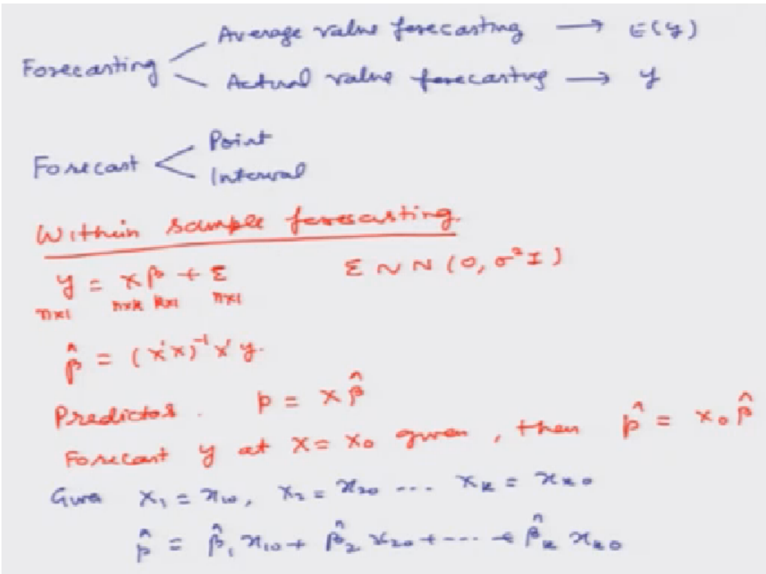
Regression Analysis and Forecasting
Prof. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology-Kanpur

Lecture: 22
Within Sample Forecasting

Welcome to the lecture you may kindly recall that in the earlier lecture we had discussed the basic concepts of the forecasting and we had understand the concepts of within sample forecasting, outside sample forecasting, actual value prediction and average value prediction. Now using the framework of multiple linear regression model, we are going to construct the predictors and we are going to understand how to find out the statistical properties of the predictors based on that we can judge weather our forecasting is going to be good or bad.

(Refer Slide Time: 00:55)



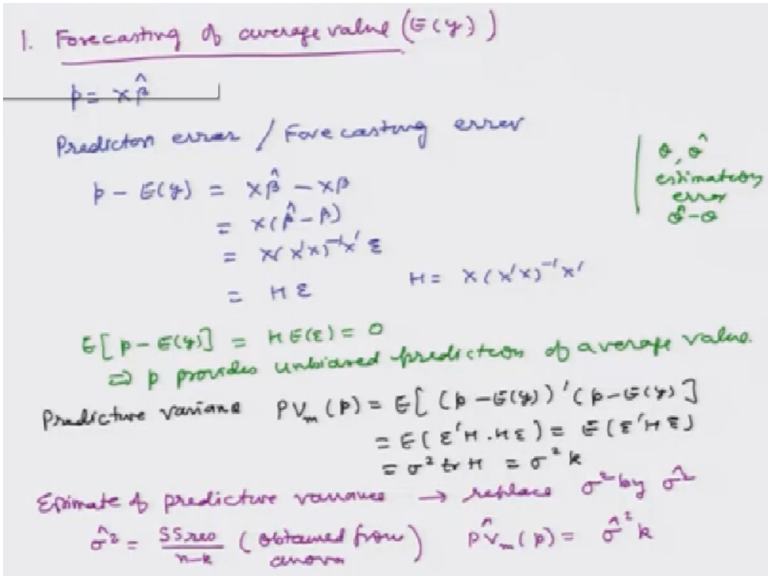
So firstly let us try to consider the within sample prediction or within sample forecasting, right. So let us try to first establish our framework, so we are going to use the model y=x beta + epsilon where is a n cross one vector of n observation on that is response variable or steady variable, x is n cross k matrix of k explanatory variable where each explanatory variable is an having an observation.

Beta is a k cross one vector of regression coefficient associated with x1, x2, xk and epsilon is a n cross one vector of the random error component and we assume that epsilon is following normal distribution with mean vector 0 and covariance matrix sigma square i and we have already seen that this beta is unknown and we estimate beta by beta hat as x transpose x whole inverse x transpose y. Now based on that we try to construct a predictor and this predictor is constructor like p=x beta hat.

Now suppose we want to forecast the value of y at x=x naught which is a given point, then forecast is going to be p hat=x naught beta hat, right. So for example if I say that we are given, suppose the first value of x1 as x one naught, second value of x2 as x2. naught and so on the kth variable is having the value xk naught then your predictor is going to be beta1 hat x1 naught plus beta2 hat x2 naught up to here beta hat k, xk naught.

So now this is our predictor which is based on the ordinary least square estimation or equivalently the maximum likelihood estimation and now we will try to use this predictor to forecast the average value and the actual value.

(Refer Slide Time: 03:34)



So first of all we consider the forecasting of average value expected value of y okay, so in this case we are going to use the predictor as p is equal to x beta hat, right. So now I would like to know the statistical properties and performance of this predictor when it is used for forecasting and average value, okay. So in the first step we try to find out the prediction error or this is called forecasting error.

You see we are using here the terminology prediction and forecasting exactly on the same line, where there are some minor difference between the two, but we are not going into that detail, oaky. So this prediction error is given by p- expected value of y, and if you try to see this is based on the same concept as in the case of an estimation error for example when we are trying to estimate a parameter theta by theta hat then we try to find out the estimation error as theta hat- theta.

Okay, so similarly here also we are trying to known the value of expected value of y using p. So if you try to simplify this is nothing but x beta hat- x beta and which is x beta hat - beta and this is nothing but your x x transpose x whole inverse x transpose epsilon and this is nothing, this is our symbolic notation this is h epsilon where h is our hat matrix given by x x transpose x inverse x transpose.

Now based on that we try to first see weather this is an unbiased predictor or not. So we try to take its expectation and expectation of p- expected value of y and this comes out to be h expected value epsilon, which we have assumed to be 0, so this is a null vector, so this implies that p provides unbiased prediction of average value, okay. Now next we try to find out the predictive variance, and this is defined as say PV and I would write here m, m means mean value okay.

So the predictive variance for p forecasting the mean value, this is given by expected value of p- expected value of y transpose p- expected value of y, okay, so this value we have already have found this is expected value of epsilon transpose h into h epsilon which is expected value of h is an idempotent matrix, so this is epsilon transpose h epsilon and this quantity is nothing but Sigma square traces of h and trace of h is nothing of k, sigma square time is k, right.

So this is our predictive variance when p is used for predicting the average value now we observe that this predictive depends on sigma square that is usually unknown in practice, so an estimate of predictive variance can be obtained by replace sigma square by sigma square hat, and this sigma square hat is obtained from the analyses of variance sum of square due to residual divided by degrees of freedom, and this obtained from analyses of variance.

So in this case we can write down that the predictive variance as an estimate as sigma hat square times k, okay. So now we have described here a predictor, we have obtained its prediction error, and we have shown that it will provided us an unbiased prediction, and its predictive variance is given by sigma square k and this can be estimated on the basis of given sample of data using sigma hat square k.

(Refer Slide Time: 08.52)

Prediction interval

σ^2 known $\rightarrow$ $\dfrac{p - E(y)}{\sqrt{PV_m(p)}} \sim N(0,1)$

$P\left[-z_{\frac{\alpha}{2}} \le \dfrac{p - E(y)}{\sqrt{PV_m(p)}} \le z_{\frac{\alpha}{2}}\right] = 1-\alpha$ $\quad z_{\frac{\alpha}{2}}$: Normal table

$100(1-\alpha)\%$ C.I. for $E(y)$

$\left[p - z_{\frac{\alpha}{2}}\sqrt{PV_m(p)}\,,\; p + z_{\frac{\alpha}{2}}\sqrt{PV_m(p)}\right]$

σ^2 is unknown $\rightarrow$ $\dfrac{p - E(y)}{\sqrt{\widehat{PV}_m(p)}} \sim t(n-k)$

$P\left[-t_{\frac{\alpha}{2}, n-k} \le \dfrac{p - E(y)}{\sqrt{\widehat{PV}_m(p)}} \le t_{\frac{\alpha}{2}, n-k}\right] = 1-\alpha$

$100(1-\alpha)\%$ CI for $E(y)$

$\left[p - t_{\frac{\alpha}{2}, n-k}\sqrt{\widehat{PV}_m(p)}\,,\; p + t_{\frac{\alpha}{2}, n-k}\sqrt{\widehat{PV}_m(p)}\right]$

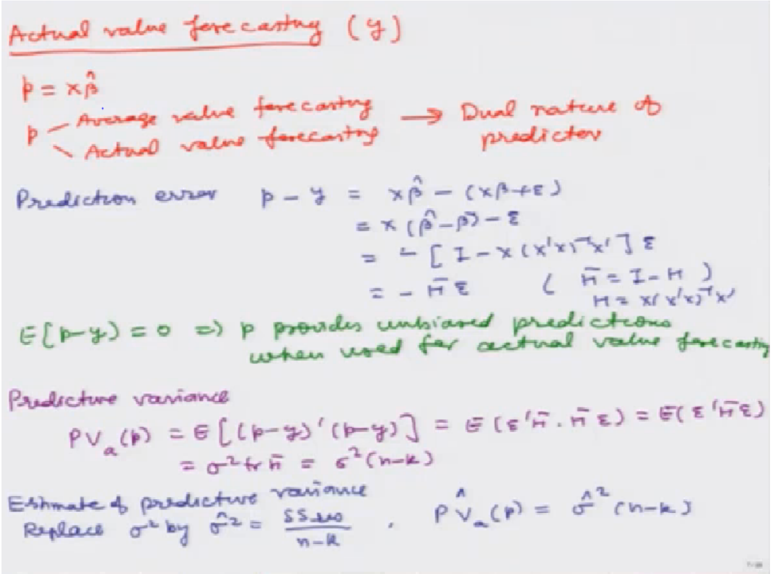
Right, now in case if you want to have the prediction interval then the prediction interval can be obtained as a like as follows, we will have a two cases when sigma square is known, then in with this case we known that p- expected value of y divided by standard errors of p, this follows a normal 0 1 okay. So now based on that we can write down here probability that - z alpha by 2 p- expected value of y and standard error lies between z alpha by 2=1-alpha.

Where z alpha by 2 are the values, which are obtain from the normal table as we done in the case of regression parameters also, and so the 101 - alpha % confidence interval for expected value of y can be obtained by solving this inequality and this gives as p- z alpha by two square root of PVmp and P+ z alpha by 2 square root of PV petty variance of p for mean values, okay.

Now there is second case, second case is that when sigma square is unknown, in this case we known that predictor minor expected value of y divided by standard error of p, this follows a t distribution with n- k degrees of freedom, and in this case the critical value are - t alpha by 2, n-k less than or =p minus expected value of y over square root of estimated predictive variance for mean t alpha by 2, n-k degrees of freedom =-alpha.

And based on that 101- alpha % confidence interval for expected value y is obtain as p- t alpha by 2 n-k, standard error of p and p+ t alpha by 2 n-k and standard error of PV. So this is confidence interval for expected value of Y, when sigma square is know and when sigma square is unknown.

(Refer Slide Time: 11:54)



The next aspect we try to considering it for actual value predictions or actual value forecasting, so we are interested in forecasting the value of y okay. So now we are going to use here the same predictor, so same predictor has been used for the average value forecasting and as well as this has been use for the actual value forecasting, so this is actually called as dual nature of predictor.

Well, that is happening because we are using the ordinary least square estimator in both cases and in practice one may use different estimators to forecast the average value or actual value. One question arises here that when we are tiring to use the same predictor for forecasting the actual n average values definitely it will give us the same value then what will be the difference, so the difference will be seen by comparing the standard errors.

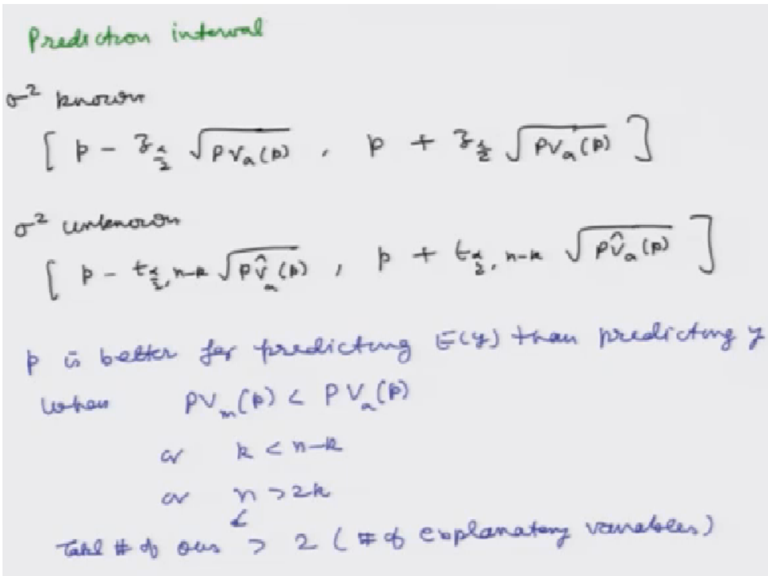
Okay, so let as try to investigate the properties of this predictor, when it is used for actual value forecasting, so first of all we try to find out its prediction error. So since P has been used to estimate or to known the value of y, so the estimation error is given by p- y, and p is our x beta hat and y is word beta + epsilon, so this can been written as beta hat - beta, - epsilon and this is - times i-x x transpose x whole inverse x transpose epsilon and this is - h bar epsilon where h bar is nothing but i-h.

And h is x x transpose x whole inverse x transpose hat matrix, right. So now from here we observe that expected value of p minus y this comes out to be null vector, so this implies that p provides unbiased predictions when used for actual value forecasting, okay. Next we try to find out the predictive variance, predictive variance in this case let us try to write down here as say a, a means actual.

So this will be expected value of p-y transpose p-y and this is nothing but your expected value of epsilon transpose h bar into h bar epsilon and this nothing but expected value epsilon h bar epsilon because h bar is an idempotent matrix and this quantity is nothing but sigma square trace of h bar which is sigma is square n minus t. Okay, now again the question arises that this sigma square is unknown.

So we can obtain here an estimate of predictive variance, so this can be obtain by just replace sigma square by sigma square hat and that is obtain from the analyses of variance as sum of square due to residual divided by the quantity and -k, right. So in this case an estimate of predictive variance of p when it is used for an actual value turns out to be sigma hat square times n minus k right.

(Refer Slide Time: 16:30)



Next we try to obtain the prediction interval in this case can also be obtain exactly as in the case of average value forecasting and we will have here two cases when sigma is square in known then this case the prediction or forecasting interval turns out to be p-z alpha by 2p predictive variance of p for actual value and p-z alpha by two standard deviation of p when it is used for actual value and the second case will be when sigma square is unknown.

And this case this prediction interval is t distribution with n-k degrees of freedom and this will be standard error of p when it is used for actual value and p+t alpha by 2 n-k standard error of PV of p when it is use for actual value prediction. Now another interesting question comes here, that we are using this p for making a forecast for average value as well as for the actual value.

So in case if you try to compare it that p is, then we can say that p is better for predicting expected value of y then predicting the actual value y when the predictive variance of p when it is used for forecasting mean value is smaller than the predictive variance of p when it is used for actual value and this gives us or k less than n-k or I can say that n is better than two k, so that means the total number of observation should be greater than twice the number of explanatory variables.

So that is very simple condition under which we can known that the same predictor is going to perform how under which type of condition, right, okay, so we stop here and in the next lecture we will try to concentrate on the outside sample prediction under a simpler setup and we will try o take one example where we would like to see how does this values comes out and how do we take an inference on that, till then good bye.