Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 103
Lecture 44C - Estimation Methods 2 -1

I just want to point out that there are other variants of OLS.

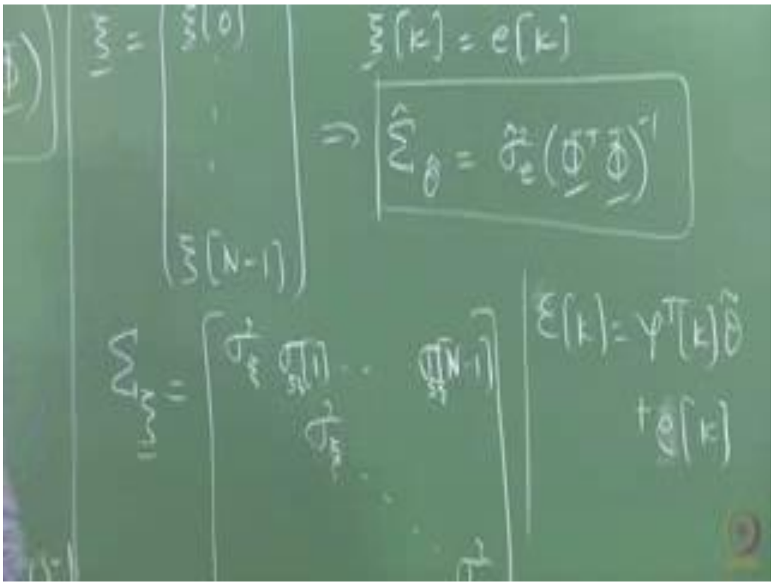
(Refer Slide Time: 00:15)

That is least squares: the original version is what we have discussed. Then came about a lot of other variants like in any other typical research, you have weighted least squares partial least squares, total least squares, there is no eclipse least squares or anything like that, but you have non-linear least squares, generalize least squares, this least squares, that least squares and so on; of which the two important ones of interest to us or the weighted least squares and the non-linear least squares.

The weighted least squares differs from the ordinary least squares in only one aspect and that aspect is instead of solving a regular least squares problem; remember I said that ordinary least squares can be viewed as minimizing the sum square errors by giving equal importance to all observations, whereas in the weighted least squares I would like to actually give different weights to different observations. When do we run into such situations? There are numerous situations, but at least if you think of two different situations: one is when you have heteroskedastic errors.

What is heteroskedastic error mean when your z heteroskedastic? The variance is changing from observation to observation; that can happen. Remember and the other thing that you should remember is although we keep using this k here as an index in time, so when I write here k for example, I write here e z k.

(Refer Slide Time: 01:47)



You should have a broader understanding of this k, this k need not be index in time it is just an index of observations. For example, I may have 10 different sensors measuring the same thing in many critical units in many processes. For example if you take nuclear reactor temperature of the reactor is a very critical variable. The industry does not rely just on one sensor; so if you go to IGCAR or any other you know plant which is running nuclear reactor you would have ten sensors reading this because I do not trust just one sensor if it fails then that is it you know there I would not be alive to see again why it has failed.

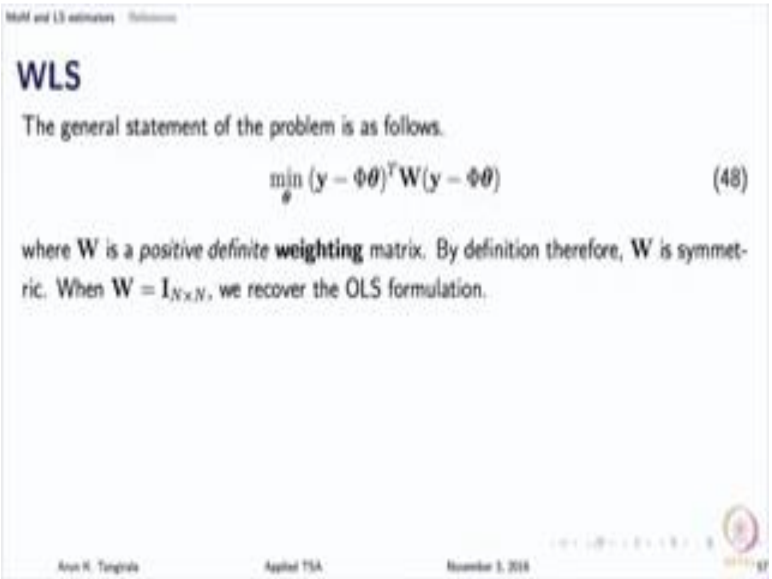
So, it is better to actually have many sensors, and then there are mechanisms to fuse this data that is coming together, and each sensor can have a different variance. So, in practice what is done I will take the average of all the sensors; least squares estimate of that average would be simple average simple mean we know that sample mean is the solution, but that is only true if each sensor as the same level of reliability.

In general that need not be true, I may have to actually give different weightings to different sensors. So, that is one situation where, and that is the case of heteroskedastic errors. So, the k need not be thinking be correlation just in time I mean sorry; k need not keep track of time, k will allow you to also keep track of sensors. But, in another situation for example; in econometrics it is very popular to have heteroskedastic errors on in some other applications. In those situations also it is important to give different weightings to different observations.

The other situation where weighted least squares would be required is when z is colored. So that there are numerous reasons why we want to look at the weighted least squares, fortunately the solution to the weighted least squares can be obtained in one shot using the solution to the OLS.

So, I will just show you the formulation here and the solution; it just takes a minute that is it and we will adjourn.

(Refer Slide Time: 04:05)

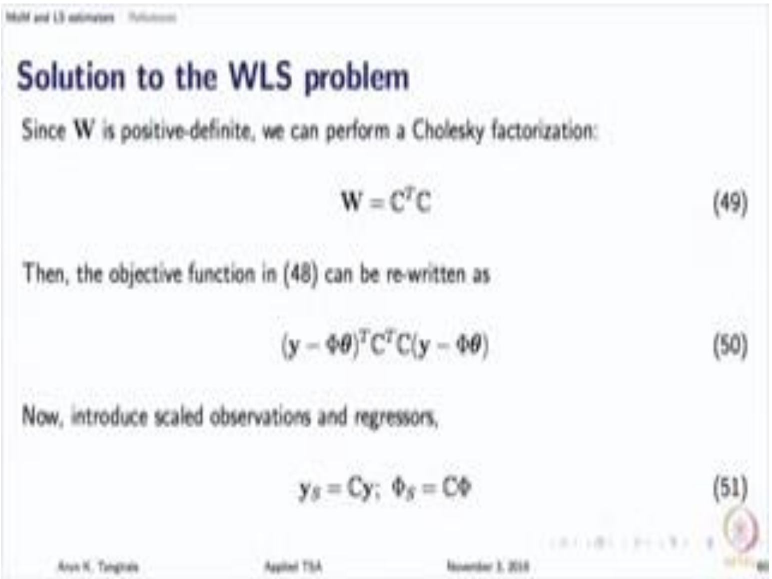


So, this is the formulation now of the weighted least squares instead of minimizing y minus phi theta transpose times y minus phi theta now we have a weighting matrix coming and sitting there. You can clearly see if the weighting matrix is identity then you go back to the OLS formulation. If the weighting matrix is diagonal with different weights then you are giving importance, attaching different importance to levels of importance to different observations. The general case is when the weighting matrix is this w is full matrix, but it cannot be any arbitrary full matrix it has to be a positive definite matrix because we want this objective function to remain convex.

Generally, when we choose weighting matrices we have to choose positive definite matrices and that is ensured by choosing symmetric at just choose w to v symmetric. And if you are in doubt look at the eigenvalues of the w they should all be greater than o. But what is the full w mean? Why should I have off diagonal elements in w, any idea? When I have errors that are correlated there are.

We talked about two different situations: one is heteroskedastic errors, but the errors are not correlated in time individually they have different variances. Then the other situation is errors are correlated. Then I may choose a w that is full matrix. Anyway, so choice of w we will talk about later on I just want to go over this solution.

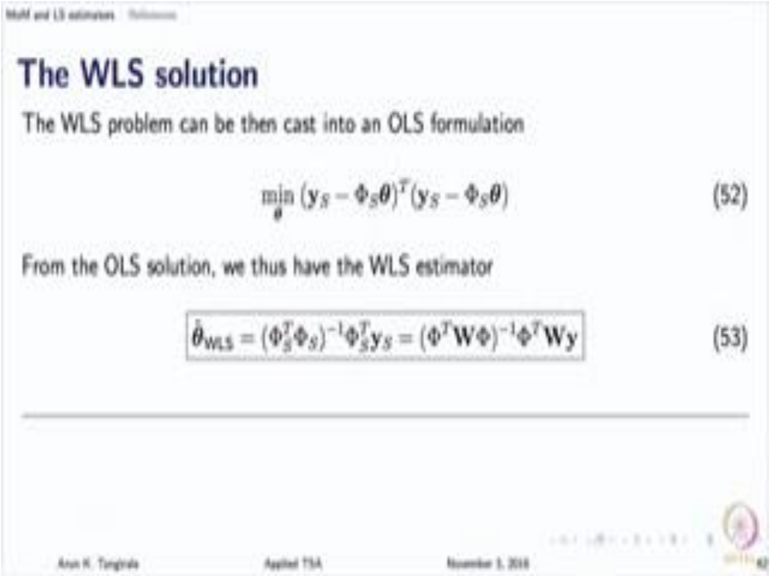
(Refer Slide Time: 05:48)



And all you do is since w is positive definite you do a Cholesky factorization. If you forget what a Cholesky factorization just go and look up a book on matrix algebra. So, you can factor a positive definite matrix into c transpose times c; where c is a Cholesky factor so that I rewrite the objective function in this form.

All I have to do now is to solve this problem introduce the scaled data.

(Refer Slide Time: 06:22)



So, introduce scaled y which is c times y and scale phi which is c times phi so that I can rewrite the WLS in terms of OLS; as an OLS in terms of this scaled variables, that is it. So, once I have done this I straight away get the solution to the weighted least squares problem.

So, that is the trick that is used widely in any estimation optimization problem. If I know the solution to one I cast a new problem into an old one by moving into a new space. But that gives you a lot of insight; it says essentially that when you have data with varying errors construct a new data by scaling appropriately so that now your new data as errors of equal variance for example.

Now, in your new domain, in your scale domain the scaled data have errors which are heteroskedastic they have same errors for example. The only million dollar question is who gives me w, because all we are saying is I assume I know w and w has to be driven as we will see tomorrow is driven by the noise covariance matrix sigma z. So, in practice we estimate this iteratively. We fit a model get an estimate of sigma c then apply the weighted lest squares and then solve iteratively.

There are some special situations in which w are known a priory, where w is not driven by the error covariance, but something else. But in general w is estimated iteratively along with theta. So, that is weighted least squares we will just discuss for few minutes tomorrow early on and then conclude with non-linear least squares.

Thank you.