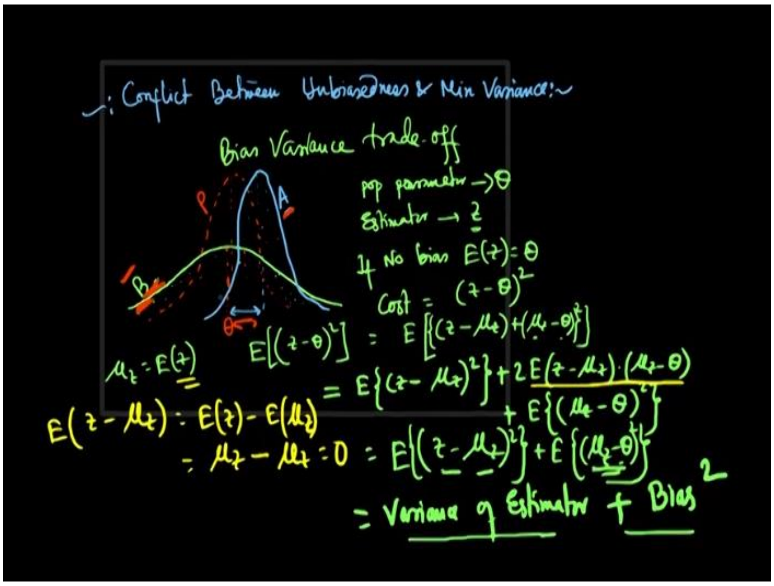
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Module - 3
Lecture - 28
Conflict Between Unbiasedness and Min Variance

Hello and welcome back to the lecture on Applied Econometrics. So, in the previous lecture, we spoke about the properties of estimator. And we have seen that, for an estimator to satisfy the BLUE condition, it has to be unbiased, it has to be of minimum variance. Now, can we always manage to achieve this unbiasedness and at the same time minimum variance? or is there any cost involved? is there any trade-off involved? So, that is what we are going to see in this lecture.

(Refer Slide Time: 00:55)



And in reality, we will actually see that there is a trade-off between bias and variance. So, I will just write down; bias variance trade-off; that is very important. And we will see the implication of this bias variance trade-off when you are trying to fit a model. Now, let us first try to just visualise the problem. So, first, let us say there is a true population distribution. As usual, we do not know anything about this true population distribution, we do not know the population mean, but we want to understand the mean of this population distribution from the sample distribution.

Let us say I have a sample distribution. Let us say, this is the sample distribution I got. Let me actually do a thick line for sample distribution, because I have already realised it, I have already got it. And then, there is another sample distribution; let us say it is of the blue line and it is here. Now, let us say I have the true population parameter, theta is here. And I got these two sample distribution, I have to decide which sample distribution I am going to choose.

Now, it depends on what exactly I am looking for or what exactly I want to achieve from the sample distributions or from this estimation exercise. So, let me actually name these distributions. So, let us say this is distribution B, the green one; and the blue one is distribution A; and the red one, let us say is the population distribution. Now, if I choose B, what I get is, I get in the long run actually, all the mean that I take is actually coming close to the population mean.

So, I get minimum error. So, my error is reduced if I actually take the distribution B. But the cost here is that, the variance is very high, all the observations I get, they are actually spread across, they are pretty pretty far. On the other hand, if I take the distribution A, what I get is that, all the observations are, actually they are very close to each other. So, they are actually, it is efficient, they are basically more reliable; because the observations are coming close to each other, the estimate is more reliable.

Whereas, there is an error; so, it is not really giving me the true estimate of the population parameter. So, there is some sort of error here. Now, depending on what I want, if I am okay with allowing some error, but I want my estimator to be really consistent, it has to be like really efficient, then I have to choose distribution A. But if on the other hand, I am okay with some error, I am okay with some variance, some inconsistency, but in the long run, I want my sample mean to be equal to the population mean, the expected value of sample mean to be equal to the population mean; so, I perhaps prefer to choose the distribution B.

Now, so, that is basically the idea behind the problem. Now, let us try to numerically express the bias and variance that is involved in a model fitting. So, let us say my model has; let me write down; population parameter is theta, whereas the estimator is, let us say Z. Now, the cost; so, if my estimator is Z, so, in case there is no bias, if no bias, then I can write, expected value of Z is going to be, is equal to theta; but I am claiming that there is some bias.

So, if there is some bias then there is a cost, and the cost is going to be Z minus theta. Now, I always express my cost as square of the error; so, I will take a square of that. So, cost is nothing but Z minus theta whole square. Now, if we want to take an expected value of this cost, so, I will write expectation of Z minus theta; let me write down like this. Now, I can expand it as expected value of Z minus mu Z; mu Z is nothing but the expected value of Z plus mu Z minus theta; and I do; let me use another bracket here; square.

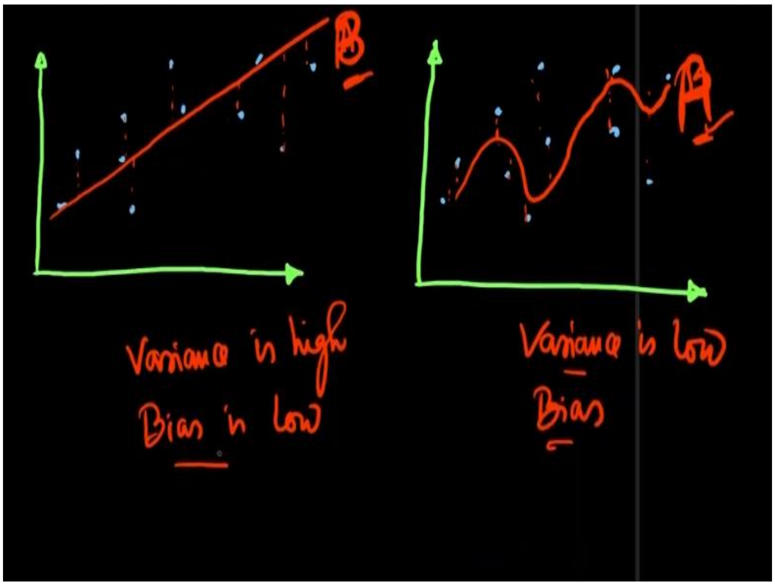
So, I will write down, here, the mu Z is nothing but, mu Z is nothing but the expected value of Z. Now, if there is no bias, my mu Z would have been equal to theta; but there is a bias, right; so, that is where the problem is. And I want to see what this expression would look like if I expand it further. So, if I expand it further, it is going to be expected value of Z minus mu Z square plus 2 into expected value of Z minus mu Z mu Z minus theta plus expected value of this; theta square.

Now, let us look at the second term here. If I actually expand it, expected value of Z minus mu Z is going to be expected value of Z and expected value of mu Z. Now, expected value of Z is nothing but mu Z; that I have already written here. So, here, I will write mu Z; and mu Z is constant there; so, mu Z; essentially, it is equal to 0. So, my second term cancels out. And what I am left with then, finally is that expected value of Z minus mu Z square plus expected value of mu Z minus theta square.

Now, that is again a very interesting observation that we have got, because, this term Z minus mu Z square and expectation of that is nothing but the variance of the estimator. And the other term, as I said, mu Z is the expected value of the estimator; so, if there is no bias, this essentially is equal to theta, which is true population parameter; but since there is a bias, the difference actually tells about the bias; so, it is essentially the bias square.

So, essentially, the cost function is the sum total of variance of the estimator and the bias square. So, that is where the trade-off is. So, it is essentially the sum total of these two. Now, if I actually try to draw it, if I actually try to show how this matters when I am actually trying to fit a model; so, let me actually try to draw that.

(Refer Slide Time: 08:47)



So, let us say I have, I actually plot lot of data. Let us say my data looks like this, something like this. And let us say I add some more points and I actually want to draw a line. Let me actually repeat the same points here, so that we can understand the differences; I will use the same pattern. So, it is pretty much like similar, if not the same. So, if I want to draw a line, let us say I want to draw a line where I am okay with allowing some variation, but I want to draw a line which will fall somewhat in the middle of all these points.

So, let us say I draw something like this. On the other hand, I want to draw a line which will try to sort of mimic all the data points in the best possible manner. So, what I will do is, I will try to draw something like this, let us say. Now, there is a difference between the two model fittings. So, this is the final regression line we got. And when I get the regression lines, what is happening in case of line A and what is happening in case of line B, that we need to understand.

So, what is happening here in case of line A is that, it is going kind of from the middle of all these points. So, what it means is that, if I actually get all these errors and take a square and sum them up, the error is going to be a little high. Whereas, on the other hand, the second line, it is actually touching up on all the different possible points. And what we get is, when I actually take the differences and square the error and sum them up, in case of the line B, it is going to be less.

So, in case of line A, the variance is high. Whereas, in case of line B, the variance is low. But what is happening to the bias? Now, what is happening to the bias is, in the case of line A, you are sort of going through all the different points, the middle of all the points; what is happening is, the bias is actually low. On the other hand, if you take B, and since it is actually trying to touch upon all the different points; what is happening is that there might be bias.

Now, essentially, this is the reflection of what you have seen here. So, or actually, I should have called it B, and I should have called it A. So, my B and A here, are basically reflection of B and A here. So, in B, in this case, I have all the extreme variances. Similarly, here, I have basically taken into account all the error terms and their square. And whereas for A, we see that the variances are less.

So, it means that all the different points in A, they are very close to each other, the model I have fit, they are kind of capturing all the different points pretty close to each other. So, the variance is low, but there could be a bias. There could be a bias here, which is there in this case, because it will not approximate the population parameter. Whereas for B here, since it is going through the middle of all the points, so, it is likely that the bias is low.

And it is similar here in the case of B; the B distribution, the distribution of the estimator is actually, the estimator is actually coming close to the theta. So, that is what is the trade-off. So, depending on what model we fit, we have to either compromise on the bias aspect, we might have to introduce some bias; or we might have to compromise on the variance aspect, we might have to allow some variance to be present in the model.

So, that is basically the concept of bias variance trade-off. And this is extremely important in statistical model fitting, in econometrics, in machine learning; we often talk about this bias variance trade-off. With this, we will end the lecture on bias variance trade-off.