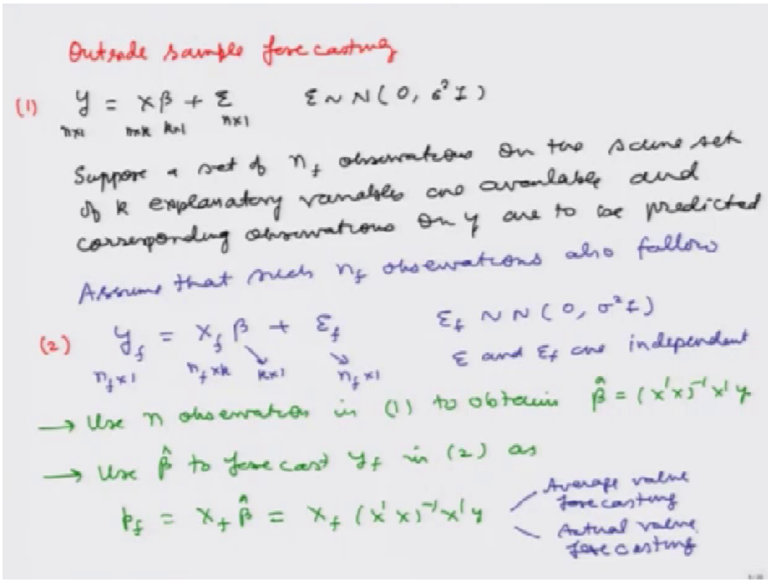
Regression Analysis and Forecasting
Prof. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology-Kanpur

Lecture 23
Outside Sample Forecasting

Welcome to the lecture you may recall that in the earlier lecture we had discussed the forecasting in multiple linear regression models, and we had considered within sample forecasting and in that case we have considered two cases, the average value forecasting and actual value forecasting. Now continuing on the same lines we are going to discuss now here the second case where we will consider the outside sample forecasting.

(Refer Slide Time: 00:46)



Right, so you may recall that we had discussed about what is the difference between within sample forecasting and outside sample forecasting in the case of within sample forecasting we tried to obtain the value of y for a given value of x, which is lying within the range of x values, and in the case of outside sample forecasting we try to obtain a value of y for a given value of x and the value of x is lying outside the range of given sample values on x.

Okay, so in this case, first we formulate our setup, so we will consider the multiple linear regression model where we assume, that we have got n observations on the response variable y in form of a n cross one vector y and we have got k explanatory variables, and we have got n observation on each of the k explanatory variable x1, x2, xk, and they are clubbed in the matrix of order n by k denoted as x.

And beta is a K cross 1 vector of regression coefficient and epsilon is n cross 1 vector of random error components and epsilon is assumed to follow a multivariate normal distribution with mean vector 0 and covariance matrix sigma square i. Now this is our basic model, and now in order to do the forecasting, we assume that suppose a set of nf observation on the same set of k explanatory variables are available and corresponding observations on y are to be predicted.

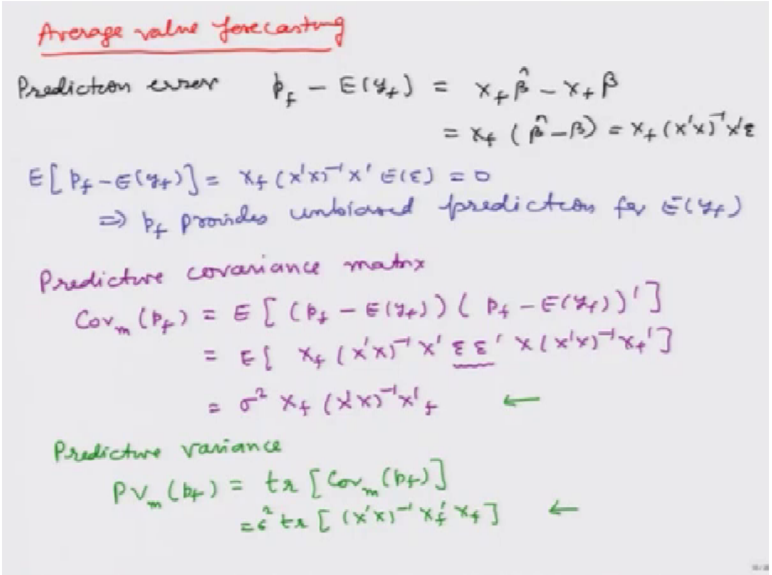
So we have here two sets of observation, the first set of observation consisting of n observations on x and y and there is another set of nf observation in which only the value of x are known but the values of y's are unknown. So we assume that for this additional nf observation, the same model continues and we assume that such nf observations also follow the same model that is y=x beta plus epsilon, but now there is a different between this model and earlier model.

So we try to denote it by here y as yf, x here as xf and epsilon as epsilon f where now yf is a nf cross1 vector of values on Y that are unknown which has to be predicated, which has to be forecasted and xf is a matrix of order nf cross k and beta is k cross 1 vector and epsilon f is a nf cross 1 vector of random error components and we can assume that epsilon f also has got the similar properties like as in the earlier case it is following a multivariate normal distribution with mean vector 0 and covariance matrix sigma square i.

And further we also assume that the elements in epsilon and epsilon f are mutually independent. Okay, one thing we have to keep in mind that the variables in matrices x and xf they are the same, same x1 x2, xk okay. So let us call this model has model number 1 and this model as model number 2. Now we have to do the forecasting the forecasting in the case of outside sample case is done in two steps, the first step is to use n observations in the model number 1 to obtain the estimate of beta as beta hat = x transpose x whole x transpose y.

And in the second step use beta hat to forecast yf in model number 2 as we try to construct the predictor pf which is xf beta hat and this is xf x transpose x whole inverse x transpose y. So now we are going to use this same predictor for average value forecasting and actual value forecasting.

(Refer Slide Time: 06:44)



So now we consider the first situation in which we are considering the average value forecasting. Now we try to adjust the statistical properties or predicator pf when it is used for average value forecasting and in order to do so first of all we find its prediction error, so as we have done in the earlier case we are trying to use the predictor pf to forecast expected value of yf.

So the prediction error is given by pf - expected value of yf this = xf beta hat-hat- xf beta and this can be written as a xf beta hat - beta, and that = xf x transpose x whole inverse x transpose epsilon, right. You may recall that beta hat - beta = x transpose x whole inverse x transpose epsilon okay. Now we try to find the expected value of prediction error and we notice that this is expect xf x transpose x whole inverse x transpose expected value epsilon this =0.

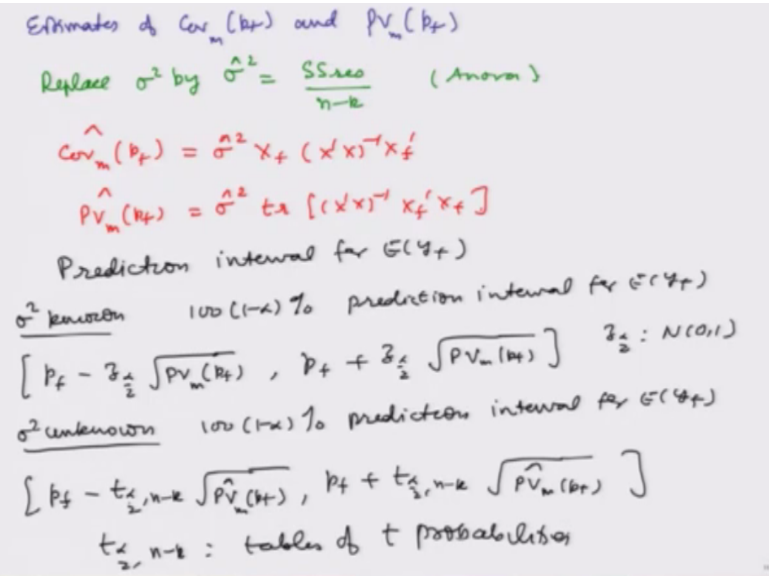
So this implies that pf provides unbiased prediction for expected value of yf that is in forecasting the average value in case of outside sample prediction. Now we have to find the predictive variance. In the earlier lecture we had found the predicator variance directly, but in this lecture just for the sake of illustration we will find the predictive covariance matrix first and then we will try to take it trace to find out the predictive variance.

So let us try to find out the predictive covariance matrix say this is defined has covariance of pf when it is used for mean value denoted by m and this is expected value of pf- expected value of yf and pf- expected value of yf and its transpose. We already have obtain these expressions, so we can write down here directly this is xf, x transpose x whole inverse x transpose epsilon, epsilon transpose x x transpose x whole inverse xf transpose.

Now when we bring the expectation operator inside the bracket, the expected value of epsilon, epsilon transpose is sigma square i and so this can be further simplified as sigma square xf x transpose x whole inverse x transpose f, and now from here we can find the predictive variance as PV of pf when it is use for mean value forecasting is equal to trace of covariance matrix of pf when it is use for mean value predication and this is nothing but trace of sigma square times trace of x transpose x whole inverse xf transpose xf.

So now we observe that both this exceptions of covariance matrix of pf as well as the predictive variance of pf they are depending on the unknown value of sigma square.

(Refer Slide Time: 11:07)



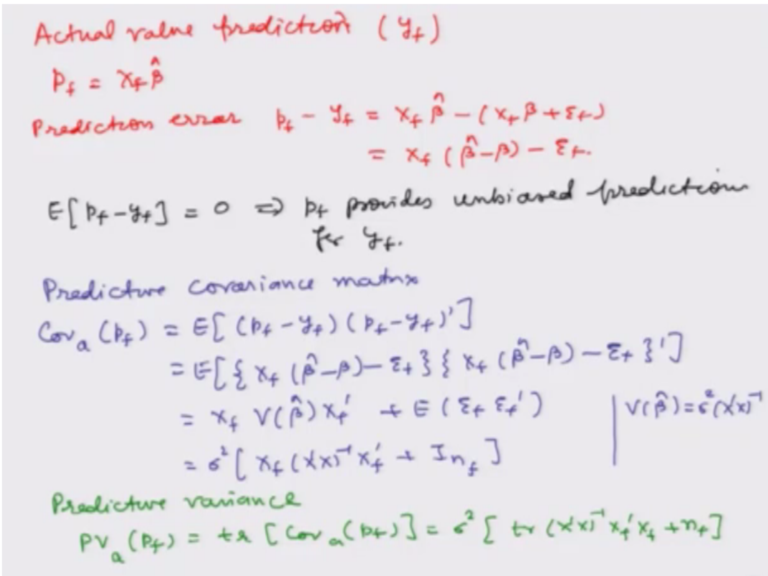
So we would like to estimate them on the basis of given sample of data and we would like to obtain the estimates of covariance matrix pf and predictive variance of pf, okay so in order to obtain the estimate the rule is very simple, simply replace sigma square by sigma square hat and that is obtain from the analysis of variance by SS residual divided by n minus k. Okay now once I try to do it, then an estimate of the covariance matrix of pf and it is used for mean value predication is obtain as sigma square hat xf x transpose x whole inverse xf transpose.

An estimate of predictive variance of pf when it is used for mean value predication this is obtained has sigma square hat and trace of x transpose x whole inverse xf transpose xf. Now we obtain the predication interval for expected value of yf. The prediction interval can be obtained exactly in the same way as we have done in the earlier lecture. So we will try to consider here two possible cases when sigma square is known.

And In this case the 101- alpha percent prediction interval for expected value of yf is obtain as say here pf - z alpha by two and square root of predicative variance of pf when it is use for mean value prediction and pf+z alpha by to square root of predicative variance of pf when it use for mean value prediction and here this value Z alpha by 2 they are obtain from normal 0 one tables.

And similarly in other case we assume that sigma square is unknown, and in this case hundred 1 - alpha % predication interval for expected value of yf is obtained as pf - t alpha by two n- k and standard error of pf when it is used for mean value and pf + t alpha by 2 n-k and standard error of pf when it is used for mean value, okay and here this value t alpha by 2 n-k they are obtain from tables of t probabilities.

(Refer Slide Time: 14:37)



So this is all about the average value predication in case of outside sample predication and now we try to consider the actual value predication in the case of outside sample predication. So here we are essentially interested in forecasting the value yf and for that we are using the predictor pf that is equal to xf beta hat. So now first of all we try to find out the predication error.

So we are trying to find or predict yf using pf, so the predication error is given by pf-yf and this is nothing but xf beta hat-xf beta+epsilon f, so this is xf beta hat – beta-epsilon f, okay.

So now we try to take its expectation and we find that expected value of pf-yf this comes out to be null vector because expected value of beta hat = beta and expected value of epsilon f= null vector.

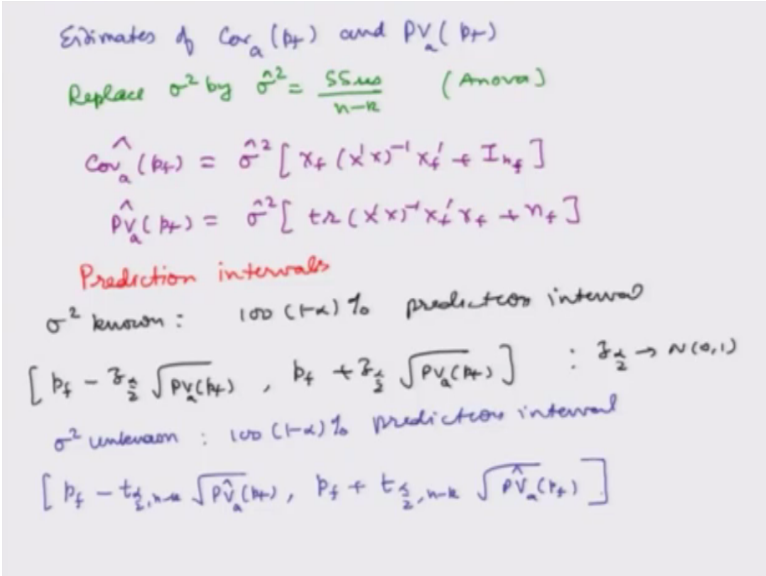
So this implies that pf provides unbiased predication for yf. Next we try to find out its predictive covariance matrix, and this is obtained as we denote it say covariance matrix of pf when it is use for actual value a, okay. So this is defined as expected value of pf-yf, pf-yf transpose and when we try to solve it this is xf beta hat-beta-epsilon f xf beta hat-beta-epsilon f and its transpose.

And when we try to open it, and we try to take the expectation of an inside this comes out to be nothing the covariance matrix of beta hat times xf transpose + expected value of epsilon f, epsilon f transpose which is nothing but the covariance matrix of epsilon f. Since we already have derived that the covariance matrix of beta hat is given by sigma square x transpose x whole inverse.

So I can substitute this value here and I can obtain it say xf x transpose x whole inverse xf transpose plus sigma square i which is identity matrix of order say nf cross nf. Now using this result we can find out the predictive variance. So we will denote the predictive variance of pf when it is use for actual value forecasting this is simply the trace of co variance matrix of pf when it is used for actual value forecasting and this is simply the sigma square times trace of x transpose x whole inverse xf transpose xf plus nf,

Nf because trace of identity matrix of order nf. So now we obtain here the predictive covariance matrix and predictive variance of pf, but we can again see that these two expressions are depending upon sigma square that is unknown.

(Refer Slide Time: 18:50)



So now we try to find out estimates covariance matrices and predicative variance, estimates of covariance matrix of pf and predictive variance of pf and they are used for actual predication. So in order to obtain them we simply have to replace sigma square by sigma square hat which is obtain by sum of square due to residual divided by degrees freedom and this some of a square due to residual that is obtain from the analysis of variance table.

So, once I try to do it here then the estimate of covariance matrix of pf when it use for actual value predication this comes out to be sigma square hat, and xf x transpose x whole inverse xf transpose plus identity matrix of order nf and estimate of predictive variance of pf when it is used for actual value prediction comes out to be sigma square time trace of x transpose x whole inverse xf transpose xf+nf,

Now our next objective is to find out the predication interval. So the prediction intervals can be found exactly in the same way as we have done in the earlier cases, so we will have here two cases one is sigma square is known and in this case the hundred one minus alpha % predication interval is obtain as pf-z alpha by 2 square root of predicative variance of pf when it use for actual value and pf+z alpha by 2 square root of prediction variance of pf when it is used for actual value.

And where the values of z alpha by 2 are obtained from the tables of normal 0 1 probabilities, similarly the second case here will be when sigma square is assume to be unknown and in this case the 100 1- alpha % prediction interval is obtained as pf-t alpha by 2 n-k and standard error of pf when it is used for actual value prediction and pf+ t alpha by 2 n-k and standard error of predictive variance that is pf when it is use of actual value.

So now you can see that we have considered the outside sample predication in this lecture we constructed a predicated based on the ordinary least square estimator and we have implied it for the average value forecasting and actual value forecasting, next question which have I would like to address is a following, in this lecture and in the earlier lecture we had use only the ordinary least square estimation for forecasting.

This possibly will give you and impression that possibly only the ordinary least square estimator can be used for forecasting this is not correct, one can actually use any estimator for forecasting the only difference is going to be that we will replace the ordinary least square estimator by the corresponding estimator, but we have to clear full after that the properties of the predictors there test of hypothesis, there prediction interval will also be changed accordingly.

Sometimes it may easy to find such properties and sometime it may be difficult to find such properties. So now we are going to stop here the only question which is left is that we have developed the theoretical properties in this lecture and the earlier lecture on forecasting, but how to use them on a real set of data, so that we would try to address in the next lecture, till then good bye.