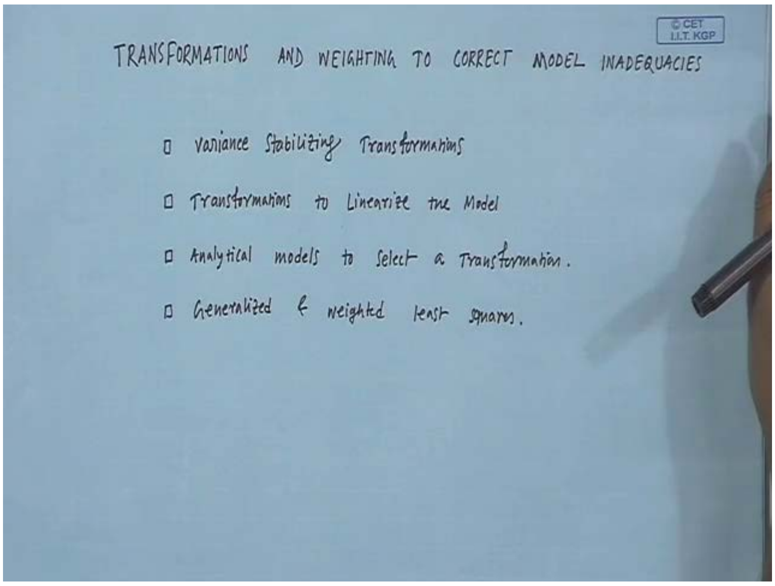
**Regression Analysis**
**Prof. Soumen Maity**
**Department of Mathematics**
**Indian Institute of Technology, Kharagpur**

**Lecture No. - 22**
**Transformations and Weighting to Correct Model Inadequacies (Contd.)**

(Refer Slide Time: 00:39)




Hi, this is my second lecture, in the topic called transformations and weighting to correct model inadequacies. In simple linear regression model or in the multiple linear regression models, we make some basic assumptions on the regression model. Like we assume that, the error terms have a mean 0, constant variance and they are uncorrelated and also, we assume that the error terms follow normal distribution. Now, given a set of data, how do you know the given data follow or dissatisfy the basic assumption we made for regression fitting? So, to check whether your data satisfy the basic assumption or not, we have learnt several techniques in the module called model adequacy checking. And in this regard the residual plot is an effective technique to test whether the model assumptions are satisfied or not.

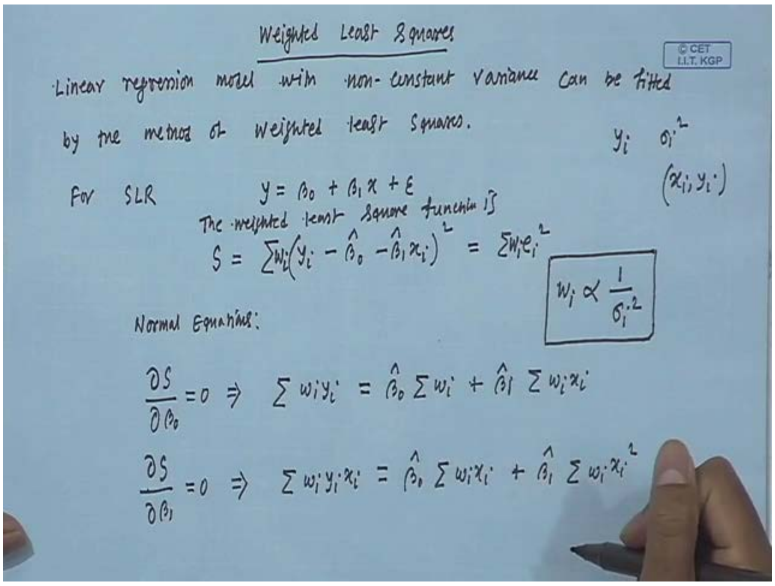
So, what you do there is that you fit a simple linear regression model and then you find the residuals and you plot residuals against the estimated response. So, if you see that the residuals are sort of scattered or centered about the line e equal to 0. Then the model is satisfactory and then we can assume that the constant variance assumption is correct.

But, if you see from the residual plot that, the residuals are short of the residual fit is similar to say, outward open funnel or inward open funnel. That means the constant variants assumption is not true and similarly, if you residual plot is similar to say double bow, then also the constant variance assumption is not correct.

So, if you have a data x i and y i and you know how to check whether a data satisfy the basic assumption or not. Suppose, your data does not satisfy the basic assumption we made in the linear regression fit. Then what we do in the current module or in the current topic is that, we learn here what to do in this situation when the given data does not satisfy the basic assumptions.

So, we talked about two techniques in the previous class and here is the content of this topic called variance stabilizing transformations. So, we take some transformation on the response variable to correct the non constant variance assumption. And also, we talked about transformations to linearize the model and we are left with analytical methods to select a transformation and a generalized and weighted least square. So, today we will be talking about this generalized and a weighted least square technique.

(Refer Slide Time: 05:11)




So, first let me talk talk about weighted least square. So, linear regression model with non-constant variance can be fitted by the method of weighted least square. So, if you can recall that for the simple linear regression model y is equal to beta naught plus beta one x plus epsilon. So, what you do to estimate? What is the ordinary least square? So, we are talking about weighted least square here. The ordinary least square is a technique to estimate this regression coefficient and there we estimate the regression coefficient beta naught and beta 1 by minimizing this quantity S equal to summation y i minus beta naught minus beta 1 x i whole square.

So, this is the observed response and this is the fitted response and this quantity is the residual. So, this is nothing but the i-th residual e i square. So, in the least square technique we minimize this quantity, we minimize the difference between the original response value and the estimated response value to estimate these regression coefficients. Now, what we do here is that instead of minimizing this quantity, we minimize this is called for weighted least square. The weighted least square function is just to multiply by W i.

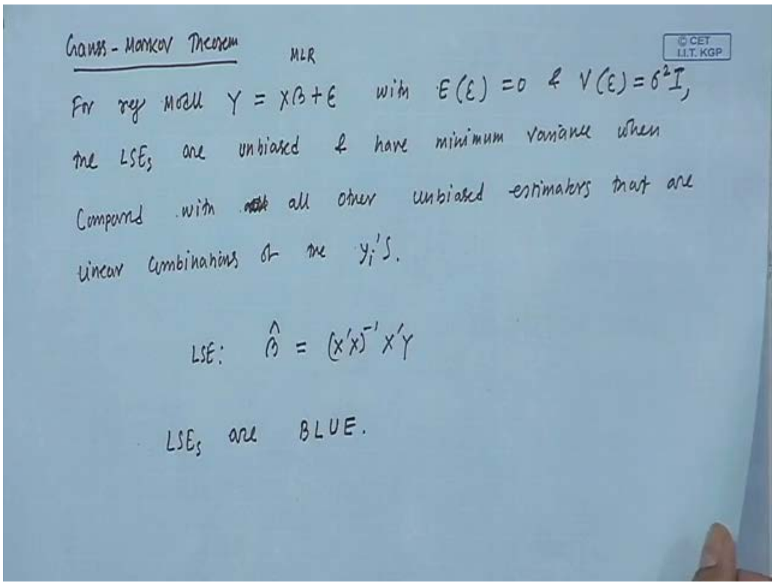
So, w i. is the weight given to the i-th observation and here w i is proportional to 1 by sigma i square. So, see here we are talking about with a model with one constant variance that means y i is coming from a population with variance sigma i square. So, w i is proportional to 1 by sigma i square and then you find the normal equation. So, you have to estimate beta naught and beta 1 by using weighted least square. The normal equations are partial derivative of S with respect to beta naught equal to 0.

This gives summation w i y i equal to beta naught hat summation w i plus beta 1 hat summation w i x i. This is the first normal equation and the second normal equation is obtained by a differentiating this weighted least square function S with respect to beta 1 equal to 0. This gives summation w i y i x i equal to beta naught hat naught summation w i x i plus beta 1 here summation w i x i square. So, you have 2 normal equations and 2 unknown, beta naught and beta 1. By solving these two equations you will get an estimate of the regression coefficient beta naught and beta one hat, beta naught and beta 1.

Now, I did not say anything about why the weight is proportional to 1 by sigma i square? And the other concern here is that, see you are given just a data set like x i y i for i equal one to n nothing else. So, you do not know what is this sigma i square for your given set of data? So, I will come back to this point.

Then first of all why this weight is proportional to 1 by sigma i square and how to get this sigma i square? So, I just give an idea of what is a weighted least square. So, this is a very similar to the ordinary least square, but here we are giving weight w i to the i-th observation and that w i is proportional to 1 by sigma i square. But, how do I get this sigma square? And why this  weight is proportional to 1 by sigma i square? That I am going to tell in this class only ok. Well now, I need some prerequisite to talk about generalized least square.

(Refer Slide Time: 13:19)

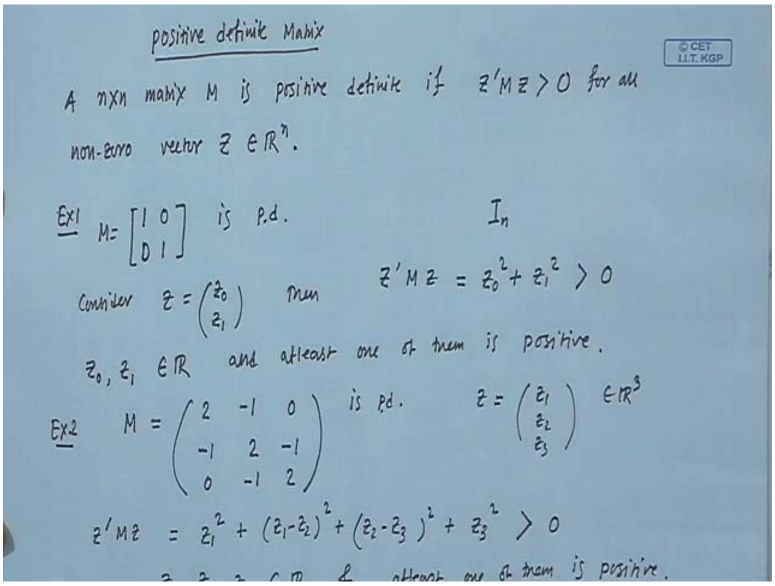



So, first one is a called Gauss-Markov theorem so, this theorem says that for regression model Y equal x beta plus epsilon. So, this is a multiple linear regression model with expectation of epsilon equal to 0 and variance of epsilon equal to sigma square i. That means it satisfy the basic assumption of multiple linear regression model. Its constant variation and the expectation are 0 and they are uncorrelated. You can see the covariance terms are all equal to 0, because this is the identity matrix ok.

For regression model with this, the least square estimate are unbiased and have minimum variance when compared with all other unbiased estimators that are linear combinations of the response value y i's. So, what this Gauss-Markov theorem says is that, for the multiple linear regression models satisfy the basic assumption. So, we apply the ordinary least quite techniques here, we know the estimate. We know that using least square estimate, beta hat is equal to x prime x inverse x prime y. So, what this Gauss-Markov theorem says that, the estimator obtained by least square estimate. This is unbiased and it has the minimum variance compared to all other unbiased estimator which is linear in y.

So, the estimator obtained by using least square technique is the best among all linear unbiased estimators. So, these least square estimators are best linear unbiased estimator. These are called BLUE so, this is what the Gauss-Markov theorem is.

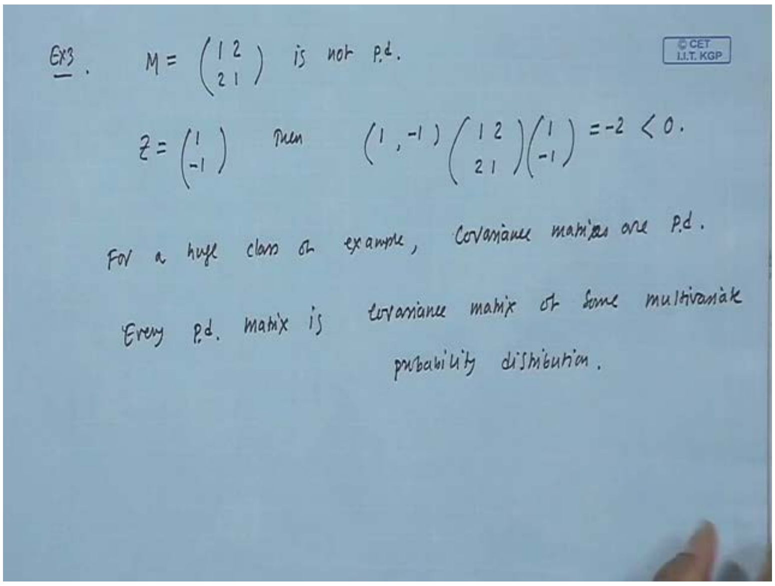
(Refer Slide Time: 17:25)



I hope that you know about positive definite matrix, but still I recall it. So, a matrix n cross n matrix m is positive definite. If z prime M Z is strictly greater than 0 for all nonzero vectors Z belongs to R to the power n ok. So, to make this definition clear I will give some example of positive definite matrix. So, example one, this is basically 1 0 0 1 consider this matrix. So, this matrix is positive definite because you considered z equal to say Z naught Z 1. Then z prime M Z equal to Z naught square plus Z 1 square and this is strictly greater than 0, because Z naught and Z 1 they are real and at least one of them is positive.

So, this is the positive definite matrix. So, similarly we can prove that I n, identity matrix of order n, this is also positive definite. Let me give one more example, example 2 consider this matrix which is equal to 2, minus 1, 0, minus 1, 2, minus 1, 0, minus 1, 2. So, this is a matrix involving some negative entry, but this is positive definite. Because you take a Z, Z equal to say Z 1, Z 2, Z 3. So, it is basically from part 3, then you can check that Z prime M Z is equal to Z 1 square plus Z 1 minus Z 2 square plus Z 2 minus Z 3 square plus Z 3 square. And this is strictly greater than 0, because of the fact that all this Z 1, Z 2, Z 3 they are real and at least one of them is positive.

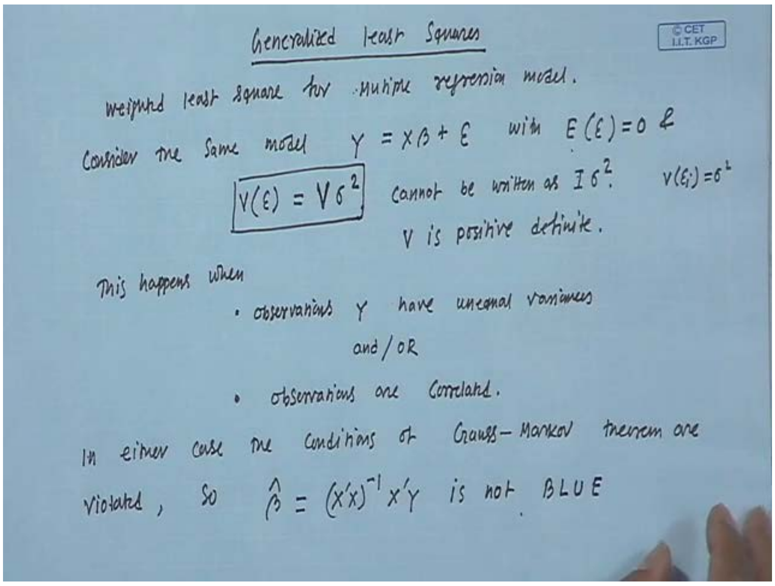
(Refer Slide Time: 21:50)




So, you may think that if there are some positive terms in the matrix, then it is positive definite. I just give one more example, say example 3, which is not positive definite. So, take this matrix M, all positive terms 1 2 2 1, but this is not positive definite. Because if you take Z non zero say 1, minus 1, then you can check that 1, minus 1, 1, 2, 2, 1, 1, minus 1. This is equal to minus 2, this is less than 0. So, this is an example of a matrix which is not a positive definite.

For a huge class of example, covariance matrices are positive definite. In fact every positive definite matrix is covariance matrix of some multivariate probability distribution.

(Refer Slide Time: 24:03)



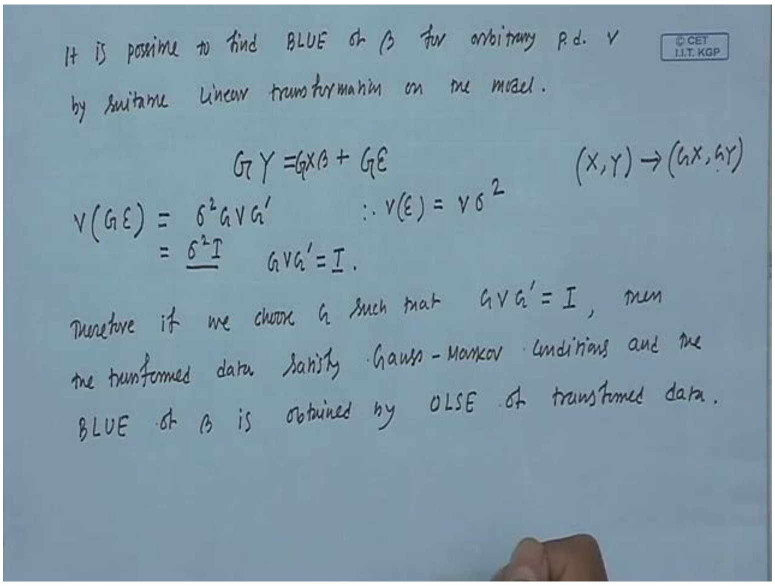

So, next I will move to the generalized a least square thing. So, here what we will do is that we will develop a weighted least square for multiple regression models. So, first recall the multiple linear regression models. So, you consider the same model say Y equal to X beta plus epsilon with expectation of epsilon is equal to 0 and variance of epsilon is equal to V into sigma square and this V into sigma square cannot be written as I sigma square. So, this is important because whatever we have done before for the multiple linear regression model, we have assume that expectation of E equal to 0 and variance of E is equal to sigma square I.

So, sigma square I means here the constant variants assumption is true that means the mean variance of epsilon I is equal to sigma square for all I and they are uncorrelated. That is why you can see here that off diagonal elements are 0 here, but here this is not true. So, this V is the covariance matrix and here V is positive definite.

So, now we are trying to fit a multiple linear regression model y equal to x beta plus epsilon, where expectation of epsilon is equal to 0, but the constant variant assumption is not here. The variants of epsilon is equal to V into sigma square, that means there exist inequality in variances and also the epsilon i's are not necessarily the y are uncorrelated, they are correlated here. That is why you can see that here V is positive definite matrix and it involve some nonzero (Refer Time: 28:25) diagonal terms. So, this happens when observations Y have unequal variances and or observations are correlated.

So, here is the violation of the basic assumption in the multiple linear regression models we assume before. Now, in either case the conditions of Gauss-Markov theorem are violated. So, beta hat which is equal to X prime X inverse X prime Y. This is the estimator of beta using the ordinary least square estimator is not the best linear unbiased estimator.

(Refer Slide Time: 30:49)



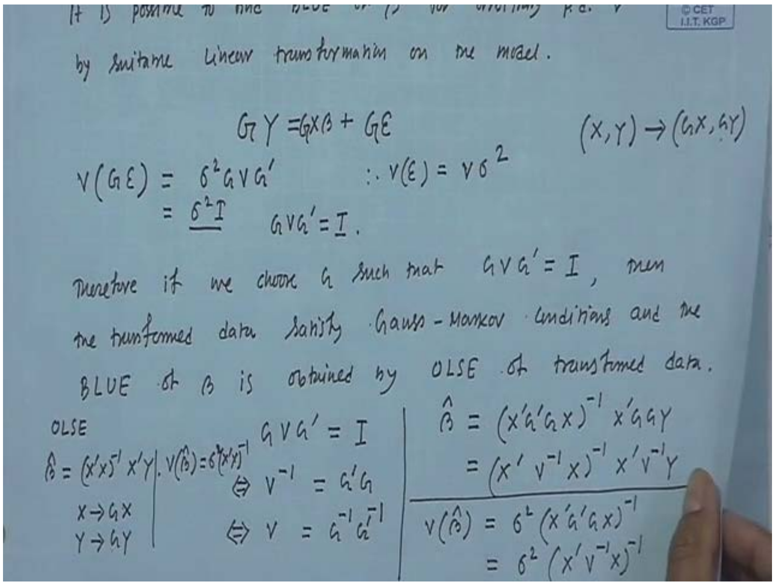

Now, in this situation also it is possible to find BLUE of best linear unbiased estimator of beta for arbitrary positive definite V by suitable linear transformations on the model. So, what we do is that, we have this model at this moment, multiple linear regression model Y equal to X beta plus epsilon. Now, we will take a linear transformation on this model. We multiply this model by G and now note that this is the transform model now.

So, the variance of G epsilon and I mean you have to choose the right G. So, how do you chose this G? Is the variance of G epsilon is equal to (Refer Time: 32:54) sigma square G V G prime. Because variance of epsilon is equal to V sigma square so, what we are doing is that you are given the data X, Y. So, X, Y the given data does not satisfy the model assumption, it has non constant variance and then we transform this data to G X, G Y. Now, our problem is to choose a correct G such that this quantity is identity.

So, you want to make you have to choose G in such a way that G V G prime is equal to identity. Therefore, if we choose G such that G V G prime is equal to I, then the transform data satisfy Gauss-Markov conditions and the BLUE best linear unbiased estimator of beta is obtained by ordinary least square estimator of transformed data. So, my original data does not satisfy the basic assumption. It has inequality invariance and they are correlated. So, we transform the data to G X and G Y and we have to choose this G in such a way that the transform error has variance sigma square I.

Since, the transform error has variance sigma square I so, it satisfied the condition of Gauss Markov theorem and that is why the BLUE of V can be obtained by ordinary least square estimator of transform data.

(Refer Slide Time: 36:51)

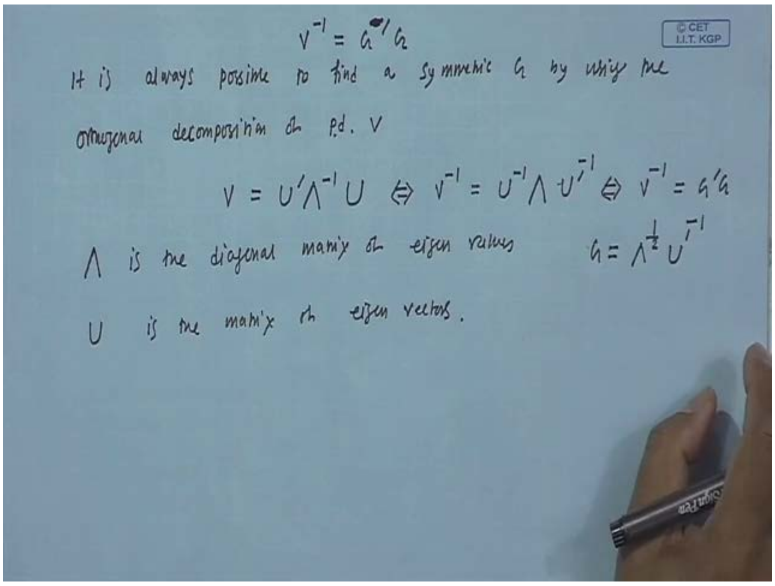



So, here this G V G prime equal to I, which is equivalent to V inverse is equal to G prime G and which is equivalent to V equal to G inverse G prime inverse. So, what is the least square? So, you have to choose the G in such a way that V equal to G inverse G prime inverse. Now, if we apply the ordinary least square estimator on the transform data, what will get is that, we will get beta hat equal to X prime G prime G X inverse X prime G G Y. How we got this thing?

Because in case of ordinary least square estimator, beta hat is equal to X prime X inverse X prime Y. So, what I am doing is that you just a replace X by G X and Y by G Y because we are working on the transform data G X and G Y. So, you replace this X by G X, you will get this formula. And finally, this can also written as in terms of V. I can write that this is equal to X prime V inverse X inverse X prime V inverse Y. So, this is the BLUE obtained using a generalized least square technique and the variance of beta hat is equal to sigma square X prime G prime G X inverse.

Before, in case of ordinary least square we had variance of beta hat is equal to sigma square X prime X inverse. So, what we are doing here is that just replacing X by G X and in terms of V, this can be written as sigma square X prime V inverse X inverse. So, this one is the generalized least square estimator of regression coefficient and here is the variance. Now, still I did not talk how to get this G? I mean we have V and then we have to choose G such that G V G prime is equal to identity matrix where we know that V is the positive definite matrix.

(Refer Slide Time: 41:38)

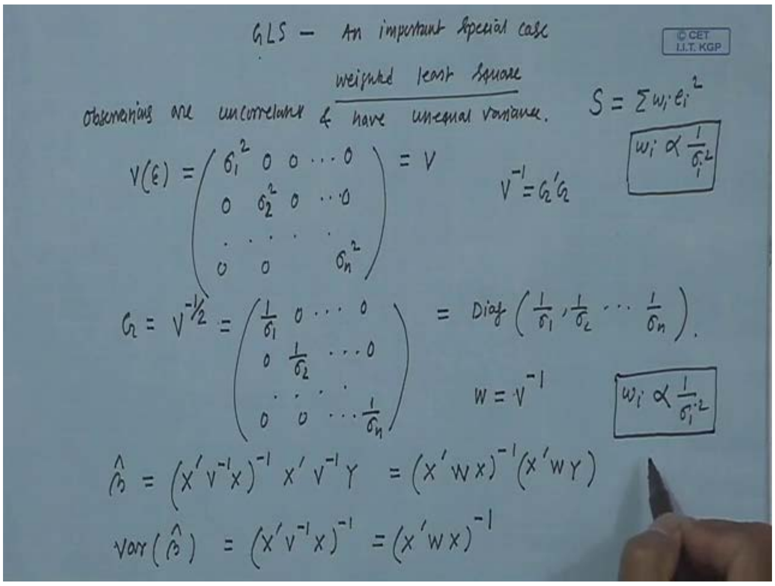



So, we have to choose G such that V inverse is equal to G prime G. Well, so V is a positive definite matrix. So, it is always possible to find a symmetric G by using the orthogonal decomposition of positive definite V. So, positive definite matrix V can be written as U prime, capital lambda inverse U where capital lambda is the diagonal matrix of Eigenvalues and U is the matrix of Eigenvectors. So, V is a positive definite matrix that we know from there and then V can be decomposed in this way. Now, we want to find G, right. So, this implies V inverse is equal to U inverse capital lambda U prime inverse and I want to write this V inverse as G prime G. Then the choice for G is, G equal to capital lambda half U prime inverse. So, this is the choice for G.

So, what we have learned here is that, if your given data does not satisfy the basic assumption of constant variance, if there are inequality in the variances and also if the errors are correlated or observations are correlated. Then variance of epsilon cannot be written as sigma square into i. It is V into sigma square, where V is a positive definite matrix. And then from that V, we can find the g such that, if you take the transformation. I just talked about here. If you take the transformation, then transformation like G Y equal to G X beta plus G epsilon.

Then you work on the transformed data. Now, the transformed data satisfy all the basic assumptions and you can apply ordinary least square technique on the transform data. So, this is what the generalized least square technique is and now, we will show how this generalized least square I mean, how we can get weighted least square technique as particular case of generalized least square.

(Refer Slide Time: 46:16)




So, we already know what is generalized least square and an important special case is weighted least square. So, at the beginning I talked about weighted least square and then I said that we have to find the regression coefficient by minimizing this w i e i square where, w i is proportional to 1 by sigma square. And then I said that I will explain this, why this weight is proportional to 1 by sigma square? Now, we know generalized least square and as a particular case, weighted least square is a particular case of generalized least square.

So, here observations are uncorrelated and have unequal variance. So, uncorrelated means the off diagonal elements are 0. So, variance of epsilon, you can take this as sigma square 0 0 0. These are the covariance terms. All the covariances are 0, because it is uncorrelated 0 sigma, 2 square 0 0 sigma n square. So, this is the variance covariance matrix. Well so, this is equal to V and I want to find a G such that V inverse is equal to G prime G. And here it is easy. You can very easily find G, such that V inverse is G prime G. You can check that the choice of G is equal to just V to the power of minus half.

This is my V, this is equal to 1 by sigma 1 0 0 0 0 1 by sigma 2 right. So, if you choose this G, you can check that V inverse is G prime G, right. This one is nothing but diagonal 1 by sigma 1, 1 by sigma 2, 1 by sigma n. Well so now, my beta hat is equal to X prime V inverse X inverse X prime V inverse Y and if we write my weight is equal to V inverse, then this is nothing but X prime W X inverse X prime W Y and variance of beta hat is equal to X prime V inverse X inverse which is equal to X prime W X inverse.

So, this is the weight we are talking about in the weighted. In the weighted least square the weight is nothing but V inverse whereas, this V is matrix and it is clear that why w i is proportional to 1 by sigma i square. So, I explained this part but still I need to explain one more thing. So, I explain why this weight w i is proportional to 1 by sigma i square. But, I will explain in my next class: how to get this sigma i square? Because see you are just given a set of observation or a data set X i Y i and from that data you identified that some of the basic assumptions like constant variance and normality assumptions are violated. And to correct those assumptions, what you do is that you are going for weighted least square.

Now, to apply weighted least square, you need to know sigma I squares right? But sigma i squares are of course, not given. So, in the next class I will take an example, which gives you just a set of observation X i and Y i. And then we will first realize for that observation some basic operations are violated and then we will try to apply a weighted least square estimator to correct that model. And of course, for that we need to find the sigma square and we will talk about how to find sigma I square in the next lecture.

Thank you very much.