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

Lecture - 19A
Spatially lagged variables in regression models

Hello, everyone and welcome back to Lecture 19 of Spatial Statistics and Spatial Econometrics.

In this lecture, we will build on the knowledge that we have accumulated till now about spatial regression models. So, at first, you know we studied the impact of spatial dependence in model error. What does that mean? That means, if I were to think about a regression model y equals x beta plus u which is a multivariate regression model where y is my outcome variable, x are my regressors and beta are the model parameters and u is the error term something that I could not explain in this model, some of these things we have seen quite many times in the past.

Then if there is spatial dependence in model error where does spatial dependence come from? We have discussed that you know what we mean by that, is that the variance of u given x is not equal to sigma squared I_N basically signifies the situation, where the variance of each u_i that is u at each location i is going to be a sigma squared and off-diagonal elements are 0.

So, there is no dependence in spatial error when we relax this and we introduce spatial dependence in regression models that is to say that variance or covariance of u_i u_j where i and j can be seen as location markers. Then this term will be not equal to 0 at least for some i not equal to j locations.

So, what we are doing is we are moving forward or we are relaxing this assumption of spherical errors and we are moving to a situation where the errors may be non-spherical in nature also termed as heteroscedastic systems, and then we talk about what is the impact of having this non-classical situation or setting. And, we have seen that this is going to be quite prevalent with spatial regression models.

In fact, when we studied spatial statistics spatial dependence, spatial clustering is one of the fundamental patterns with spatial data, right? So, when we come to regression modeling, spatial dependence in errors is not surprising. In fact, it is a fundamental feature of the spatial data.

If with our example of the price of homes, you know prices of homes are clustered together as well in their levels in the sense that you know highly priced homes are likely to be located in regions where all other homes are also highly-priced, right? And, that is how we have these you know terms like posh communities which means that in this community all homes are highly priced, right?

So, this phenomenon then turns out as this feature where mathematically we can say covariance of this unaccounted term in my regression model is exhibiting spatial dependence. We looked at inference properties that is you know what happens if this is a situation to my regression estimator and we saw that the regression estimator will still be unbiased, but it will now be inefficient.

We then introduce this FGLS estimator with the variogram model that we looked at in the previous portion of this course, right?

The second type of spatial dependence that we have seen is through Manski's dependence reflection problem, right? Manski's reflection problem applies to settings where let us say, you know in a peer network when I look at the performance of any student, it is not just a reflection of their own aptitude and their effort levels with regards to studying the course, but it also reflects the peer group in which the individual belonged, right?

So, in a way, an individual's performance is a reflection of the peer group's performance and vice versa. In such situations, we have seen that it is harder to argue causal inference and then we studied methods for reconciliation of this reflection problem.

And, finally, we introduced this idea or this notion of spatial lags in regression models to generalize how to work with spatial dependence in specifying spatial dependence either in the outcome mean or in the error term in a more generalized setting like the irregular lattices, right? So, you can go back to previous lectures and revise these concepts, right?

So, we have seen this earlier, but I am just presenting this idea of spatially lagged variables in a regression model. I have this first thing called the spatial lag or spatial autoregressive model, right? It is a model that comprises spatially lag-dependent variables.

An example here is written as W times G something again we have seen in the past that this is suggesting W is called as the weights matrix, the weights matrix characterizes neighborhood properties. It generalizes this notion of spatial lags, right, and W times G is accounting for the average behavior of what is happening in the neighborhoods vis a vis this outcome variable G, right?

So, the spatial lag model looks like the following we have an outcome variable G in this case we have seen an example for groundwater data. So, let us say we are trying to model groundwater dynamics and we have a column of data on groundwater levels at different locations that is let us say different wells, right? And, this G is then modeled as a function of it is lag G_L, which is the lag bearing, it is a weighted average of what is happening around each with regards to each location I.

There is a parameter sitting in terms in front of this G_L parameter, this G_L variable. So, this parameter rho is a measure for the extent of correlation or dependence between the outcome variable and it is the neighborhood average plus our traditional x beta plus u, right? This lagged dependent variable is then included as a regressor or as an independent variable we call this model a spatial lag model.

Note that G_L is nothing but W times G which is to say that there is a N by N weights matrix that characterizes neighborhood properties something that we have seen in the previous lecture. I will just revise it very quickly in a minute and then G is a N by 1 matrix vector. So, overall G_L is nothing but an N by 1 vector, right?

We will emphasize the need to check matrix and vector dimensions going forward when we work with these models because you know space by its own virtue brings in heavier notation keeping track of neighbors. you know I have neighbors as j, k, and l; j may have neighbors as k, l, but, and m, but not i, and so on.

So, if you have those situations arising where some of my neighbors are not you know neighbors of my own, other neighbors then to generalize that notation we use matrices, which are very convenient, but then when we actually do the math and we specify these things we should keep a track of the dimension of these matrices something we will emphasize as we go forward.

The second type of spatial regression model that we have seen is called a spatial cross-regressive model or the SLX model in this case I will be modeling G as a function of X beta, but also including W X which is nothing but the lagged version of X, right?

So, if I am thinking about my example of let us say groundwater data and one of the regressors is rainfall. Now, rainfall is very important in explaining groundwater levels. If it rains more there will be more regeneration and more sort of water entering the groundwater reserve. And, the groundwater level will come up and hence my observation of groundwater levels will change.

Now, you can imagine that when you look at groundwater level or groundwater recharge at any given location it is not only going to be the rainfall at that very location. It is also to account for the rainfall levels around or in the neighborhood of that locations because of various geographic regions, right? The location of interest by itself may be in a low elevation.

So, you know rainfall that happens around might all run off to this location and hence feed into groundwater levels at this given location of interest. So, in those cases a weights matrix can be applied to these variables and we can define a variable XL which is nothing but the lag of the covariance. And, then we can have a vector gamma and u this is an SLX model.

The final form here is called the spatially lagged model where what we say is that we have G equals X beta plus u such that u exhibits spatial dependence. Again, we will carefully one by one review these models. So, there is nothing to worry about, but I just wanted to provide a quick exposition of these so that in enough coming discussions, these notations start to become more and more comfortable with these notations ok, alright.

So, just before we move forward I just want to recall this idea of spatial weights matrix matrices. So, you know the weights matrix characterizes a device that characterizes neighborhood connections. So, we have an abstract situation on the left, which has been adapted from Anselin's lectures. So, we have these 6 polygons, some of them share borders some of them do not share borders and we can focus on let us say polygon 6 and polygon 1, right?

The weights matrix on the right-hand side is representing the neighborhood structure for this given polygon structure, right? So, because I have 6 spatial units of interest, the size of this weights matrix is 6 by 6 because the row represents each unit 1 2 3 4 5, and 6 and columns are representations of neighbors, right? So, columns also have representations of neighbors.

Now, because 1 is not her own neighbor so, the diagonal element is 0, right? It is 0 2 is not her own neighbor, 3 is not her own neighbor, 4 is not her own neighbor, 5 is not her own neighbor and 6 is not her own neighbor. So, all the diagonal elements are 0. The off-diagonal elements switch from 0 to 1, only when these spatial units share a common border, right? So, that is the characterization of spatial linkage in the data, right?

Now, with 1 2 4, and 5 are neighbors. So, we have one sitting right you know at column 2 row 1, column 4 row 1, and column 5 row 1. Note that we still have 0s in columns 3 and 6 with respect to rho 1 because they do not share a border with unit 1, right? My second unit of focus is 6. So, 6 is neighbors with 1 only, has only one neighbor which is 3.

So, in the row that characterizes the neighborhood structure for unit 6. I have 0s everywhere except for column 3. So, I hope this makes things clearer something that we also discussed at length earlier was the row standardizations of these weights matrices and we discussed why that is a critical idea.

So, I am not going to go over it you can go back and look at it, but the point that I am trying to make here is that every time going forward we see a W, I actually imply W tilde that is I am always implying the use of a row standardized weights matrix. We cannot use the row rho weights matrix and it is good homework for you to understand this lecture, you can go back and review the previous lectures and see why is row standardization a very critical entity.

So, now we are going to start with what is called the spatial lag model and look at its characteristics in detail. So, let us write out the spatial lag model and go from there. So, we have y equals rho W y plus x beta plus u. So, we are including a spatial lag, and like I said whenever and wherever I use this matrix W, I mean a row standardized version of it. So, you can go back and check, but it is very important to understand that we are always using a row-standardized weights matrix.

So, I have a column of data that is N by 1, y is N by 1 because I have data that has ID location markers, you can have the coordinates and then I have y, right? So, I have let us say 10 data points then I have my x y coordinates for these 10 data points and then I have values let us say 10, 9, 7, 6 and let us say 10 again, right? So, I have this N by 1 column which characterizes my dependent variable that is what this y in this matrix form of this regression signifies.

W times y we know that if we have 10 entities the size of W will be 10 by 10. So, N by 1 entity the size will be N by N, y we know is N by 1. So, these two combined will be an N by 1. X is an N by K. So, there are K different covariates and you have N data points for each covariate and for each covariate I have my parameter vector given as beta, right? And, u is again N by 1.

So, I have an N by 1 on the left-hand side and an N by 1 entity on the right-hand side which is a must because I cannot really sum two vectors of different dimensions. I cannot be summing N by 1 with some m by 1 or l by 1, right? So, I need exactly the same dimension for each element that is linearly summed together. So, it is a nice check for us as analysts when we are doing these things analytically.

So, here rho is called the spatial auto-regressive coefficient, like I have said this coefficient measures the extent of spillover effects that I have from its neighbors and vice versa, right? So, I also exert impact or spillovers on her neighbors, right? So, the spillovers are both ways right we have talked about that in spatial data it is not like time series, it is not unidirectional, it is multi-directional, right?

And, also we do not have a regular lattice to work with we have a regular spacing of entities in space and they are of different sizes and so on. So, you have to account for those situations right? So, we can call this equation 1 and we can rewrite equation 1 as under we can say y minus rho W y is equal to X beta plus u.

So, now, if we look at the left-hand side as a consolidated regressed and or the dependent variable then on the right-hand side I have something that I am very used to with regression analysis, right? So, this is an important enough equation. So, I am going to call it equation 2. I can rewrite equation 2 as follows. I can say I minus rho W y equals x beta plus u.

So, this I is an identity matrix, now the question is what should be the size of I? Let us think about it now. So, y is N by 1, x is N by K, and beta is K by 1. So, I have an N by 1, and u is N by 1, right? So, I need an N by 1 on the left-hand side, right? W I know is an N by N, rho is just a scalar. So, it is just 1 by 1, right? So, it is just a scalar entity, I because W and I, we have a linear operator which is a subtraction sign in between I must have an identity matrix of size N.

The identity matrix is a square matrix, N by N matrix where the diagonal elements are all 1s and off-diagonal elements are 0, right? So, I minus rho W is some kind of a filter, it is kind of a filtering of the spatial impacts from the overall I, right? So, I have this I minus rho W times y. So, I have an N by N sitting here, N by N times N by 1 is an N by 1 so, kudos, right? So, I have N by 1 again as each element in this equation. I am going to call it 2 Prime.

So, now, this I minus rho W is known as a spatial filter, right I am going to say this is similar to the detrending device or the detrending activity in case of either time series data if you are used to, if you have seen time series data in the past or even when you use panel data, you can detrend your data. So, this is a spatial filter.

As a next as next piece of activity, I am going to pre-multiply 2 prime with I_N minus rho W inverse. What does that give us? It gives me y, it returns back my original dependent variable equals I minus rho W inverse X beta plus I minus rho W inverse u. Again I will check dimensions very very important very very important, check dimensions because we are working with a heavy matrix.

Matrices you know checking dimensions will provide us a nice quick check on whether we should move forward or hold on and make sure that we are, you know whether we should revisit these things. So, I am going to do that I have a N by 1, I am used to this. I minus rho W we saw here it is an N by N matrix. So, it's inverse the inverse of an N by N matrix is also N by N. So, I have N by N sitting here, I have N by K and K by 1 as X, and beta again I have an N by N and an N by 1 as u.

So, now I have, this will give me N by 1 N by N times N by 1 is again N by 1. So, nice, and N by N times N by 1 will be N by 1. So, we are done. So, we have what we can work it looks like a legitimate regression equation. So, I started with model 1, and all I am providing you are some algebraic manipulations of it, but as we move forward these will come out to be crucial in terms of providing interpretation to these devices.

So, with that we are going to move to a question that is usually the question that applied econometricians ask or applied social scientists ask in general or applied sciences you know scientists are also interested in is that what is the change in y as a result of a small or marginal change in X. So, if I were to shock this regressor or covariant matrix X by a little bit, let us say I move from a lower rainfall to a higher rainfall world lower, population to a higher population world a little bit what is the impact.

So, of course, when I am changing in X, I can change anyone covariate. So, I am talking about these things in the spirit of all else held constant. So, my query is in mathematical language so, I have written an English language term, now I am going to write down the corresponding mathematical language term is basically saying what is the expectation of y given a small change in X. So, I will change X and I want to know as a result what is the average y.

So, this will be equal to, because all else is held constant, so, u will be kept constant. So, as X shocks, as I get a shock in X, I do not see any shock in u. So, I have I minus, by the way, I am looking for a change in y condition on change in X. So, I have I minus rho W, I is of size N because we have seen times delta X into beta, remember beta is a parameter which given data we can estimate from the model, right?

So, you can imagine that let us say you have beta and then you are trying to evaluate some policy or some shock in this variable X. Now, we have this I minus rho W with an exponent of minus 1. So, this can be then further expanded using the power series expansion of an inverse function, and for that, I have a specific resource for you to read where I am going to request you to read inverse series expansion and more generally Taylor series expansion.

If you have heard of this mathematical concept of Taylor series expansion you are basically going to apply that, but as a special case, you know you can read power series expansion and try and understand where we are coming from. So, I am going to just say here to apply power series expansion to I minus rho W inverse. So, I minus rho W inverse can be then written as I plus rho W plus rho square W square rho cube W cube keep going as an infinite series.

And, then obviously, I am going to just put a marker here just a second very very important read Taylor series expansion. So, I am going to just name this new important equation which is the change equation in expectation as equation 3. And, then with this idea of Taylor series expansion, I am going to take it to assume that you guys are going to go back and read what the Taylor series expansion is, if you are a student of economics it is really important.

In fact, Taylor series expansion is what links the comparative statics of the utility maximization model or the expenditure minimization model, profit maximization model, the comparative statics of these models or these algorithms the linkage between them and our regression models and our exercising or mobilizing the causal inference idea is via the Taylor series expansion.

So, that is a matter for separate discussion, but what I am trying to motivate you to do is to definitely read and study the Taylor series expansion at your time it is very important. So, then 3 will become the following. So, the expectation of change in y conditional on a small change in X will be can be written as I plus rho W plus rho square W squared plus rho cube W cube plus keep going times delta X beta.

I mean you can set delta X equals 1, if you want and you will see how my model parameter beta will then explain the total change, right? Now, I can rewrite this as equal to delta X beta because once I multiply I with anything right any matrix multiplied by the identity matrix is the matrix itself plus all of this remaining term of the rho W plus rho square W squared rho cube W cube keep going till times delta X beta.

Now, this is a very tractable intuitive entity. Why? Because delta X beta is something that we are used to if rho was equal to 0, if there was no spatial impact there was no spatial lag term, then all we are left is delta X times beta. This is something that we are very much used to while conducting an understanding of the change in y due to the change in X.

So, in the case of spatial regression analysis, we will interpret delta X beta as the impact of a marginal change in X at the location of this change or shock, right? So, at the location of change you know this is how much impact I will have if there were no spatial impacts there were no spatial spillovers by virtue of this model parameter rho, then all I have is delta X beta, I am very much used to it.

This is further defined as what we call the direct effects of change in X. The second effect you can imagine is called the indirect effect and that is this remaining term which says the impact of a marginal change in X on first-order neighbors, second-order neighbors, third-order neighbors, and so on and this I am interpreting as an indirect effect.

Now, I have written these first-order neighbors, second-order neighbors, and third-order neighbors in different colors why? Because if you pay attention to what you are looking at, rho times W, W characterizes linkage with neighbors. Now, this rho W is a marker for first-order neighbors, this is the first-order spatial spillover rho square W square has this term, W square. What is W square? W times W, right?

So, now, a square matrix multiplied by itself is going to express you can work it out on your own time, but this expresses a second-order spillover effect. So, if I have an impact at the location I, rho W explains an impact on locations which are immediate neighbors of I, but by virtue of change at locations which are immediate neighbors of I will have also an impact on the neighbors of those first order neighbors of I that is this captured by this rho square W square. So, this is the second-order impact and then we have this third-order impact.

So, we can then interpret expectation delta y given delta X as we will rewrite rho I plus rho W cube dot dot dot dot delta X beta this entire entity is called the spatial multiplier effect. So, it is a device that is able to measure the spatial multiplicity we will look at some of the properties through some, you know we will just formalize these things through some notes.

So, the first note that I want to make here is that whenever note that whenever rho is less than 1. In fact, I mean it can be positive or negative we are usually working with positive spatial spillovers I mean that is the idea of spatial contiguity. So, we are usually working with a positive rho, but when rho is less than 1, the spatial multiplier effect declines exponentially with distance, right? What does that mean?

So, if I go back to my formulation, let us say, if I were to draw a graph, where I have distance on or some kind of a lag measure on the x-axis and I have the impact measure on the y-axis, then the origin on the x-axis will mean own location, where the shock has happened. I will then have first-order neighbors, farther than them will be second-order neighbors, and farther than them will be third-order neighbors and I can keep going like this on the x-axis.

What is the impact at own location? The direct impact measure, which is delta X beta. What about at the next first-order location? Well, it is going to be slightly lower, it is rho time's delta X beta. So, you know it is going to be rho W delta X beta, right? At second-order neighbors I am going to have rho square W square delta X beta, and then so on and so forth right?

So, I will have an impact here then an impact here, an impact here, here, and so on. So, what p by virtue of the specification of I model because I have spatial dependence through this row parameter through the specification of this because of the way we specify the spatial lag model the spatial multiplier effect takes the takes an exponential shape in terms of the decline of the effect through a distance which is intuitive, right?

I mean if you are going to have a shock in terms of spatial spillover of course, the spatial spillovers are going to spillovers by themselves are going to become weaker and weaker and weaker as we go away from the entity of impact right? So, that is the first pointer the second note or the second pointer is that in the presence of a spatial lag that is W y, we have the total effect of change in X that is delta X is greater than delta X beta, right?

So, if you ignore spatial dependence, you are going to account for all of this you know you are going to inadvertently either this will all come to the direct effect.

So, it will be a biased effect because it will also account for what is coming as a spillover, right, further, it might actually enter you know as a confounder and hence articulate the bias, right? So, they are the same things, but the idea is that you will be misinterpreting the effects as totals will not come out to be exactly as delta X beta.

So, notes continued, the third note is on the direct and indirect effects of a spatial lag model or in a spatial lag model, right? So, you have seen these things earlier. So, the direct effect due to the change in X, I am just going to say delta X is just delta X beta, and the indirect effects due to delta X is given as rho W plus rho square W square plus rho cube W cube plus keep going times delta X beta, right we have seen this.

This can be written as I minus rho W inverse minus I times delta X beta. So, just simple algebraic manipulations are able to provide us with these understandings. What is interesting maybe is the visualization. Let us visualize these effects. So, let us say I am talking about an entity I right the entity I can have the following neighbors, right?

All these entities could be just say j_1, j_2, j_3, j_4; entities j can themselves have their own neighbor's right. So, j's can have their own neighbors and we can then term them as k_1, k_2, k_3, k_4, k_5, k_6, k_7 right these entities can then further have their own neighbors. So, of course, in space, every entity will have neighbors. The direct effect is the change that happens delta X beta due to shock at the location I, right?

What spills over to its immediate neighbors is given by this W times rho you know multiple this is the first order effect. The second-order effect is characterized then the second-order indirect effect is characterized as rho square W square delta X beta, right? So, these mathematical entities have very interesting characterizations so as to say, spatially.

Now, as an application for the policy, you know what this really means is that the spatial lag model is able to measure the effect of change in a policy variable X at a certain location I that extends beyond I, right? So, a policy shock on location I can have an effect beyond I to its first order neighbors to its neighbors or neighbors of neighbors or neighbors of neighbors of neighbors.

So, which order neighbors that is not the point is a shock at the location I can have an impact on locations beyond I right in other words the spatial lag model will allow us or help us to simulate the spatial imprint of a policy change. So, I am going to write it down that a spatial lag model helps us simulate the spatial imprint of a policy change ok should think about it a little bit.

So, the last topic in spatial lag models is the impact of the misspecification of OLS estimates on inference. So, we have written down this model, we have to say y equals rho W y plus X beta plus u. Here this W y is nothing but y lag, of course, you know the vector notation is used differently between what I am saying, and what is expressed on your screen. So, I am just going to say that y under the bar is the same thing as y vector all I am trying to say is we have a column of data, alright?

So, the question is what if I were to ignore this spatial regression spatial lag term when it indeed should have been in there by virtue of the population process? Well, if it is ignored it is going to go and sit in the unobservable term as if you know I did not observe it because I have ignored it. So, what is the implication?

So, you know Luc Anselin provides a very nice depiction of you know what happens if you ignore these things through a simulation. So, what you see on your screen is a scenario where this parameter rho varies from 0 to 0.9. I have said earlier if rho equals 0, the total effect is just delta X beta, but if rho is not equal to 0 total effects is greater than delta X beta, right?

So, this density function in blue or in purple characterizes the expected value of beta hat when rho is 0. This is the precise effect parameter when there is no spatial order correlation data. So, if there is no spatial order correlation data and we ignore it we are fine. But, let us go back and look at an extreme case where let us say rho was equal to 0.9 which is this brown-looking or red-looking curve.

Here the expectation when we do include beta hat comes out to be much larger than the case when we did not include when rho was equal to 0. Why? Because the total effect is now also accounting for the spillover effects, something that can be a spatial spillover.

So, you know theoretically, we saw total effect is greater than delta X beta when you have a higher when you have a spatial spillover through the spatial lag term. This graph tells me if I were to ignore if I had rho equals 0.9, but I altogether ignored this term rho X rho W y in my regression model I will still end up with the blue curve, right? So, I will still be this representation provided by the expectation of beta hat when rho was equal to 0, right?

The distance between these two articulates the bias. So, what happens if I ignore spatial lags and they should have been included is that I will have a bias in my beta hat estimator right? So, the OLS estimator bias is more and more severe as rho goes up right another thing to look at is that the direction of bias is not just that it will be underreported but may even be overreported.

And, we go from rho equals 0 to rho equals 0.2, and the total effect actually comes down. So, it does not provide me with a direction automatically. So, these things are you know they depend on context they depend on the setting, and so on.

The point here that I am trying to drive home is that we cannot really avoid these effects when they are indeed present in the population process. If we do, then we have a sort of misrepresentation of the population that we are trying to provide an explanation for, and so on. So, that is it for the spatial lag model as a next step we are going to look at the spatial error model in detail.

Thank you for your attention.