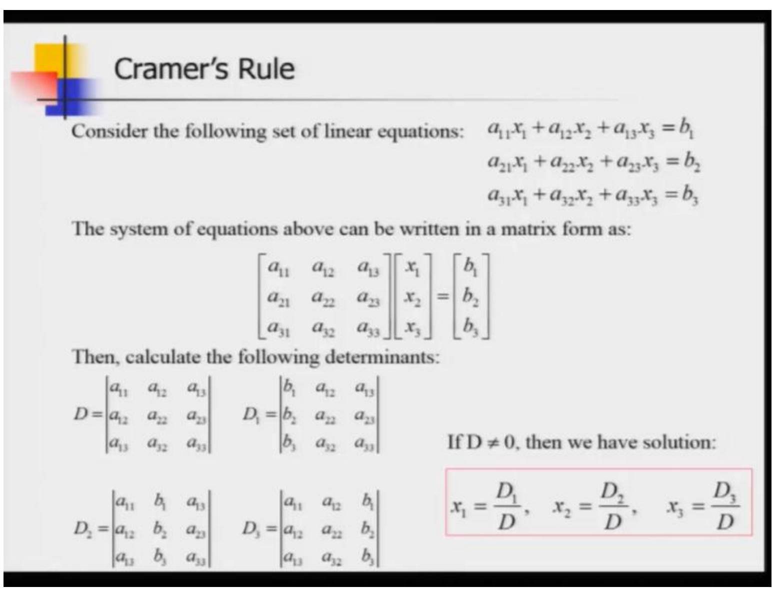
So, here in this slide step 5 and step 6, I am going to talk about that Cramer's rule and in step 5 now you see what I am doing. I am making use of step 4 equations. Now I am writing this in linear system of equations into matrix form and you see the 1st matrix which is a 3 by 3 matrix. It talks about the quantities that I can compute from the data. So, that is basically my 1st 3 cross 3 matrix.

Then this matrix is multiplied with the column vector and here I have 3 unknown quantities that I need to solve. So, namely a naught, a1 and a2 and then if you multiply this matrix and this vector. So, the matrix is 3 cross 3 and the column vector is 3 cross 1. So, of course you can get the 3 cross 1 matrix again as the output of this matrix multiplication or outcome of this matrix multiplication.

And these 3 components will be equal to the components in the column vector that you see in the right hand side of this equation and here the quantities are again known to me because I can calculate to the data. So, now in step 3 you can actually apply Cramer’s rule. So, what is it? So, it is an explicit formula for the solution of a system of linear equations with as many equations as unknowns.

And note that this Cramer’s rule is a valid method only when the system has a unique solution. So, for many of you Cramer’s rule may seem very Greek because you may not have done that in your schools or college level studies. So, I am going to spend couple of minutes to describe the Cramer’s rule method in some detail of course I cannot give you examples etc. So, I leave it to you. You can develop further based on this simple introduction that I am giving you but you must know how to calculate determinant of a matrix.

(Refer Slide Time: 44:37)



So, here in this slide I am going to show you how Cramer’s rule works. So, let us start with the following set of linear equations. You see I have started with a very general set of equations and there are 3 and I assume that the values of a11 a21 a31 these coefficients and I also know the values of b1, b2 and b3.

The only things that are unknown to me are basically the variable values x1, x2 and x3. So, I need to solve for x1, x2 and x3. Now this system of equations can be written in the matrix form as I have shown in the previous slide. So, I want to prepare a matrix involving numbers that are coming from the set of linear equations as note that these are basically the coefficients a1 a21 a31 etc. and then the column vector will consist of the unknown quantities that I need to solve here namely they are x1, x2, x3.

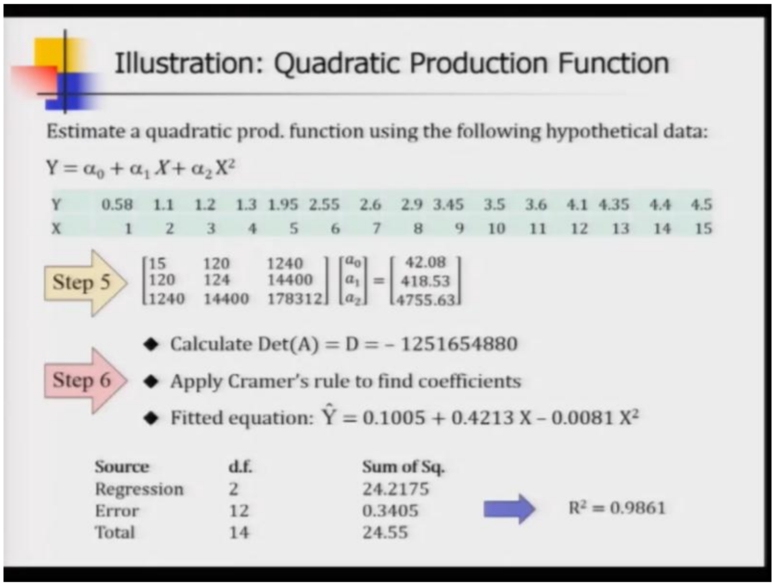
And then the outcome of this matrix multiplication will be equal to again one column vector and this has the elements b1, b2 and b3. And note that I also know the values of b1, b2 and b3. So, in this context the Cramer's rule suggests us that we can actually find the values or optimal solution for x1, x2, x3 by calculating some determinants of the matrices. And I am going to show you the 4 determinants that you need to calculate.

First you have to calculate the mega determinant or that is denoted as D and this determinant is basically calculates the determinant for the coefficient matrix involving a11 to a33 and then basically you have to calculate another determinant say D1 and now note that what I have done. So, D1 determinant actually is going to help us to find the optimal solution for x1. So, for that what we will do?

So, we will now replace the 1st column in the 1st determinant formula capital D by the values of beta vector. So, b1 b2 and b3 and the other coefficients I will keep them intact. And now I calculate the determinant value and I get the value for D1. So, now note that for if we now want to solve the variable x2 then we have to calculate another determinant D2 and here as I am trying to solve for x2 I will not tamper the 1st column and the 3rd column of the determinant expression D, I will change only in the middle column/

I will replace all the coefficient values by the column vector values b1 b2 and b3 and then calculate the determinant. And similarly I will follow the same strategy for solving x3. So, here as I am solving x3, I will not change the columns first 2 columns of D determinant I will only change the last column by b1 b2 b3 and I will calculate the determinant value. And you see that if the determinant of the main coefficient matrix which is D is not equal to 0, I can now write the solution for x1 x2 x3. So, here I am showing you the formula in the box.

(Refer Slide Time: 48:35)



So, now we know we are at the final slide of today's lecture. So, here I am going to show you an illustration how we can fit the quadratic production function. So, here is a hypothetical dataset. So, there are 15 data points coming from a hypothetical firm and if we apply 1st order derivative to this fitted equation then what will be the equation for my marginal product of variable input x?

So, the intercept is constant, so, it will fall out from the equation and then I will have 0.4213 minus 2 times the slope coefficient for the variable X square. So, finally I will have minus 0.016. And if you plug the sum of squared values into the R square formula you will get the value of R square equal to 0.9861.

And again I suggest you to try your hands within numbers and then match your calculated value of R square with the R square value that I am reporting here in this slide. So, this is it for the time being. In the next lecture we are going to continue our discussion on linear regression analysis. Thank you.