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

Lecture - 16B
Spatial Regression Analysis

$We\ established\ that\ \beta_1 being\ a\ causal\ impact\ of\ R_i on\ P_i\ we\ need$

$(A_2)\quad E(u_i|\ R_i) = 0$

$i.e,\quad E(u_i|\ R_i = 1) = E(u_i|\ R_i = 2) = \ldots = E(u_i|\ R_i = 8) = 0$

$\Rightarrow when\ R_i\ changes\ to\ R_i + 1\ all\ else\ is\ held\ constant$

$\beta_1 = \frac{\partial P_i}{\partial R_i}\ implicitly\ assumes\ all\ else\ is\ held\ constant$

$y_i = \beta_0 1 + \beta_1 R_i + \beta_2 x_{2_i} + u_i$

$Now,\ say\ u_i\ comprises\ of\ an\ idex\ for\ "public\ amenities"\ in\ the\ neighborhood\ of\ house\ i.$

$s.t.\ Corr(R_i,\ A_i) \neq 0\ (infact,\ Coor(R_i,\ A_i) > 0$

$\Rightarrow Corr\ (R_i,\ u_i) \neq 0\ (\because A_i is\ in\ u_i)$

Welcome back. So, in the previous part of the 16th lecture, we established that beta 1 will represent a causal impact of the number of rooms on the price that is realized in the market for a particular home, that is the causal impact of R_i on P_i is given by beta 1; if we have a crucial assumption A2 satisfied which means expectation u_i given R_i is 0. This means that whatever the value of R_i, be it R_i equals 1 that is it is a 1-room home, it is 2 rooms home, it is 3 rooms home, 4 rooms and so on and so forth.

The expectation of u_i will remain the same if I were to change the level of R_i right? And in fact, the expectation of u_i will not only remain constant, but it will also be equal to 0, that is the scope of assumption 2. Now, this implies that when R_i changes to R_i plus 1, all else is held constant.

Everything observable and unobserved in u_i is all health held constant, that is why beta 1 in the regression model equals del i by del R_i; this partial differential notation implicitly assumes that all else was held constant.

Now, if this was a deterministic equation, if we only worked with y_i equals beta 0 plus beta 1 R_i, this mathematical formulation that del P_i by del R_i will automatically suggest that all else is held constant right? Now, what does it mean to say that all else is held constant? This means, that if I change R_i to R_i plus 1, then the factor sitting in front of beta 0 remains 1. So, whatever change I see in y_i is going to come from the change in R_i right?

And, this can be true even if I had let us say beta 2 x 2i, right? I had another covariant x 2 i. The formulation that beta 1 equals del y_i or del P_i over del R_i will implicitly automatically hold x_i value to its original x_i. It does not let it move from there, that is what beta 1 measures in a deterministic equation.

The trouble is that in regression we have this random term, this random draw coming out apart from all the deterministic variables. And, the only way to be able to hold the u_i value constant when we change R_i from R_i to R_i plus 1 is this assumption A2, right?

That is why it is perhaps the most crucial assumption of a regression equation because it ensures the causal impact of R_i on y_i, y_i being represented as P_i here right? Now, for example, u_i comprises an index for public amenities, public amenities in the neighborhood of house i right? Now, I am going to sort of use a mathematical notation for these public amenities as A_i such that the correlation of R_i and A_i is non-zero. I am going to go one step forward and, I am going to say R_i and A_i are positively correlated.

That is to say that, if you have better amenities in a neighborhood, you are likely to find larger homes in that area, which is kind of a posh community understanding in the Indian society or any real estate market. Typically, if you have lots of public parks surrounding, very good sanitation, infrastructure, and good roads, the crime rates are under control, and there are a lot of security arrangements. Those are the communities that typically have larger homes and not 1 bedroom condos right?

And in areas which are dense and have fewer public amenities, people live in less secure conditions, the quality of roads is poor, you find smaller 1 bedroom apartments in a city like New Delhi or in general in India right? So, smaller homes or a smaller number of rooms, if that is the index for space will be found in areas where the index for amenities is lower right?

Now, with that understanding, you can say that this will imply that the correlation of R_i and u_i is also by extension equal to 0 because Ai is in u_i. It is not controlled explicitly, we do not have a measure of it. Just because we do not measure A_i does not mean it does not affect P_i. It affects it because, I cannot measure it and I cannot control for it in my regression equation, as an analyst it sits in the error term u_i.

Now, that I have talked about the interpretation of A_i, it says that if you have a higher degree, better degree, and better public amenities, you are likely to have a more spacious room, you know more spacious homes or apartments. And, that would sort of provide you know an understanding that there are bigger apartments and more rooms right?

$Then\ if\ \Delta R_i \neq 0 \Rightarrow \Delta A \neq 0,\ we\ cannot\ simulate\ a\ \text{“}ceteris\ paribus\text{”}$

$experiment\left(wherin\ upon\ change\ R_i \rightarrow R_i + 1\ we\ could\ hold\ all\ else\ constant\right)$

$\widetilde{P}_i - P_i\ is\ attributable\ solely\ to\ R_i \rightarrow R_i + 1\left(i.e.\ \Delta R_i = 1\right) only\ if\ Corr\left(R_i,\ A_i\right) = 0$

$\Updownarrow$

$Corr\ \left(R_i,\ u_i\right) = 0$

$\Updownarrow$

$E\left(u_i |\ R_i\right) = 0$

$This\ also\ the\ diff \frac{b}{w} correlation\ or\ association\ and\ causation.$

$Moreover,\ data\ driven\ estimate\ of\ \hat{\beta}_1 is\ unbiased\ only\ if\ E\left(u_i | x_i\right) = 0$

Then, if delta R_i implies is not equal to 0 implies delta A_i is not equal to 0, we cannot simulate, what is called a ceteris paribus experiment, wherein upon changing R_i to R_i plus 1, we could hold all else constant. So, a ceteris paribus experiment is one, where if I change a covariate x_i or in this case R_i, I can hold everything else constant right?

If I am not able to hold everything constant and something moves in the error term; that means, the expectation if u_i was given R_i is non-zero, then that would imply that the impact that I see on P_i upon this marginal change in R_i is not causal, but merely an association between P_i and R_i ok. So, we have seen this formulation that if I have a change P_i tilde minus P_i, you know is attributable solely to R_i changing to you know R_i plus 1, that is delta R_i equals 1.

You know only if the correlation of R_i and A_i were 0, this implies a correlation of R_i and u_i is 0 and this implies an expectation of u_i given R_i is 0. These things will go both ways. So, the implication is it's double implied. Now, this is also the difference between correlation, or what we say association, and what is called causation. So, when we talk about spatial regression analysis, we would then interpret these items in terms of spatial dependence.

So, if spatial dependence interferes with expectation u_i given x_i that is when we have a problem in terms of establishing causality in a spatial regression model. If spatial dependence sort of is the one which is correlated with one of the $x_{i's}$, then we are in trouble, right? If prices were correlated in a community, let us say they are spatially correlated in the sense that higher-priced homes are clustered together and we do not control for it, right?

Then, you know that will also be an index for higher room-sized or higher more spacious homes clustered together. And, if we do not control for such a phenomenon, then that phenomenon is sitting in my u_i term and that will start to interfere with my causal inference. And, this is something we will study in a more conceptual setting you know going forward. Moreover, the data-driven estimate of beta 1 that is beta 1 hat is unbiased only if you know the expectation of u_i given x_i is 0 right?

So, the regression estimator that is the data-driven estimate of beta 1 hat is also unbiased, that is it is a good representation of the true model beta 1; only if you have an expectation u_i given xi equals 0. Why is that? Because look the definition of beta 1 arises from the fact that P_i or you know is sitting on the left-hand side and R_i is sitting on the right-hand side. So, beta 1 by definition measures a unidirectional relationship which goes for it starts at R_i and ends at P_i.

If the expectation u_i given x_i fails; that means, the causality is not in the direction of R_i to P_i rather it is merely an association that can come from P_i to R_i as well. So, it is a circular loop, a loop where we cannot identify which direction the impact is, that is by definition merely correlation or causation, right? So, if the expectation u_i given x_i fails, then that is we cannot simulate a ceteris paribus experiment, that is we cannot establish causality. We are merely dealing with the correlation estimate.

Then, we might as well not go through this pain of conducting regression analysis and may just rely on a correlation you know statistical correlation between P_i and R_i, which is sufficient. We do not need a regression right? So, everywhere and anywhere, you see a regression analysis is done going forward. You should critically analyze whether or not this is a correlation or association or it is indeed a causation impact, causal impact. In social sciences, causal impacts are really important.

It is kind of important because you see there is you know if you think about even the real estate market if developers or builders are building spacious homes; should they be investing in larger space homes?

You can get 1 unit for a very large space versus 4 units in that equal space which is a good deal by the policy of a real estate business right? If the premium of having a larger house does not translate into better pricing then perhaps we should not find larger homes in the market, but that is not the case, right?

So, this beta 1 is not merely a mathematical unit, it has a real-world interpretation and real-world implications, on which you know a lot of investment is riding right? Similarly, if you have programs like the Skill India program, did it create more jobs right? If you do not conduct a causal analysis, then your investment might not produce, and it might go to waste.

If you provide higher MSPs for crops, does that translate into better welfare for farmers? You need a causal inference, correlations are not enough. And, that is where social scientists and more particularly economists are focused in terms of evidence-based policy.

$A3$: $V(u_i|x_i) = \sigma^2$ $(i.e\ it\ does\ not\ depend\ on\ i,\ but\ a\ const)$

$\overline{\curlywedge}\ Cov(u_i,\ u_j|\ x_i) = 0\ \forall\ i \neq j$

$Variance - Covariance\ matrix\ for\ u_i is\ given\ as\ V\left(\frac{u}{nx1}\right) = \left[V(u_1)\ Cor(u_1,\ u_2)\ \cdots\ Cor(u_n,\ u_n)\ Cor(\right.$

$= \left[\sigma^2 \quad \sigma^2 \quad \ddots \quad \sigma^2\ \right]$

$\Rightarrow E\left(u_i^2 |\ x_i\right) = \sigma^2\ \forall\ i$

$and\ Cov(u_i,\ u_j|\ x_i,\ x_j) = 0\ \forall\ i \neq j$

$I_n\ matrix\ form$: $E\left(u\,u'\right) = \sigma^2 I_n$

$= E\left[u_1\ u_2\ \vdots\ u_n\ \right]_{nx1}\left[u_1\ u_2\ \cdots\ u_n\ \right]_{1xn}$

$= E\left[u_1^2\ u_1u_2\ \dots\ u_1u_n\ u_2u_1\ u_2^2\ \dots\ u_2u_n\ \vdots\ u_nu_1\ \vdots\ u_nu_2\ \ddots\ \dots\ \vdots\ u_n^2\ \right] = \left[Eu_1^2\ Eu_1u_2\ \dots\ Eu_1u_n\ Eu_2u\right.$

$$\Rightarrow E\left(u_i^2 | x_i\right) = \sigma^2 \ \forall \ i$$

$$E(u_i, u_j | x_i, x_j) = 0 \ \forall \ i \neq j$$

So, moving the discussion forward, we have looked at two assumptions till now. We will now quickly look at a few more assumptions of the regression model and end this lecture. The 3rd assumption is that the variance, the variance of u_i given x_i is sigma squared right? That is it does not depend on i, it is a constant. But, a constant which is the same for all i's.

Moreover, it says the covariance of u_i and u_j given x_i is 0 for all i not equal to j, that is the variance-covariance matrix for u_i is given as variance of u, where u is a n by 1 collection of all the error terms from 1 to n. This is going to be n by n matrix right? We have seen that the variance-covariance matrix is of is a square matrix of size n by n, if it is a variance-covariance matrix of an n by 1 vector; where the diagonal terms are just the variance terms and the off-diagonal elements are covariance terms and also it is symmetric.

So, you know if you have covariance u_1 u_2, you will have a symmetric term covariance u_2 u_1, where it is exactly equal. It does not matter which order you are calculating the covariance at. So, u_n u_1, covariance u_n u_2 and keep going like this.

So, we have this variance-covariance matrix, and assumption 3 says, that this matrix looks like the following. So, all off-diagonal elements are 0, and diagonal elements are the same, that is sigma squared which can be simply written as sigma squared I_n right? So, the variance-covariance matrix is written as sigma squared I_n.

This could also translate into this implies that the expectation of u_i squared, this is the second moment of u_i is sigma squared for all i right? This can then also be written as so, and the covariance again u_i u_j given x_i x_j is 0 for all i not equal to j. In matrix form, we have to always learn to translate between matrix and scalar forms, because they are very useful in terms of reading papers and textbooks right; most textbooks are written in vector form.

So, this module, this recap is not only giving you a recap of the regression model. But, also I hope it is giving you this translation between the scalar form and the vector form. So, in the matrix form these two conditions can be then compressed into an expectation uu prime equals sigma squared I_n. What does that mean? Let us look at the LHS. So, we have an n by 1 and a 1 by n. So, I have an n by n which is what I should be looking at.

So, I have an expectation operator sitting here, I have u_1 u_2 to u_n, and it transposes u_1 u_2 to un as a row vector. This is when we multiply as n by 1 and 1 by n they are conformable. If we multiply, we get an expectation of u_1 squared u_1 u_2 to u_1 un, then u_2 u_1 u_2 squared, all the way to u_2 u_n. Then, you have un u_1 and keep going u_n u_2, all the way to unsquared. So, all the diagonal elements are squared values. So, the expectation operator is a linear operator that simply goes in and applies itself directly to each term here.

So, this is going to be expectation u_1 squared expectation u_1 u_2 keep going expectation u1 un expectation u_2 u_1 expectation u_2 squared, keep going expectation u_2 u_n. Similarly, all the way to expectation u_n u_1 expectation u_n u_2, and tail until expectation unsquared. Now, this matrix in front of you that is the expectation uu prime is nothing, but the variance-covariance matrix, right? Because, if you look at it carefully expectation u_i squared is nothing, but the variance of u_1 given that expectation u_1 u_i is 0.

Expectation u_1 u_2 corresponds to the covariance between u_1 and u_2 and similarly, expectation u_2 u_1 is corresponding to this. This will imply that expectation u_i squares directly correspond to this you know factor sigma squared and expectation u_1 u_2 are all going to be 0s right; that means that this matrix is nothing, but you know sigma squared and 0s everywhere else. This will imply that expectation u_i squared, given x_i is 0 for all i, and expectation u_i u_j given x_i x_j will be 0, sorry above one is sigma squared for all i not equal to j.

These two combined are also sometimes you know we say errors are spherical. So, the 3rd assumption that we take is the assumption that the variance of each u_i term is a constant and equal across all $u_{i's}$ and the errors are not correlated. So, home 1 and home 2 whatever we have not observed is not correlated by itself. This is the 3rd assumption. Let us move on to assumption 4.

$A4$: $Rank\ of\ the\ covariate\ matrix\ X_{nxk} = K$

$|y_i = \beta_0 1 + \beta_1 x_{1_i} + \beta_2 x_{2_i} + \ldots + \beta_k x_{k_i} + u_i$

$i = 1, 2 \ldots, n$

$$X_{nxk} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}$$

$\rightarrow All\ columns\ of\ X_{nxk}\ are\ linearly\ independent.$

$\Rightarrow Rann\left(X_{nxk}\right) = k \Longrightarrow n \geq K$

$i.e.,$ $we\ have\ at\ least\ as\ much\ data\ as\ we\ have\ variables$

$Relates\ also\ to\ "Dummy\ variable\ trap"$

$Woolridge's\ Advanced\ Econometric\ Analysis$

$A5$: $\quad X_{nxk}\ is\ a\ non-\ stochastic\ matrix.$

$A6$: $\quad u\ is\ normally\ distributed\ i.e.,\quad u_i \sim iid\ \ N(?,?)\ \ \forall\ i$

$A1,\ A2,\ A3\ and\ A6 \Rightarrow u_i \sim iid\quad N\left(0, \sigma^2\right) \forall\ i$

Assumption 4 means the rank of the covariate matrix X which is let us say an n by k matrix is equal to k. What is this X n by k matrix? Well, you can think of this as the X fact vectors arranged in columns. So, my regression equation is y_i equals beta 0 plus beta 1 x_i beta 2 x $_{2i}$, all the way to beta k_x k_i plus u_i. Now, i goes from 1 to n; that means, that all these you know xi's are arranged in columns. So, the first column is just a column of 1s.

How many of these are? There n of these are, there are n rows of 1s as the first covariant. The second is x_{11} x_{12} all the way to x_{1n} and this is the first covariate. So, in the case of my price and you know the portal for house prices the real estate market x_{11} is the number of rooms in home 1 or house 1 x 1 2 is the number of rooms in house 2, we have data for it.

Say we have more data, let us say the size of the kitchen I do not know right? That might be the second covariate x_{211} x_{222} x_{22} and x_{2n}. Similarly, we have x_{k1}, the k^{th} covariate, the kth covariate would be the crime rate near that house in the neighborhood of that house; x_{k1} x_{k2} till x_{kn}. These are all, this is the entire data that is included on the right-hand side of the model.

When we say that this is this has the rank X, what we are saying is that you know all columns of X n by k are linearly independent, that is to say, that you know because I do not have superficial data. If I have let us say the number of rooms and the number of bathrooms, let us say there is always an attached bath to a room. Then, you know including the number of rooms and the number of bathrooms is separate covariates is sort of is redundancy.

So, the point of this assumption is there is no redundancy in the data. The second thing it says is that you know ultimately it means the rank of this matrix X is equal to k implies that n is greater than or equal to k, that is to say, that we have at least as much data as we have variables. Remember, to each k to each k is attached a coefficient beta. So, that is what we are saying that you know to each variable there is at least one data point that corresponds to more than that, that makes sense right?

We cannot have 10 data points and 10 covariates, that is going to be a very difficult situation. Now, this relates to also to what is called a dummy variable trap. I am going to leave you with this interesting term dummy variable trap and I am going to give you a home assignment to study this from the book on econometric analysis by Wooldridge right? You can say Wooldridge's Advanced Econometric Analysis.

You must read the first two chapters and if you read them, you will figure out what is dummy variable trap. You should find this term and study it from the book. This will also give you a very nice review of this what we have covered in this you know in this class till now. Quickly, the 5th assumption of a regression model is that X, this matrix X contains factors that are you know non-stochastic, that is to say, when we say that the number of rooms is 3, we do not mean 3 with some error, we say it is 3 right? It is a degenerate matrix.

Assumption 6 is that the error term is normally distributed, that is to say, that ui is normal with a mean and a variance for all i and these are all iid. If we look at our assumptions; assumption 1, assumption 2, assumption 3, and assumption 6 will imply that ui is iid normal 0 comma sigma squared for all i. This is a crucial assumption.

$Least\ Squares\ Estimator$

$y_i = \beta_0 + \beta_1 x_{1_i} + \beta_2 x_{2_i} + u_i\ ;\quad i = 1, 2...., n$

$Need\ to\ find$: $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\ with\ n\ data\ points\ and\ the\ knowledge$

$of\ above\ true\ model.$

$$\min \beta_0, \beta_1, \beta_2 \quad \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_{1_i} - \beta_2 x_{2_i}\right)^2$$

$yields$: $\hat{\beta}_{0,LS}\ ;\ \hat{\beta}_{1,LS}\ ;\ \hat{\beta}_{2,LS}$

$Predicted\ value\ from\ model\ estimation$: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1_i} + \hat{\beta}_2 x_{2_i}$

$Model\ residual\ term$: $y_i - \hat{y}_i = \hat{u}_i$

$Given\ A_1 - A_6$

$\hat{\beta}_{1,LS}\ is\ B\ L\ U\ E$

$\hat{\beta}_{0,LS}\ is\ B\ L\ U\ E$

$\hat{\beta}_{2,LS}\ is\ B\ L\ U\ E$

$\because Gauss\ Morkov\ Theorem$

$B\ L\ U\ E\ means,\ \ Best\ Linear\ unbiased\ estimators$

$unbiased\ means,\ E\left(\hat{\beta}_{k,LS}\right) = \beta_k$

$Spatial\ dependence\ will\ interfare\ with\ A2\ ;\ A3$

Now, finally, I want to just in this recap, I want to talk about what is called the least squares estimator. I have been saying that ultimately, we want to be able to estimate this beta 0s and beta1 and so on and so forth. So, I have a model, a true representation of the truth given as the following. Let us say I have this representation of the truth, that I want to explain y i based on a model equals beta 0 plus beta 1 x_{1i} plus beta 2 x_{2i} till u_i.

Now, I have data on these things right? I have data on 1 2, all the way till n. I have n values and I want to figure out what is the database representation, right? So, we need to find or estimate beta 0 hat beta 1 hat, and beta 2 hat with n data points for both y for y x_1 and x_2.

We have n data points and the knowledge of the above true model, right? So, the true model is giving me what parameters to be estimated. And it is also giving me the specification, that both x_1 and x_2 are entering as linear regressors. And, you know there is a constant of 1s and there are three parameters.

The model parameters, and coefficient parameters that is beta 0, beta 1, and beta 2 are to be estimated. So, least squares estimator, the algorithm is to minimize what is called the sum of squared errors. So, we take the errors at every time, every instance of my data observation. We calculate the error and minimize it, that is we want to minimize y_i minus beta 0 minus beta 1 x_{1i} minus beta 2 x_{2i} the whole squared. So, I am conducting this minimization exercise while I choose beta 0 beta 1, and beta 2.

It is an optimization problem. I will write my first-order conditions, but ultimately this exercise will yield beta 0 hat least squares, beta 1 hat least squares, and beta 2 hat least squares right? The model representation, the prediction, or the predicted value from model estimation is given as y_i hat which is nothing, but beta 0 hat plus beta 1 hat x_{1i} plus a_2 hat x_{2i}.

So, if I were to visualize this on a scatter plot, you know I have my x_{1i} and y_i of course, there is also x_{2i} which can be viewed in the third dimension. But, I am restricting myself to the cross-sectional representation, scatter plot representation of $y_{i's}$ and $x_{1i's}$. Let us say I have my stylized scatter plot, which I have been working with till now. And, I am able to estimate a regression where this represents beta 0 hat and the slope of this line is beta 1 hat.

Now, to every value of x_i, x_{1i} rather you know I have a true value and a predicted value, and another true value. So, here is the prediction which is y_i hat which is given as beta 0 hat plus beta 1 hat plus beta 2 hat x_{2i}. And, the truth is y_i, there are two truth values here y_{i1} and y_{i2} let us say. Then, you know the model residual term, that is what this model could not explain the distance between the truth and the predicted value that is y_i minus y_i hat is nothing but the value u_i hat right?

Now, you can represent u_i hat values in this fashion. To each x_i we can have more than one u_i hat. In fact, we have typically more than one u_i hat value, if we have a dense enough data set. Now, given very very crucially this is the point where I am going to end this review of the regression analysis.

But, I am going to then articulate the departure from the typical traditional regression analysis to spatial regression analysis from this point that I am going to make next. That is given the assumptions A1 to A6, right? The 6 assumptions that we articulated today, we can say that beta 1 hat least squares is BLUE right?

Equivalently, we will also say beta 0 hat least squares is BLUE and beta 2 hat least squares is also BLUE. This idea of BLUE estimator, you should study from what is called the Gauss-Markov theorem. Again, articulated very well in Wooldridge’s book. Otherwise, also Gauss-Markov theorem is so, popular that you can you know we do not have time in this course to go over it. But, any introductory econometrics course introduces the Gauss-Markov theorem.

So, by Gauss-Markov theorem given assumptions A1 to A6, the beta hats that are arrived from the least squares algorithm are BLUE. What does BLUE mean? Well, BLUE means that they are the Best Linear Unbiased Estimators. They are the best guess of the truth and when it means, what it means by best, we will look at as we sort of you know walk through different parts of this course. What unbiased means that the expectation of beta k hat least squares is exactly equal to the true beta k [FL].

So, you have beta hat least squares is exactly equal to beta k which is the truth. Remember, we are looking at the truth as well as we are looking at its estimate. The estimate is a function of data. Truth is just the model representation of the real world, where y_i can be linked with x_1 and x_2 right? Once, we are given the data, we are practically somehow getting a representation of beta 1 and beta 2, and beta 0.

Is it a good representation? Well, it is good given that we have these assumptions satisfied. And, as soon as we depart from these assumptions in any way, that is where we are going to look at spatial dependence. So, spatial dependence is going to specifically interfere with this. So, spatial dependence will interfere with A2 which is the causal inference, very the most perhaps the most crucial assumption.

And, A3 is the assumption that the variance-covariance matrix is diagonal, off-diagonal elements are 0 and all the diagonal elements are the same. Spatial dependence will interfere with these two assumptions. And, this is where we are going to first introduce spatial dependence or spatial regression as in the next lecture. And, then using this recap, we are going to then depart from or relax the two assumptions that are A2 and A3.

And, then we are going to figure out how we can estimate these models; even when there is spatial dependence and they interfere with the assumption of the least squared estimator, making it a not best linear unbiased estimator. So, it will with A2, it will introduce bias, and spatial dependence will introduce bias. And, with A3, it will not, it will interfere with the fact that the estimator is best or not.

What we will see is that when A3 fails due to spatial dependence, we will have a better estimator. We will have something else, that will be a better representation of the true estimator. So, thank you very much for your attention. I understand that this recap could be a little bit fast-moving for some people, who have not seen regression analysis before. I highly encourage you to go over these lectures again.

This set of lectures 16 again and read Wooldridge's book, read the introduction of regression analysis from that book. Even if you have seen regression analysis before, please read Wooldridge's book. And, after that, I believe the upcoming lectures will become much easier to understand conceptually.

So, thank you very much for your attention and we will see you in lecture 17 next.