Spatial Statistics and Spatial Econometrics
Prof. Gaurav Arora
Department of Social Sciences and Humanities
Indraprastha Institute of Information Technology, Delhi

Lecture - 18B
Casual Inference in spatial regression models

So, welcome back to the second part of the 18th lecture on Spatial Statistics and Spatial Econometrics. And in this lecture, we will be discussing Causal Inference in Spatial Regression Models. And when I speak of causal inference I am specifically referring to assumption 2 of the linear regression model that we covered in lecture 16.

So, a very quick recap for your benefit, you know the linear regression model a simple linear regression model looks like the following. It is y_i equals beta 1 plus beta two x_i plus ui right? Assumption 2 states that expectation u_i given x_i is equal to 0, that is to say, given every value of x_i the expectation the value u_i the random error will be in expectation 0, that is to say, that model error is going to be 0; for every given value of x_i.

This is critical for us to simulate the Ceteris paribus experiment for causal inference, right? What was a Ceteris paribus experiment? Well, we were looking at an example of the impact of the number of rooms on housing prices as an example. And we said that if the number of rooms goes up by one unit that is R_i went to R_i plus 1.

Then whether or not this change in the space or the spaciousness of a house can be causally related or causally linked to the change of price from let us say P_i to P_i tilde. This change can only be causal only if all else is held constant and to be able to ensure that all else is held constant what we must also ensure is that the expectation of u_i is the same at R_i and R_i plus 1 right?

This meant that we wanted to ensure that the expectation of u_i given R_i is the same as the expectation of u_i given R_i plus 1, both of them are going to be 0 by the definition of the assumption. So, the idea is to be able to attain causality a causal linkage. We also saw and covered in that lecture that this is also key to the unbiasedness of least squares estimate.

So, data-driven estimates of these beta 1 and beta 2 are only unbiased if you have A2 satisfied. So, now, we will figure out how this manifests in a spatial regression model. So,

this is a classical traditional regression model as we are used to. Now, we are going to move to a spatial regression model and see how this condition comes about.

So, in order to see that we are going to study a seminal paper by Charles Manski, who is Charles Manski who is an economist who wrote a seminal paper on identifying the endogenous spatial effects. And he called this problem a Reflection Problem. This reflection problem as stated by Manski relates very closely to the endogeneity of spatial effects or the unidentifiability of the spatial effect.

So, let us define these effects let's go over Manski's definition and then adapt it then we will adapt it to the spatial regression models. So, Manski defines endogenous spatial effects as the propensity of an individual to behave in some ways that covaries with the prevalence of that behavior in some reference group containing the individual now that is a very heavy sort of a statement.

What it is saying is that an individual the way an individual behaves or the decisions that an individual takes if it covaries or correlates with the behavior or the decisions of a group in the population that we are working with, right? So, an individual's behavior is somehow mirroring the behavior of a group.

Then what happens is that we are not able to understand whether the group is impacting the individual behavior or simply reflecting it. So, this is the core reflection problem. Wherein, if I am trying to understand the behavior of an individual and I find that behavior to be similar to other individuals in some reference group then we do not know who caused that behavior, is the group causing that individual to imitate that behavior or is it simply a reflection of the way the individual is performing or behaving in a population.

What is this reference group? Well, the reference group first of all is within a population of interest. So, we are looking at population, we are trying to understand population dynamics, and we are searching for a reference group. It could be a socioeconomic identity a family background or even an economic class.

It is a strata, a group within a population where people behave in similar with similar traits in terms of their consumption patterns, in terms of their occupational patterns, and in terms of their day-to-day everyday living choices. Now, that is what you know an endogenous social

effect is where an individual's behavior, there is a propensity of that being reflected in a group's behavior and vice versa.

Now, endogenous social effects are also called as now these social effects, these endogenous social effects have been very popular in sociology, in social psychology. And now also in economics, Manski wrote this paper in 1993. So, economists were grappling with social effects for quite some time and we will see how.

So, the point is that the terminology with which endogenous social effects appear in different literature is quite varied and it has many names. Those names are social norms. So, sometimes there are these social norms or peer influences that people are impacted by their peer group, right? Neighborhood effects are something that we have seen with housing prices conformity, imitation, contagion, bandwagons, and herd behavior.

All of them are referring to the fact that people are behaving as herds, you know everybody tends to converge with the same ideas same decisions, and so on, right? So, there is a tendency to conform to a social group. It is hard to defect from you know what everybody else may be doing in a particular social group and an ethnic group or a socio-economic group or so, or family background, you know it drives a lot of how people make decisions in the real world.

Then there are things like interdependent preferences, as economists we study these devices called preferences which are about how individuals would order or rank elements of their choice set.

Now, if an individual is ordering or ranking elements of a choice set a choice set can be consumption goods it could be, it could have recreational parks, it could have anything you know sports, which sports channel to watch, and so on. if my preference of how I order the sports channels that I like to watch is dependent on other individuals in my peer group then that exhibits endogenous social effects.

So, as I have said economists have been concerned about endogenous social effects that are mediated through markets. So, typically when economists talk about endogenous social effects they are talking about the effects that are generally mediated through markets.

The first and most common is the price-mediated effect where how does an individual's demand for a product vary with market price. This is a typical question that an economist asks is to say how will the demand for milk vary by its market price? How will the demand for cigarettes vary by their market price? These are questions that are found in economics literature there is a large literature of this kind.

Now, this market price; however, is determined at least partly by the aggregate demand in an irrelevant market. So, the prices that we see in a market for any good exhibiting a general population level or a group level or a city level willingness to pay for that entity, for that object, for that good.

Now, if an individual is buying that good they are somehow imitating the behavior here of their reference group, this group which is also willing to pay that price for that very object right? So, there are these endogenous social effects, but they are mediated through market prices. So, through a market equilibrium. Again, there are models of oligopoly that posit reaction functions that link an individual firm's output with aggregate industry output again.

It is a very similar idea, but now on the using a production function rather than the consumption or consumer preferences. Now, if decision-making is costly, if it is costly to make decisions, if it brings stress to someone in making decisions, if someone feels that they have to acquire a lot of knowledge they have to read a lot before they make any decision, right? Then the tendency is to rely on someone else's to copy someone else's decision, right?

This is also a form of endogenous social effect that economists have documented. I am providing you with a reference to 1980. So, in 1980, Conlisk was documenting decisions that were very costly either on time or some other factor knowledge capital or some other factor. It is a costly decision. So, an individual or a group of individuals are found to imitate others just because it is very hard for them to find out all the information about the relevant decision that they are making.

There is also documentation of residents choosing not to live in neighborhoods where the percentage of residents from their own race or ethnic group is below a threshold. Again, an individual decision somehow imitates the decision or reflects a decision of a reference group. Here in this case the reference group is not through markets, but by race or a social class a social strata which could be which could be ethnic groups or races or whatever, right?

So, all these examples existed in the economics literature which were then formalized by Manski as endogenous social effects as follows.

So, Manski states that the reflection problem arises when an econometrician is a researcher or an analyst that is us observing the distribution of behavior in a given population and desires to infer whether the average behavior in some group influences the behavior of the individuals that comprise that group.

Now, this is a heavy statement. So, let us break it down into pieces. So, we are talking about an econometrician who is desiring to infer something. So, the econometrician is desiring to infer whether average behavior in some group will influence the behavior of individuals in that group. And what is the information set for the researcher or the econometrician? It is the distribution of behavior in a given population.

So, you are able to see the population at an aggregate level, if you are able to for example, see the groundwater extraction at the country level for different districts or different villages then you are trying to understand whether an individual village is somehow exhibiting the average behavior of all the villages in a given state or a given region and so on, right?

So, when we are trying to say that average behavior in some group will influence what we are interested is a causal link and not a correlation, right? This is the key in social sciences you know we are mostly interested in causal links and not just correlation. So, now, the term reflection signifies simultaneous movements of an individual and her reflection in the mirror here, the mirror being the reference group.

So, the fundamental query is does the mirror cause the individual's movements or merely reflect them, right? So, the problem of causality versus correlation is that, if we are not careful about which direction the effect has gone they are simply associations, they could go either way. It could be the mirror causing an individual’s movements or it could be a reflection of the individual’s movements, right?

So, this is when we are trying to put a causal link where we say, no the group behavior, average group behavior is influencing individual decisions that are directional you know the establishment of a link in terms of a quantitative effect is called causal inference. So, the reflection problem is that I am behaving in a certain way, my peer group is behaving in a

certain way from a distant to a distant observer we both look quite similar in terms of our behavior.

Let us say in terms of investing time in studying econometrics or in times the number of hours I invest studying in a week versus leisure. So, if me and my peer group are exhibiting similar features in terms of the time we invest in studying and in recreation then it is very hard to figure out whether I am reflecting the group's behavior or the group is reflecting my behavior, right?

It is like two mirrors kept in front of each other right? And, but we are trying to understand a directional effect that you know what is the peer effect on my day-to-day activities.

So, in this spirit, we are now going to with this background and understanding, we are now going to consider a linear model. And we are going to look at the endogenous social effects through a linear model.

So, we are going to work with a population that is characterized by a value y x z and u; y is the outcome variable or dependent variable like for example, house housing price; x is the characteristic describing an individual's reference group. Now, a reference group could be an ethnic group, it could be a family background, could be an economic class, right? It is a reference group to which this individual affiliates. It could even be a union right of some kind, it could be a student body of some kind, right?

y is my decision, my outcome variable that I am wanting to, I am trying to understand, I am trying to explain. x is the group in which I being observee, you know belong to or affiliate with; z are the factors, z and u they both affect y directly. For example, the number of rooms in the case of the housing market affects y directly the public amenities in a neighborhood effect, y directly that is housing price directly right?

The observed ones in these direct impact factors are denoted as z and the unobserved ones are denoted as u. Now, u we are working with regressions is going to be a model error. So, u is going to be a random variable. So, we specify y equals alpha plus beta, expectation y given x. What is expectation y given x? This means it is the mean or the average group outcome right?

Whereas, on the left-hand side what you have is individual group outcome. How is it group outcome? Because it is conditional on some characteristic of the group itself right? Then plus

gamma expectation z given x, this expectation z given x is the attribute, the average attribute or factor size for the group that is let us say in the housing price if I am looking at a certain group which is let us say near the railway station.

Then I am going to see the number of rooms average number of rooms in homes or houses located near the railway station. If near the railway station is x then that is what E z given x, this is the mean of explanatory variable z in the reference group. And eta z plus, now, eta is model parameter z is the direct impact that is the number of rooms of that particular house that we are studying the price for.

So, z is the individual-level explanatory variable, right? And u is the model error something that we have talked about many times now right? And also I am saying that expectation u is conditional on z and x is delta x right? So, the mean of error, conditional on the group characteristics, and the explanatory factor right is equal to delta x.

Remember, this is saying that this is not equal to 0 if delta is not equal to 0. This is going to spell trouble in terms of the second assumption. Now, beta not equal to 0 refers to what is called the endogenous social effect. What is the endogenous social effect? The endogenous social effect represents the propensity of an individual to behave in some way that varies with the behavior of the group this is the endogenous social effect that we are talking about right?

This is something we have defined till now; gamma on the other hand is termed as the exogenous effect. The exogenous effect is also called a contextual or contextual effect. Context is something that sociologists and anthropologists take very seriously right. So, gamma here is representing what is called the contextual effect, the contextual effects are wherein the propensity of an individual to behave in some way varies with the exogenous characteristics of the group.

So, now an exogenous characteristic that is explaining why we take this exogenous characteristic aggregated onto their group level and then look at the effect of that on the individual y. So, this is a variable that is saying now the average number of rooms in homes located near the railway station where near the railway station is x, which is characteristic of the group that we are studying.

The last one which is delta which is coming from the model error is called the correlated effect; wherein, individuals in the same group tend to behave similarly because they have

similar individual characteristics or face similar institutional environments. So, this is a group-level characteristic; obviously, x is about groups, and it's coming through the expectation of u which is the error unobserved terms.

But in mean or on average value this is the effect, the propensity, or the component that is causing equivalency or similarity in behavior between the individual and the group by the fact that individuals in the group are similar right? They face similar institutional constraints and so on.

So, now as I said this condition that expectation u given z comma x is non-zero is where A2 is violated that is we no longer have causal inference we merely have a correlation or an association, how do we establish this mathematically?

So, now, let us look at the model that we have y equals alpha plus beta expectation y given x plus gamma expectation z given x plus eta z plus u. One thing that I have not said till now alpha, beta, gamma, eta, and delta are all model parameters. Then, the mean regression of y on x and z, mean regression of y on x on z is talking about a situation where you have a plot a scatter plot of y and x. And the regression line is expectation y given x right. This is the regression line equation or regression line.

Now, instead of y given x, it is saying to run this regression the mean of y conditional on x and z. So, now, in the conditional statement, I have two entities x and z and I will take this and I will run this expectation on both sides. So, I am going to take expectation on the left-hand side, I am also going to take expectation on x z, and then expectation everything x z. That is what I am writing here it is alpha plus beta expectation y given x plus gamma expectation z given x plus eta z plus delta x.

Now, integrating both sides with z reveals what is called the “social equilibrium” equation; z is the exogenous factor, right? If I integrate it all, I am left with the group characteristic x and the endogenous social effect.

So, now I am going to relate these two, and I am going to find expectation y given x. So, from here to here see that z has been integrated out so; that means, that I am doing something like this. So, I have f of z d z on the domain outside. So, I have gone from this to this I have integrated both sides in this fashion I have integrated z out of this.

I am going to have alpha over one minus beta, how did that happen? Because now if when I have integrated this out, I have expectation y given x here, and expectation y given x here, I will bring it here I get 1 minus beta on the left-hand side. And I divide the whole thing by 1 by 1 minus beta. So, alpha over 1 minus beta plus gamma plus eta over 1 minus beta expectation z given x plus delta over 1 minus beta x right? So, you can write 1 and 2, I am sure that you will be able to figure it out.

Now, expectation y given x is the linear function of 1. So, the 1 is you know intercept column of 1's. It is also linearly related to expectation z given x and x. So, it is linearly related with 1 expectation of z given x and x. As a consequence, the model parameters alpha beta gamma, and delta are all unidentified.

This implies that the endogenous social effects cannot be distinguished from either the exogenous effect which is z or the correlated effects which is x. So, the reflection problem basically says I cannot distinguish between the betas, the gammas, and the deltas. I cannot tell beta apart from either gamma or tell it apart from the delta. This is the problem of this effect, right?

So, I am going to end this part of this lecture of lecture 18 here. And in the next part, we are going to now adapt this problem of endogenous social effects, the spatial regression model. And then we are going to see it with the spatial lag probably giving a similar flavor as expectation y given x which is what caused the problem in the previous case. And then from there, we are going to see how we fix this problem.

So, thank you very much for your attention. See you in the next part of this lecture.