

## **Applied Econometrics**

**Prof. Tutan Ahmed**

**Vinod Gupta School of Management**

**Indian Institute of Technology-Kharagpur**

**Lecture No. #03**

### **Correlation and Causality (Contd.)**

(Refer time: 00:27) Hello and welcome back to the lecture on applied econometrics and we have been talking about the difference in the concept of correlation and causality. And this is a continuation of the previous lecture, and we are talking about the difference in what and why and you know how that and that is actually related to the concept of correlation and causality.

So, we have basically seen a series of what questions are actually leading to a why questions and there is a hierarchy of this what question. So, you know you keep on uncovering the what's. It is like, how you peel the onion, right. So, there are multiple shells that you keep on uncovering, so one what leads to another what, and that what leads to another what and so forth.

So, the more, the the deep you go, the more you know sort of clarity you get, you know so far as clarity is concerned, okay. Now we will try to understand the structure as in how we really go to the process of discovering causality. How we actually know the the fact, okay. So let us uh, let me actually, you know try to represent it diagrammatically or give you some ideas, how to actually do this uncovering, okay. (Refer time: 01:30)

So let us say, A is causing B, okay. So that is, let us say, that is, that is, that is an idea, that is the hypothesis. We really do not know if A is causing B, but we have, we have some idea that perhaps A is causing B. Now how do we know for a fact that A is causing B, okay. So, there could be, you know like, we can see that, you know when B is happening A is present, or when A is present, B is happening.

One example, could be smoking cigarette and you know and getting cancer, okay? So, if people are smoking cigarette, you people are also getting cancer, or people who are getting cancer, if you sort of do a diagnosis, you will see that they are smoking cigarette, but it is not always the case. But we come to that. But let us say that is what you try to understand if you know for a fact if smoking cigarette is causing cancer.

Now how do we really come to the causal part? How do we really come to establish that this is actually happening? So just the two factors are happening together, just that two factors are varying together does it ensure causality? Now I will give you some scenarios. So one could be, let us say, you know there are so many other factors. Let us say there is a factor x, there is a factor y.

And there is also our A and there is a factor z. And all of these, let us say they actually causes B, okay. There could be some other possibilities. So, there are so basically, cancer can be caused by some other factors, including A. Now let us say there there is another possibility that all right, so A is actually causing some, you know u and u is causing B, whereas your x, y, z can also cause B, right?

That could be another possibility. Let us say there is a factor, which is let us say v, v is causing A, and v is causing B, and there are x, y, z, which also cause B, okay. Now pause for a moment and see the the structures that I have just drawn. So, in the first case there are many factors, which are causing B, including the factor A. In the second case, A is responsible for B, but not directly, via u.

So, A is causing u, u is causing B, okay. In the third case, v is causing, so there is a factor called v and v is causing A. Whereas v is also causing B. So, because of the presence of v, both A and B are created, okay? And there are other factors like x, y, z, which can also cause B. So, it is very important for us to actually understand the role of A in in causing B, okay.

There could be another scenario, let us say where, you know your A and B are just covarying just randomly just basically by chance. You have no idea if if they are actually related or not, but for whatever reason, they are actually varying together. It is just let us say it is an accident, okay. Some some, you know accidentally, all of the observations you have got you found that both A and B are varying together.

Now how do we really ascertain the causal link? This is our question. How do I really know if A is for a fact causing B, okay? Now let me actually try, let us try to see how, you know how do we further conceptualize the problem. (Refer time: 04:58)

So let us say this is our solution space and there are many possibilities, many factors, just I have drawn. Let us say, A is one and there is x, y, z, u, v and all of these, you know whenever

we see their presence, we also see the presence of B, okay. So that is that is kind of the problem statement. Now let us say A is actually causing B.

Now let us say cigarette smoking as actually causes cancer, but not all cigarette smokers are actually having cancer. So that is another problem. So, A can cause B, or you know some for some people A is not causing B, okay. So that is basically an understanding which you have to use the, you know we have to understand probabilistically, okay. Whether there is a probability that A can cause B, you know.

And you know what is the probability and when you can actually say, the causal part is involved, but we will come to that later, you know the probabilistic part. But just to understand, you know whether something is actually causing A, you know causing something is you know causing something else like whether A is causing B. So, we have to have some process.

We have to have some uh path delineated for us to actually uncover the whole uh process. So let us say, one way to do that is to repetition. (Refer time: 06:16)

So, one has to see the repetition. So let us say, you know previous example of cancer and cigarette smoking. Let us say, I have seen, let us say, some researcher is interested in this question. And let us say he has taken data from uh from say UK Population Survey, okay. And they have this data on people having, you know people who are addicted to cigarette and who got the cancer.

And they have, the researcher group has actually seen that, you know people who are addicted to cigarette are also likely to have cancer. But does that ascertain the fact that cigarette smoking is causing cancer? So, you need to actually see whether that fact is true in other context as well. So perhaps the researcher group will say, all right, so there are some reasons, this particular cohort, we see these kinds of results.

But perhaps we need to see more. So, what they will do, they will go back to UK Population Survey for past 100 years, let us say. And in past 100 years, let us say that this data on people having you know cigarettes, smoking cigarettes, and getting cancer, and they actually see that it is indeed true that, you know people who are actually smoking, they are actually getting cancer.

And that is true, not just for this cohort, but all the cohorts they have seen for the past 100 years. All right, so then the question will come that could there be something in the genes of people living in UK, that actually is causing cancer. You know like, that is there is a particular variant of gene they have, and you know that is just causing cancer and the fact that, you know cigarette smoking causes cancer may not be true, if they go to Asia or say United States or some other continents, okay.

So, then the researcher group is further sort of, they will expand their research, they will go to Australia, they will take data from Australia, they will take data from Asia, they will take data from United States. And they will see that actually, for all these continents uh, it is actually the people who are smoking cigarette, that actually causes cancer, okay.

So even now, now that the the correlation, the correlation between cigarette smoking and cancer has become a little more prominent, because we can see the relationship across different context. So, we are basically confirming our belief about this correlation. We are yet to reach to the causality. We are still talking about correlation, but we are confirming our belief about this correlation, all right?

Now how whether this uh repetition of certain thing actually helps you to prove causality, perhaps not. And here I will talk, I will give you an example. So, the previous example was cigarette smoking, cancer, and it was for say UK and other continents for a long-time horizon, okay. And all I am saying here is that, if that is the case, then it is only going to confirm my belief that cigarette smoking is causing cancer.

It is still not proving. And why it is not proving and I will come to that. So, this is another story, okay. And the story is about a philosopher Bertrand Russell, you might have heard of Bertrand Russell. So, this is a, there is a term called Russell's chicken. Now what is Russell's chicken? So, it is the story of a chicken, okay? The chicken is living in its owner's home and the owner actually feeds the chicken pretty nicely.

So, every day in the morning, it gets his breakfast and other meals the chicken gets and chicken you know it gets a sort of very good treatment. So, it is like, it becomes a routine for the chicken to get that. To get that or you know meals in time, right. Now so the chicken would believe that perhaps this will continue. You know the the meal will come in the given time until till the date he was actually it was actually slaughtered, okay.

So, the chicken, you know you know it was getting the meals in right time and then one day it got slaughtered. So, then the question will come, you know the, you know why the owner has

actually, you know the chicken might ask why the owner was actually stopped the meal and it can only ask when its dead already, right? So, the, it is actually a very tricky question because that actually talks about the question on hypothesis.

You know like, which which one is true, you know which one for a fact is true. It can be sort of, you know called true unless proven otherwise. So, you know the chicken can think that the food will come as per the schedule unless it was proven that it is not, okay. And this is something you can also you know you might have also heard of; you know Karl Popper who talks about the theory of black swans.

So, all swans are white till the time you see a black swan, right? So, you really do not know what is causing what, okay. So here the example, to illustrate it further, I will see the I will say that, let us say in the previous example, where the owner was feeding the chicken, the owner was at the same time, let us say was feeding his dog, or let us say it was also feeding his pigeon, okay.

Now the owner actually ended up killing the chicken, okay. So, it was feeding, it was feeding, and then it killed the chicken whereas it did not kill, the owner did not kill the dog or pigeon. Now why the dog and pigeon were not killed and why the chicken was killed.

So, till the time all the animals all the animals were getting regular feed, so that that means perhaps the owner is doing out of love okay, out of love for the chicken and out of love for the dog and out of love for the pigeon. But when the owner actually ends up killing the chicken, so it was perhaps you know the the causal, the causal link.

The why owner is feeding the chicken is perhaps, you know proven that the owner is owner was not feeding the chicken because out of because it loved the chicken but rather, the owner actually wanted to eat the chicken after after, you know whatever time. So now we can further ask that, you know why owner actually did not love chicken, whereas the owner actually loved the dog or the pigeon.

So, you have to go to that, you know human brain, like why certain, perhaps certain certain behaviour of the dog or the pigeon actually, uh you know the owner actually sort of, you know wanted that their company and the owner might have liked the dog or the pigeon, but it the for for when it came to chicken, it was not that that you know sort of that chemistry, that that brain chemistry that actually happened.

So, owner actually wanted to eat the chicken. It might also be because of the historical learning of the owner. So, you have to go much deeper into neuroscience to understand the reason why owner loved the chicken. Why owner loved the dog and pigeon and did not kill. And on the other side, it did not love the chicken and killed, okay. So that is basically, uh you know how you go to understand the layers of what questions, okay.

So, what owner did and then you ask why. And the moment you ask why you again ask, so what exactly is the brain chemistry that happened for the owner, so that it killed chicken, but not the dog or not the pigeon, okay. So that is how we will try to understand these layers of questions. Now, so that is basically you know how we sort of uncover the causality. So, it is just a continuation of the previous lecture.

And in the next lecture, we will further see how why it is so important for us to you know have an understanding of causality. And uh you know what is the role of econometrics, which, which we are trying to learn in uncovering the causality, okay, the the scientific part of it. So, with that, we will actually conclude the lecture on causality. Thank you.