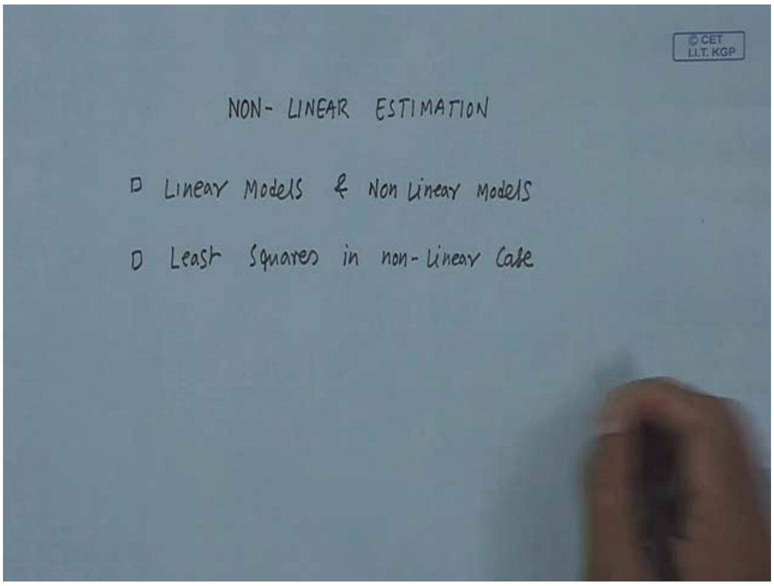
Regression Analysis
Prof. Soumen Maity
Department of Mathematics
Indian Institute of Technology, Kharagpur

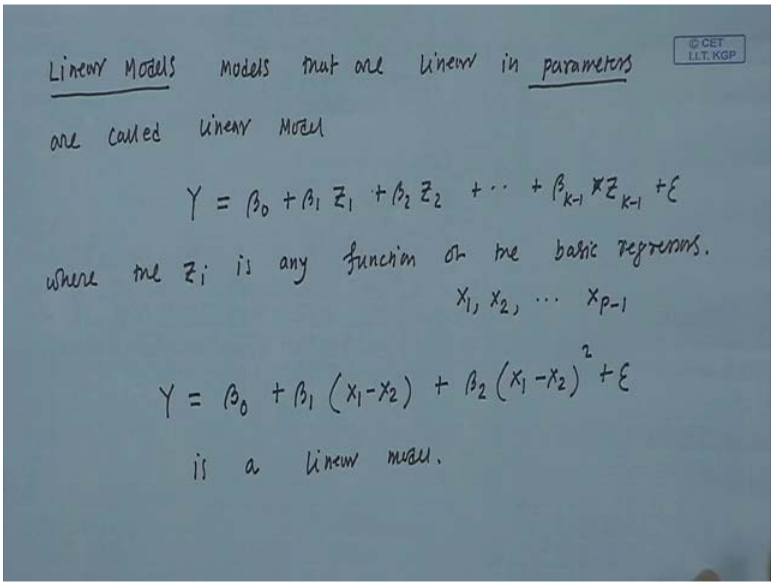
Lecture - 32
Non-Linear Estimation

(Refer Slide Time: 00:27)



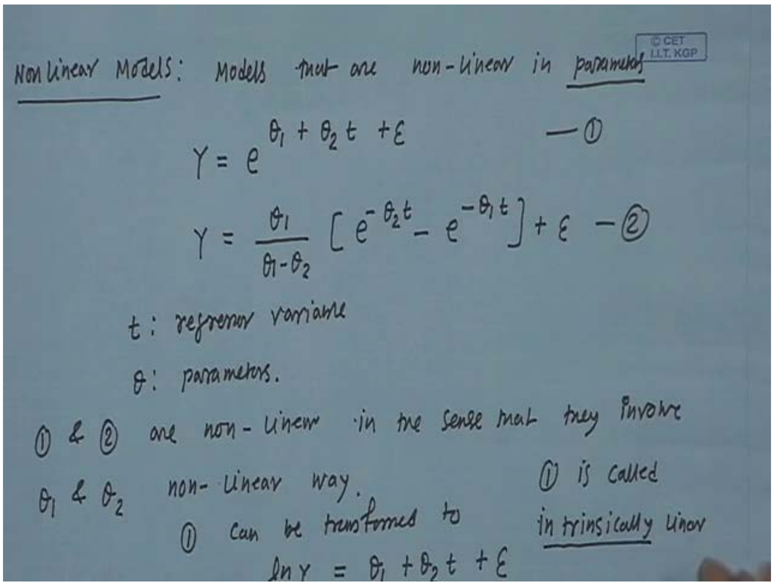
Hi, today we will start non-linear estimation, and here is the content of this topic linear models and non-linear models. So, we will give detailed definition of non-linear models and then we will talk about least square in non-linear case, so how to fit non-linear model using least square technique. So, given a set of observation say x i, y i, the starting point is that we start with the simple linear regression model or if we have more than one regressor we start with multiple linear regression model. Of course, those models are linear model. Also you must have realized that the polynomial regression is also a linear model. So, let me give the detailed definition of non-linear model and also linear model.

(Refer Slide Time: 01:43)



Linear models, models that are linear in parameter are called linear model, this is important linear is in parameters not in the regressor. The model for multiple linear regression is beta naught plus beta 1 z 1 plus beta 2 z 2 plus say beta k minus 1 z k minus 1 plus epsilon. Where, the z i is any function of the basic regressor variables. What I mean by this, the basic regressors are say x 1, x 2 say x p minus 1. Then y is equal to beta naught plus beta 1 into x 1 minus x 2 plus beta 2 into x 1 minus x 2 square plus epsilon. So, this is a linear model. Because here, z 2 could be any function of the basic regressors. But, this model is linear in parameters, the parameters are beta naught, beta 1 and beta 2. So, this model is called linear models.

(Refer Slide Time: 04:28)

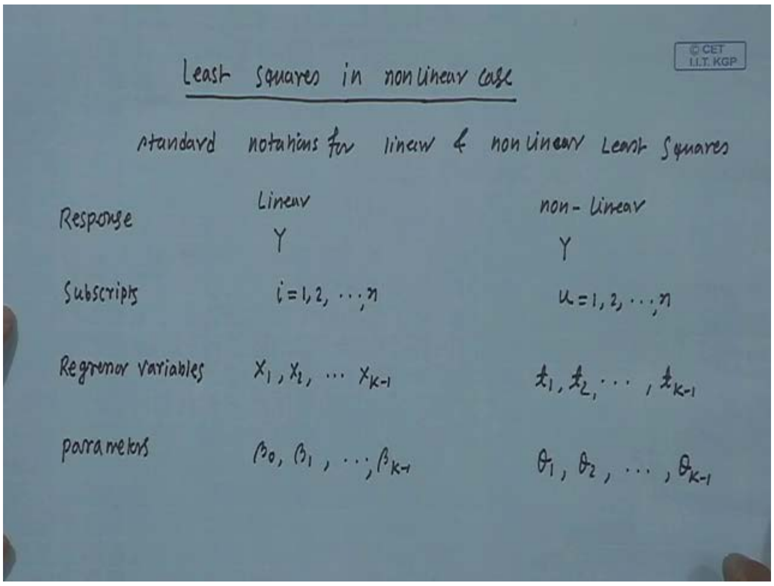


Non linear models, models that are non-linear in parameters are called non-linear model. Let me give examples, Y is equal to e to the power of theta 1 plus theta 2 into t plus epsilon. So, here y is the response variable and in non-linear case generally we represent the regressor variable by t instead of x. So, this is a non-linear model because it is non-linear in theta 1 and theta 2.

Let me give one more example, say y is equal to theta 1 by theta 1 minus theta 2 into e to the power of minus theta 2 t minus e to the power of minus theta 1 t plus epsilon. So, this is very clear that this model is non-linear in parameters, parameters are theta 1 and theta 2. So, these are the non-linear models and here t is regressor variable and theta are parameter. So, let me call it say 1 and 2. 1 and 2 are non-linear, in the sense that they involve the parameters theta 1 and theta 2 in non-linear way. That is why these two models are called non-linear. And now you can see that 1 can be transformed to say log base e, so l n y is equal to theta 1 plus theta 2 t plus epsilon.

So, once you take this transformation this non-linear model becomes linear, linear on the data l n y and t. This type of non-linear functions are called intrinsically linear. So, 1 is called intrinsically linear, because you can very easily transform this non-linear model to a linear model by taking some transformation. Well, so we understood what is linear model and what is non-linear model.

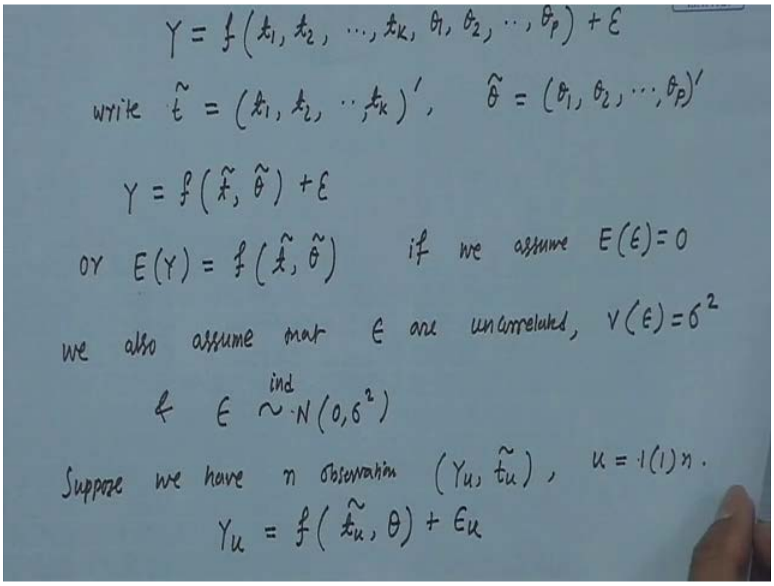
(Refer Slide Time: 09:08)



And, this is the notation for the linear and the non-linear case. So, given a non-linear model we are going to estimate the parameter of the non-linear model. Here, are some notations, you can see that for response variable in the linear case we represent it by y. So, y generally stands for the response variable and in non-linear case also y stands for response variable, and subscripts for linear case it is i is from 1 to n, and then here you know just notation, and here we will use instead of i, that is instead of calling y i we will say y u and u is also from 1 to n, so n observations.

And, in case of linear models we use the regressor variables standard notations x 1 x 2 x k minus 1. But, here we will be using t 1 t 2 t k minus 1. And in case of linear the parameters are beta naught, beta 1 and beta k minus 1 and in case of non-linear we use theta 1 theta 2 and theta k minus 1, this notation you know and to make your life difficult. Well, now let me start with the non-linear model.

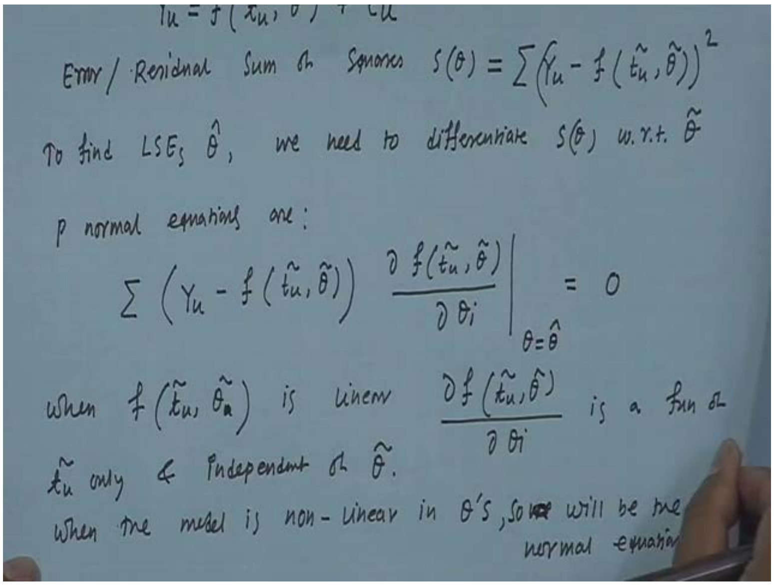
(Refer Slide Time: 10:48)



Say, general model y equal to f t 1 t 2 t k and theta 1 theta 2 up to theta p plus epsilon. So, this is non-linear model, f is non-linear in the parameters theta 1 theta 2 and theta p, and in vector notation we write t equal to t 1 t 2 t k and theta vector is theta 1 theta 2 theta p. And then in terms of this vector notation, we can write y equal to f t theta plus epsilon, or also we can write, expectation of y, this is the non-linear model we are talking about. Expectation of y is equal to f t theta, if we assume expectation of epsilon is equal to 0.

We also assume that epsilon are uncorrelated and also variance of epsilon is equal to sigma square, and epsilon vector follows normal 0 sigma square so independent. So, these are the basic assumption we do in case of simple linear regression model, simple or multiple regression model also, we are making the same assumption for the non-linear model. And suppose we have n observations like y u t u, like we had before you know y i x i for i equal to 1 to n. Here, notation difference is u from 1 to n. Then we can write y u for the u th observation is equal to f t u theta plus epsilon u.

(Refer Slide Time: 14:57)



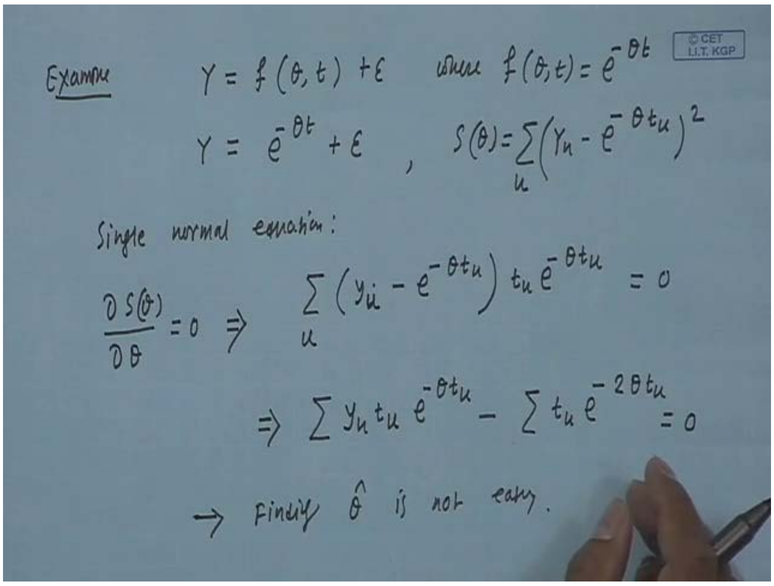
Now, we have the model say y u equal to f t u theta plus epsilon u and we have to fit this model. Fitting this model means you are given a set of observations 1 to n and you have to estimate the regression coefficients. So, by using the least squared techniques we consider the residual or error sum of squares, call it, s theta is equal to y u minus f t u theta whole square, basically you know if you put hat here this becomes the observed response and this is the estimated response. So, this is the uth residual and residual square is s theta, so this is the quantity you want to minimize in order to find the least squared estimates.

To find least squared estimates theta hat, we need to differentiate s theta with respect to theta, theta is a vector, it has p components. If you differentiate with theta 1 theta 2 theta p, from there you will get p normal equations, p normal equations are summation y u minus f t u theta. Differentiating this function with respect to theta i and then the partial derivate, this involves theta i, so the partial derivate of this one f t u theta with respect to theta i and that at the point theta equal to theta hat.

This is the point you want to estimate so this equal to 0. This is the ith normal equation and if you differentiate with respect to theta j you will get the jth normal equation, this way you will get p normal equations. And now you should realize that when f t u theta is linear then this partial derivate f t u theta by delta theta i is a function of t u only. If it is linear in t i then when you differentiate it does not involve theta i. So, it is a function of regressors only and independent of theta or specifically independent of theta i.

When this model is non-linear, means f is non-linear. When the model is non-linear in thetas, so will be the normal equations. If the model is linear then the partial derivatives are independent of theta but if the model is non-linear then the partial normal equations are also non-linear.

(Refer Slide Time: 21:00)

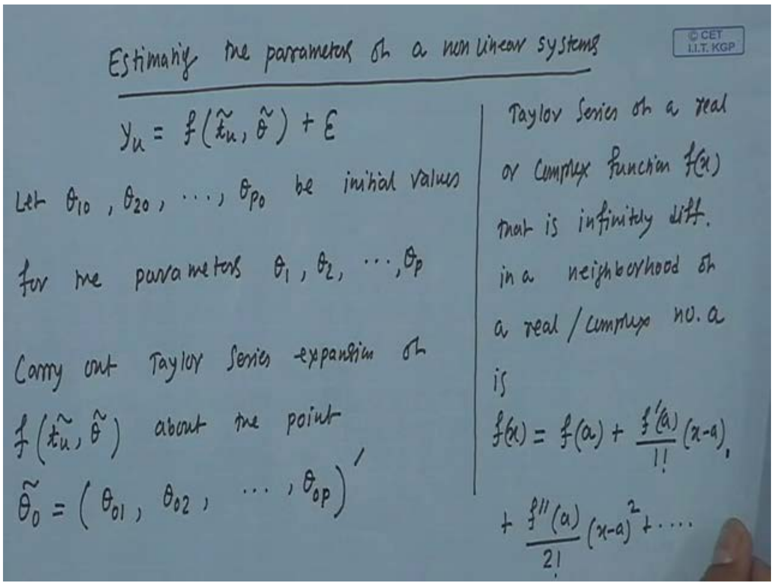


Let me give you an example, consider this model, say y equal to f theta t plus epsilon, where f theta t is equal to e to the power of minus theta t. That means we are considering the model e to the power of minus theta t plus epsilon, so this is the model. So, what is my s theta here, my s theta is equal to y u minus e to the power of minus theta t u, its only one regressor say t, this is whole square sum over u, this is my s theta and we want to minimize, we want to estimate theta in such a way that this is minimum.

That means, the single normal equation is obtained by differentiating s theta with repect to theta this equal to 0, implies summation y u minus e to the power of minus theta t u and then the derivative of this one with respect to theta, so this is t u into e to the power of minus theta t u, this is equal to 0 and sum over u. So, this gives that you know, your normal equation is y u t u e to the power of minus theta t u minus t u e to the power of minus twice theta t u equal to 0.

So, what I want to say here is that if the model is non-linear and the normal equations are also non-linear, this is considered a normal equation and it is non-linear in theta. So, finding theta which satisfy this equation is not easy. So, finding theta hat, means you know you solve this the theta you got is the theta hat, so this is not easy here. Now, how to estimate the parameters of non-linear systems.

(Refer Slide Time: 24:22)

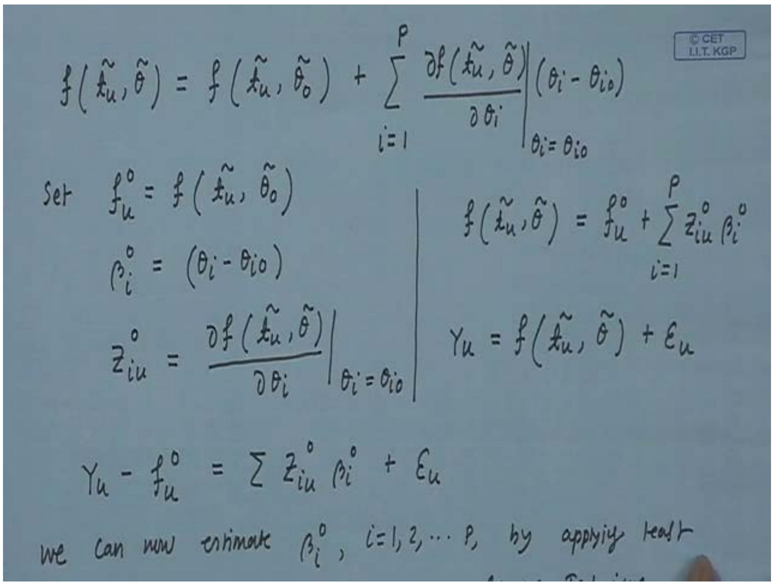


Estimating the parameters of a non-linear systems, Notation ally it may look very difficult but the idea is simple here. So, I am given a model like this, what I am given is that, I am given the model y u equal to f t u theta plus epsilon. This is non-linear in theta and the idea here is that to estimate the parameter, as you must have observed that if the model is non-linear then the normal equations are also non-linear and then solving the non-linear systems equations are difficult.

So, what we will do is that we will approximate this non-linear function by linear functions using Taylor series. This is non-linear and we will approximate this one by a linear function, for that let me talk about the Taylor series. As you know Taylor series of real or complex function f x that is infinitely differentiable in a neighborhood of real or complex number a is f x equal to f a plus f prime a by 1 factorial into x minus a plus f double prime a by 2 factorial into x minus a square. So, here you are expressing a function, any function in terms of polynomial.

So, here what we will do is that, similarly we will approximate this non-linear function by a linear function, so we will take up to this term using the Taylor series. Let theta 1 0 theta 2 0 and theta p 0 be initial values for the parameters theta 1 theta 2 theta p, and then we carry out this Taylor series expansion about this initial value. So, you carry out Taylor series expansion of this non-linear function f t u theta in the neighborhood of this point, about the point say theta naught, which is equal to theta naught 1 theta naught 2 theta naught p. We will carry out the Taylor series expansion of this non-linear function about the point or in the neighborhood of this point. So, we will take up to this term because we are looking for a linear approximation for this non-linear term of this non-linear function.

(Refer Slide Time: 30:04)



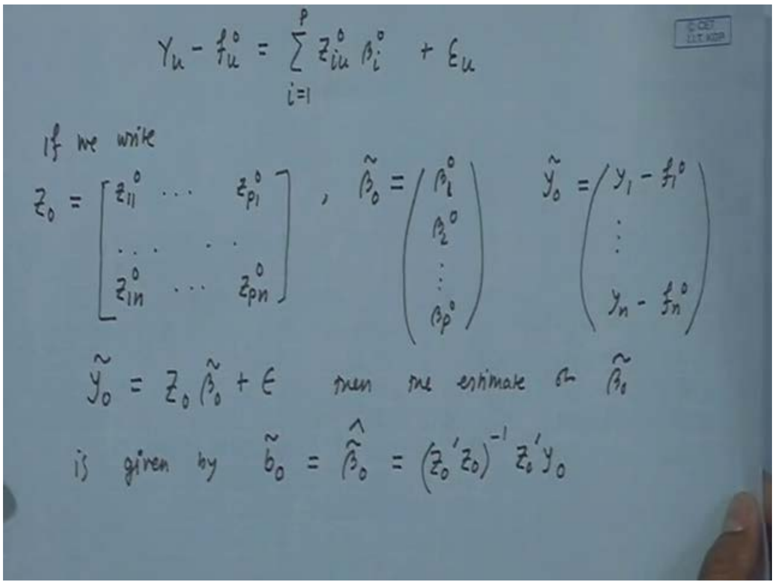
So, here is the Taylor series expansion, f t u theta is equal to f t u theta naught plus the partial derivative of this f t u theta by delta theta i, theta i minus theta i naught and this derivative at the point theta i is equal to theta i naught, and this i is from 1 to p. You must have understood that this is the Taylor series expansion of this non-linear function up to the first of the second term, linear term.

Now, let me use some more notations, set f u naught is equal to f t u theta naught, and this term we will denote by beta i naught which is equal to theta i minus theta i naught, and then this term we will denote by z i u naught, so that is nothing but partial derivate of this one theta at the point theta equal to theta i naught.

Now, if you use all this notations in this linear approximation you can write this one as say, f t u theta, so non-linear function is equal to f u naught because f u naught is this plus z i u naught beta i naught and i is from 1 to p. So, this is the linear approximation of the non-linear function in theta so here it is in terms of beta. So, y u this is the model we started with, y u was f t u theta plus epsilon u, now if i plug this linear approximation of this non-linear function here, what I will get is that, I will get y u minus f u naught is equal to z i u naught beta i naught plus epsilon u.

Now, you can see that this is a linear model in beta, this is same as multiple linear regression model. So, in this model we can estimate this parameter, the transformed parameter beta i naught using the least squared technique. So, we can now estimate this beta i naught for i equal to 1 to p by applying least square technique. So, let see how to do this thing.

(Refer Slide Time: 35:41)

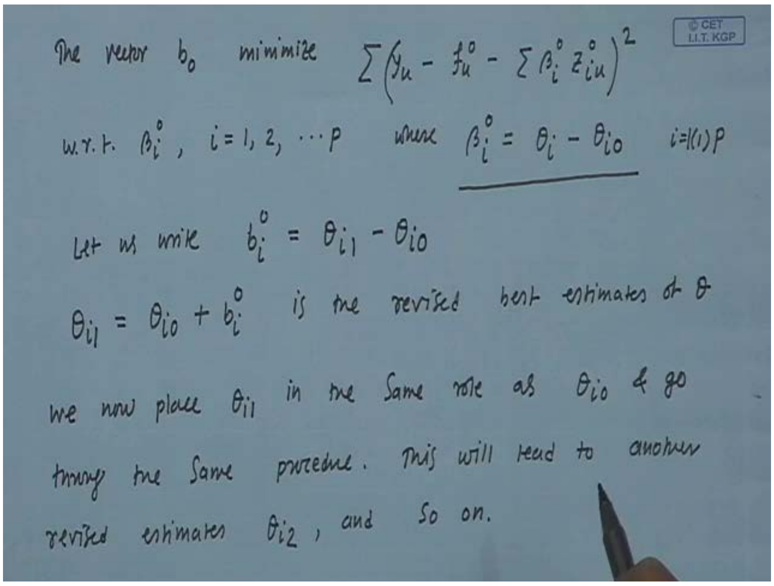


We have the linear model y u minus f u naught equal to z i u beta i naught, i is from 1 to p plus epsilon u. So, we started with the non-linear model and then we have a linear model by using the Taylor series approximation. And we are trying to solve this linear model now, so we can write in matrix form, if we write my coefficient matrix z naught which is equal to say z 1 1 naught z p 1 naught and then z 1 n naught and z p n naught, so this is my coefficient matrix, and beta naught is my coefficient vector, that is beta 1 naught beta 2 naught and beta p naught and let me write my response as y vector, so y vector is y 1 minus f 1 naught and then y n minus f n naught.

So, using this matrix notations I can write my simple linear, I mean this basically is a multiple linear regression model. Now, I can write it in the form y naught is equal to z naught beta naught plus epsilon, then the estimate of this beta naught, we know, now we can apply a least squared technique of this one is given by beta naught, which is equal to basically beta naught hat is equal to z naught prime z naught x prime x inverse, x prime y. So, here z naught is nothing but x in the multiple linear regression model so z naught prime y naught.

So, what we have now is that, we have the, let me write these in the vector notation only. We have the fitted model, so the fitted model is, I mean. Of course, this is not the final estimate, we wanted to estimate theta and now we are estimating beta, where beta is sort of theta minus theta naught. And what we will do is that we will improve this estimate iteratively.

(Refer Slide Time: 39:52)



The vector beta naught, which is estimate of b naught, which is estimate of beta naught, this minimize y u minus f u naught minus beta i naught z i u naught square. This is nothing but s theta basically with respect to beta i naught, so i is from 1 to p. Where, this beta i naught is equal to theta i minus theta i naught.

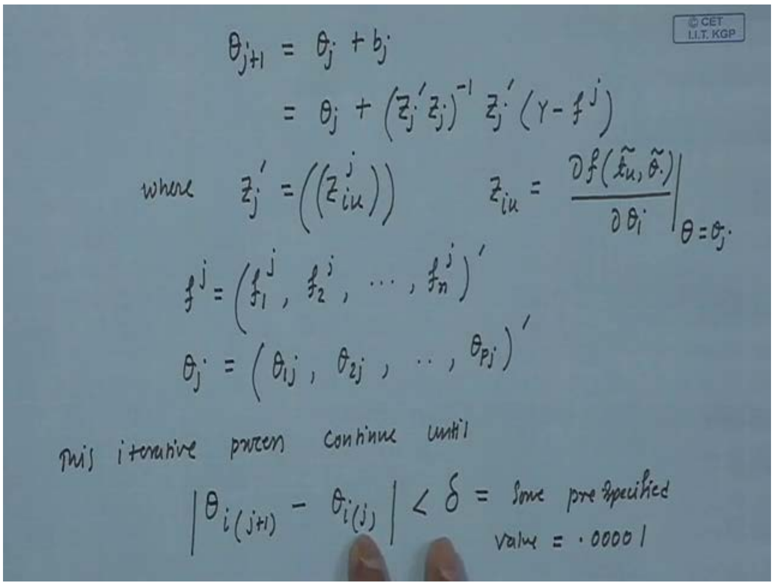
We need to understand this part, we want to estimate theta i's or theta 1 theta 2 theta p, so this i is from 1 to p. We started with some theta naught, we want to estimate theta, and then we have used this Taylor expansion about this point theta naught to make the non-linear function linear. And after making the function, from non-linear to linear, we have estimated this difference, so we have estimated this beta i naught and that is nothing but b i naught.

So, let us write b i naught is equal to theta i minus theta i naught. We started with theta i naught as a estimate of theta i and then we see that this quantity, this difference which is nothing but beta i naught and we have estimated this difference. Now, what I am trying to say is that, you know we started with the initial point and then we will try to improve this estimate iteratively. So, let me put a 1 here, this keeps my estimate of theta i at the first iteration, so theta i 1 is equal to theta i naught plus beta i naught. We started with theta i naught then we improved this theta i naught by theta i 1 and this is the revised best estimates of theta.

Now, again we will place theta i 1 in the same role as theta i naught and go through the same procedure. So, this will lead to another revised estimate, theta i 2 and so on. So, we started with theta naught. Let me just give little idea about what I am doing here, is that, we have a non-linear model and that is a non-linear in theta. First, what we do is that we take initial estimate of theta, that is, theta naught and then we consider the Taylor series expansion of the non-linear function about theta naught and we approximate the non-linear function by a linear function.

And once we have the transformation from non-linear to linear, we can use the result of simple linear regression model. And we estimate the linear model and from that estimate of linear model using the least squared technique, what we get is that from theta naught we get theta 1. Now, we put this theta 1, the revised estimate in the same role as theta naught was initially. So, we will do the same thing again, we take the Taylor series expansion of the non-linear function about theta 1. And then once you have the linear function, once you transform the non-linear model to linear model, we can apply the results of linear regression and from theta naught from theta 1 you will get theta 2 and so on.

(Refer Slide Time: 47:14)



At some point after the jth iteration, what we will get is, that we will get theta j plus 1 equal to theta j plus beta j. So, in the j plus 1 iteration we improved theta j by using the same technique and improved on other revised one is that theta j plus 1. So, this one is nothing but theta j plus this b j, we know what is this b j, b j is z j prime z j inverse z j prime y minus f j. Where, this z j prime is equal to z i u to the power of j, I mean the notation not to the power of j. I am sure that you understand what is this, this one is z i u is the derivative of that non-linear function t u theta with respect to theta i and theta equal to theta j. So, j result of the jth iteration.

And my f j is equal to f 1 j f 2 j f n j prime and my theta j is equal to theta 1 j theta 2 j theta p j. So, you understood from every iteration we are improving the estimation and there should be some stopping criteria, so when the result of jth iteration is not much different from the jth plus oneth iteration we stop there. This iterative process continue until this difference is i, this is the result of j plus oneth iteration minus theta i j, jth iteration, when this is less than some delta which is a small quantity, some pre specified value say for example, 00001.

So, when you see that the difference between the result obtained from jth plus oneth iteration and the jth iteration is very small, very significant difference. That means you can stop there. This is about the non-linear estimation. So, you understood what is the non-linear model, non-linear means non-linear in parameters theta 1 theta 2 theta p and given a non-linear model, you now know how to approximate that non-linear model by a linear model using the Taylor series expansion. And you also know how to estimate the parameters of non-linear estimation, non-linear model, using the least squared technique. That is all for today.

Thank you.