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

Lecture - 17C
Generalized least squares estimation in the presence of spatial dependence

Alright. So, welcome back to part 3 of lecture 17. Now, we have established that the least squares estimators are no longer efficient if you have spatial dependence in model errors, and that as consequence the least squares estimators are no longer blue.

$GENERALIZED\ LEAST\ SQUARES\ (GLS) ESTIMATORS$

$IN\ THE\ PRESENCE\ OF\ SPATIAL\ DEPENDENCE$

$We\ moved\ from\ a\ specialized\ var\ -\ covar\ structure:\ \sigma^2 I_N$

$= [\sigma^2 \quad \sigma^2 \quad \ddots \quad \ddots \quad \sigma^2\ ]_{NXN}$

$To\ a\ GENERALIZED\ var\ -\ covar\ structure: \Omega\ = [\sigma^2\ \sigma^2\rho\ \sigma^2\rho^2\ \ldots\ \sigma^2\rho^{n-1}\ \sigma^2\rho\ \sigma^2\ \sigma^2\rho\ \ldots\ \sigma^2\rho^{n-2}\ \vdots\ \vdots$

$Efficient\ GLS\ estimator: \hat{\beta}_{GLS} = \left(X^1\Omega^{-1}X\right)^{-1}\left(X^1\Omega^{-1}\underline{y}\right)$

$\hat{\beta}_{LS} = \left(X^1X\right)^{-1}\left(X^1\underline{y}\right)$

$Scatar\ form: \quad \hat{\beta}_{2,LS} = \frac{\sum_{n=1}^{N}(P(.)-\bar{P})(R(.)-\bar{R})}{\sum_{n=1}^{N}(R(.)-\bar{R})^2}$

So, then what we do is we move on to a new estimation strategy and to what we call a Generalized least squares estimator. So, in general, you know what we have done is we have moved from a specialized variance-covariance structure for model errors, that is in front of your screen which is the one where the variance of each delta S_i is exactly the same which is sigma squared.

And, the covariance between these deltas that is model errors at two different locations is 0. So, the off-diagonal elements are 0 and the diagonal elements are the same which is a constant sigma squared. We have moved to a generalized variance-covariance structure where we introduced non-zero off-diagonal elements which are nothing, but the covariance non-zero covariance between errors of two different locations. When rho is greater than 0, what it is exhibiting is a positive spill over from the unobserved factors onto each location of interest.

So, each location of interest is getting a positive spill over from its neighbour. If the neighbour is nearby the spill over is strong which is sigma squared rho. If the neighbour is further apart the spill over decreases in intensity by a factor of rho to the exponent of how far apart, they are, this is simplistic; however, very realistic type of spatial dependence structure, and that is what you would expect in space.

When you see clusters in a locality of equal intensity, but as you move away there are clusters of different intensity. So, there is less dependence on values that are further apart than relative to the values that are closer together, right? So, this simplistic starting point variance-covariance structure is powerful enough to capture such a real-world spatial dependence structure.

In the presence of this structure when it is exactly known when we are able to exactly write down what our rho variance-covariance matrix says which is an N by N matrix, which is no surprise we are working with N-sized columns for deltas and R's for P's. So, basically, we have an N sized column for each data set we have N observations and the variance-covariance matrix is N by N right? In this case, the way the GLS estimator is written as x prime omega inverse x times x prime omega inverse y.

Now, the beta hat is written in a matrix form, remember till now we have written it all in the scalar form, right? Just to do a slight translation; so, if I were to write an ordinary least square which is the non-generalized form, I will write beta hat LS equals x prime x inverse x prime y. Now, what is to be understood here is that when you move from least square to GLS; what is happening is that in my estimator I am somehow normalizing or weighing for this omega inverse, you know matrix right?

Now, in the case of least squares I could envision doing the same, I mean I could write sigma squared I_N inverse and sigma squared I_N inverse as a sandwich and what will happen is if you solve it, what will happen is that these two terms will cancel.

But when the matrix is a bit more complex that is omega, this cancelling does not happen. So, the generalized least squares make the least squares formulation a special case of itself, right? Now, I am going to just get rid of this sandwich form just for our convenience going forward.

If you want to see the correspondence between the scalar form and the matrix form, you can say, we can just write down the scalar form and you will start to see it.

So, beta hat 2 LS was equal to summation N equals 1 to N P minus P bar R minus R bar divided by summation 1 to N R minus R bar; sorry about the spelling mistake minus R bar squared. So, now, P is my y and x is my R right? So, this numerator directly corresponds to x prime y whereas, the denominator directly corresponds to the x prime x inverse.

So, the matrix form is just a consistent concise form of data, right? The matrix x, what is matrix x, what is matrix y; all of that let us look at it with our example going forward. But, here I just want to sort of provide you a functional form for beta hat GLS and how it is different from the beta hat LS right? Beta hat GLS explicitly accounts for the complex variance-covariance structure which accounts for spatial dependence by the way into the estimator definition by itself.

$\textit{For our example}: P\left(\vec{S}_n\right) = \beta_1 1 + \beta_2 R\left(\vec{S}_n\right) + \delta\left(\vec{S}_n\right)$

$Cov\left(\delta\left(\vec{u}\right), \delta\left(\vec{v}\right)\right) = \sigma^2 \rho^{\|\vec{u}-\vec{v}\|};\ \rho\epsilon(0, 1)$

$\left[\hat{\beta}_{1,GLS}\ \hat{\beta}_{1,GLS}\right] = \hat{\beta}_{GLS} = \left(X^1\Omega^{-1}X\right)^{-1}\left(X^1\Omega^{-1}\underline{y}\right) \ = A\,y \quad where\ A = \left(X^1\Omega^{-1}X\right)^{-1}X^1\Omega^{-1}$

$X\ Nx2 \quad \equiv \ \left[1\ \vec{R}\right]$

$\vec{S}_n$	$P(.)$	R	$Intercept$
$\vec{S}_1\ \vec{S}_2\ \vdots\ \vec{S}_N$	$\left[P_1\ P_2\ \vdots\ P_N\right]$	$\left[R_1\ R_2\ \vdots\ R_N\right]$	$[1\ 1\ \vdots\ 1\ ]_{Nx1}$

$$y \equiv \left[P_i\right]_{i=1}^{N}$$

$V_X\left(\hat{\beta}_{GLS}\right) = \sigma^2\left(X^1\Omega^{-1}X\right)^{-1}$

$=\left[V\left(\hat{\beta}_{1,GLS}\right)\ Cov\left(\hat{\beta}_{1,GLS},\ \hat{\beta}_{2,GLS}\right)\ Cov\left(\hat{\beta}_2, \hat{\beta}_1\right)\ V\left(\hat{\beta}_{2,GLS}\right)\ \right]$

$Issue$:

The above formulation has assured the knowledge of *Ω*. However, in most applications, we will not know *Ω* a priori.

So, for our example that is P Sn equals beta 1 beta 2 R Sn plus delta Sn with covariance delta u delta v equalling sigma squared rho u minus v where rho is between 0 and 1. So, it is positive, but less than 1 greater than 0 right? So, it is a positive spill over, if rho were negative then it would be a negative spill over, that is to say, if I see a high value at a given location in its neighbourhood, I am likely to see a lower value.

So, spatial order correlation can be positive as well as negative. We will probably in most cases you will find cases of positive autocorrelation, it depends right? So, now, when I say beta hat GLS as x prime omega inverse x inverse x prime omega inverse y. What I am saying is that x is nothing, but a vector of 1s that is this coefficient 1, and the R's. Now, I have data where I have data on n locations.

So, I have let us say location S_n so, S_1 S_2 till S_n, and for each location I have a value P which is let us say P_1 P_2 all the way till P_n and then value till R that is R_1 R_2 till the way all the way till R_n. This coefficient of beta 1, beta 1 is a coefficient of this value 1 which is a constant. So, I have an intercept which is nothing, but a column of 1s. So, this fact this 1, this column 1 is what this 1 is representing.

So, we will draw it as a bold just to say it is a column vector and; obviously, its size is N by 1 and so, is the size of this vector R which is also N by 1. Hence, x is nothing, but N by 2, and y is my vector of prices which is again an N by 1 vector right? So, y is nothing, but you know P_i is going from i to equals 1 to N organized as a column vector. So, y itself is an N by 1.

So, what I have is an x transpose is going to be 2 by N, omega is N by N its inverse, inverse of a square matrix is the size of the matrix itself and x is N by 2, x prime is 2 by N, omega inverse is N by N and y is N by 1. When I multiply these, of course, you know there is conformity. So, there we go so, we have 2 by N and again I have a 2 by 2 matrix.

So, the first matrix is a 2 by 2 matrix. The second matrix lets us evaluate is going to be an again 2 by N and a 2 by 1. So, the second matrix is a 2 by 1 matrix. Again, there is they are conformal and the final matrix is going to be a 2 by 1, right? So that means, that beta by itself is a 2 by 1 matrix implying that it is comprising beta 1 hat GLS and beta 2 hat GLS.

So, I will, if I am very concisely able to summarize the solution for beta hat GLS in this 2 by 1 matrix. So, we are moving away from the scalar representation to a vector representation or matrix representation, but they are translated easily, right? So, it is not so hard at the end of the day.

So, now we have our estimator beta hat GLS, we know that it is a random variable y. Why? Because you know it is itself a function of this random variable y which is nothing but the prices right? x's are non-random, what is in the variance-covariance matrix is non-random sigma squared and rho are model parameters. But what is random is y right so; that means, beta hat GLS is also random right?

Another thing is beta hat GLS is linear in y. It is very important to see that it's linear in y because all you are looking at is a constant, a constant here, a constant here. So, basically what we are looking at you know somehow beta hat GLS is being written as a constant A times y right? You can reformulate this whole thing in that format and view this in this format. So, the y is N by 1 and here I am going to have a 2 by N matrix A, where A is nothing, but x prime omega inverse x inverse x prime omega inverse.

Interesting, you know the way to view these objects. So, having understood that I can write down the variance of beta hat GLS. Of course, the variance is conditional on the data that is given to me which is R, I am just going to use x just to keep myself in the same general format. This is going to be x prime omega inverse x. Sigma squared is the model parameter, it is coming from right here in the variance-covariance structure.

I have my sigma square; it is the same sigma squared here, right? It is a scalar entity, a constant, and inside I have x prime omega inverse x, the whole inverse; sorry about that. Now, let us evaluate the size of this matrix. I have 2 by N, N by N, and N by 2. So, I have a conformable 2 by N and then finally, I have 2 by 2. So, the variance of this matrix is a 2 by 2. So, the variance matrix is 2 by 2 which should not be surprising, because, at the end of the day, it is the variance-covariance matrix of beta hat GLS.

So, the variance of beta 1 hat GLS is the variance of beta hat 2 GLS are the diagonal elements. And, the off-diagonal elements are nothing, but the covariance of beta hat 1 GLS and beta hat 2 GLS. This is a symmetric matrix because the covariance of the flip order which is the covariance of beta 2 hat and beta 1 hat is exactly the same as the covariance of beta 1 hat comma and beta 2 hat, right?

So, this is exactly the same here. So, the covariance of beta 2 hat and beta 1 hat. Now so, we are done, right? So, if we have a spatial dependence, a structure with spatial dependence rather than working with beta hat LS which is the ordinary least squares, we should work with beta hat GLS which is the generalized least squares format.

However, there is a slight issue. The issue is that the above formulation is assuming the knowledge of omega, right? This above formulation has assumed the knowledge of omega. However, in real-world cases, in real-world situations, we may not know omega. In fact, we will not know omega a priori, right?

However, in most applications, we will not know omega a priori. In such a case, in such a scenario, we have to back out omega from the data itself. We suspect spatial dependence so; we know there might be a general variance-covariance structure going on.

We are also aware that if there is a general variance-covariance structure going on with the non-zero off-diagonal elements or even perhaps non-constant diagonal elements, then we know for a fact that least squares are not going to be as efficient. So, we must come to generalized least squares, but we do not know omega. Then what do we do?

Well, what we then do is called the feasible generalized least squares strategy. So, going forward in the next sort of slide, we are going to start looking at an example of how you implement the generalized least squares estimation when in fact, you are not aware of the exact omega spatial dependence structure through omega, that which is the variance-covariance matrix.

$F\ G\ L\ S\ strategy\ in\ the\ presence\ of\ spatial\ dependence.$

Say, we have data $\left\{P\left(\vec{S}_n\right),\ R\left(\vec{S}_n\right),\ A\left(\vec{S}_n\right)\right\}$

We are interested in estimating the following regression model:

$$P\left(\vec{S}_n\right) = \beta_1 + \beta_2 R\left(\vec{S}_n\right) + \beta_3 A\left(\vec{S}_n\right) + \delta\left(\vec{S}_n\right)\ ;\ \vec{S}_n \epsilon \left\{\vec{S}_1, \vec{S}_2, ..., \vec{S}_N\right\} \epsilon R^2$$

$$\left(Cov\left(\delta\left(\vec{S}_i\right), \delta\left(\vec{S}_j\right)\right)\right) \equiv \Omega_{NXN}$$

$We\ suspect\ spatial\ dependence\ in\ house\ prices.\ But\ we\ do\ not\ know\ the\ exact\ form\ of\ it.$

So, I am going to sort of title this slide as the feasible generalized least square strategy in the presence of spatial dependence. So, say we have data that looks like the following. Let us say we have data on P which is the prices, we have data on the spaciousness of this housing entity at location n which is R and we have data on public amenities let us say which is A. So, of course, you know we have the price of the house, we have the spaciousness or the number of rooms index and we have a public amenities index.

You know this could be the quality of schools in a neighbourhood or you know that is attributable to that particular house in question, public parks. You know people care about parks when they buy homes, care about school quality around that area etcetera and etcetera.

Now, we are interested in the following model in estimating the following regression model. So, we have P S_n equals beta 1; so, I have the intercept beta 2 R S_n plus beta 3 A S_n. And finally, the model error delta S_n, such that we know that S_n is basically locations S_1 S_2 and keep going till S capital N. All of these are in the second two-dimensional real space, right?

And, we also know that this is not the complete model. The complete model is that we have a covariance of the delta at S_i and delta at S_j, where you know i could go from 1 to N and j could go from 1 to N is a general variance-covariance matrix omega right? So, here what is happening is that we suspect spatial dependence in data, dependence in housing prices right; because you know we live in the real world.

So, if we see housing prices with higher-priced houses located in a cluster together and lower-priced houses located in another cluster at a different location, then we know that prices exhibit spatial dependence. Now, when we have specified the systemic portion of this model, we have not specified spatial dependence in any way. So, where is it all going to be? It is going to all reside in the error term something that we could not or did not account for in the given model, right?

So, as an analyst, we suspect that there will be spatial dependence on housing prices, right? But we do not know the exact form of it, exact we do not even know the approximate form. So, we must use the data because my data exhibit it, then I must be able to pack it out from the data right; that is called the FGLS strategy. So, we will study this strategy in steps.

Step1: Run the Least Squares regression for estimating $P\left(\vec{S}_n\right)$ as a function of $R\left(\vec{S}_n\right),\ A\left(\vec{S}_n\right)\ \frac{w}{intercept}$.

$$\left[\hat{\beta}_{1,LS}\ \hat{\beta}_{2,LS}\ \hat{\beta}_{3,LS}\right] \equiv \hat{\beta}_{LS} = \left(X^1X\right)^{-1}X^1y$$

$$X\ NX3 = \left[1\ \vec{R}(.)\ \vec{A}(.)\right]\ ;\ \ y = \vec{P}(.)$$

Note that $\hat{\beta}_{1,LS},\ \hat{\beta}_{2,LS}\ \hat{\beta}_{3,LS}$ *are still unbiased.*

$$Obtain:\ \hat{\delta}\left(\vec{S}_n\right) = P\left(\vec{S}_n\right) - \left(\hat{\beta}_{1,LS} + \hat{\beta}_{2,LS}R\left(\vec{S}_n\right) + \hat{\beta}_{3,LS}A\left(\vec{S}_n\right)\right);\ \vec{S}_n \epsilon \left\{\vec{S}_1,\ \vec{S}_2,\ \ldots,\ \vec{S}_N\right\}$$

So, step 1 is to run the least squares regression for estimating the price P S_n as a function of R S_n and A S_n with intercept. So, we do not know omega so, we cannot write that we cannot evaluate beta hat GLS. So, all we are left with is beta hat LS that is you know we get to our beta hat LS which is x prime x inversed x prime y.

Remember, x now is 1 which is an N by 1 vector, we have R which is again an N by 1 vector; just like data in an Excel column, and then we have A which is again an N by 1 vector. So, x by itself is N by 3 and y is nothing, but the price vector which is N by 1 and so, y is also N by 1. So, beta hat LS is made up of 3 by N and N by 3. So, overall, 3 by 3 inversed which is 3 by 3 times 3 by N and N by 1.

So, I am looking at a 3 by 1, they are conformable and perfect. I have finally, a 3 by 1 vector, this is a 3 by 1. So, what I have been able to back out is beta 1 hat least squares, beta 2 hat least squares, and beta 3 hat least squares. Now, note that the beta hats, beta 1 hat least squares, beta 2 hat least squares, and beta 3 hat least squares are still unbiased.

We have established in this lecture, at the beginning of this lecture that you know even when there is spatial dependence, even when there is heteroskedasticity of the form of spatial dependence in data; it does not affect the unbiasedness property of a blue estimator, right? So, it is still unbiased. So, we are good to go using beta 1 hat beta 2 hat, and beta 3 hat as good guesses of true beta 1 beta 2, and beta 3, right?

So, using these I am going to propose that we obtain delta hat sorry S_n as P S_n minus beta 1 hat LS beta 2 hat LS R at location S_n minus beta 3 hat LS times A at location S_n. So, I can evaluate it at each location. So, we know that S_n is as all S_1 S_2 all the way to S capital N. So, I have a very good guess of also model residuals that is the difference between the predicted value, the predicted value, and the true value, right?

So, this is the true value, the true price which we observe in the data, and this here is my predicted value, right? This gives me now because beta hats are good to go, they are a good guess. So, is a predicted price a good guess because it is just a linear function, right? And so, if I deduct this predicted price which is a good guess from the actual price, I get a very good guess of the delta which is delta hat S_i.

This is all happening with the least squares estimator which we are we can do; we can work with you know quite efficiently without even worrying about the omega structure. So, we suspected there is omega. We do not know it so, we are doing what we can do best, that is we can run the least squares estimator.

The least squares estimator is falsely assuming that omega is sigma squared I_N, we know that as an analyst I am aware of that. But, with that in mind, I know I can still back out beta hats that are unbiased right? So, with these beta hats I will now back out of delta hat you know S_n.

Step2: Evaluate the stationarity of $\left\{\hat{\delta}\left(\vec{S}_n\right)\right\} \quad ; \quad \vec{S}_n \in \left\{\vec{S}_1, \vec{S}_2, \ldots, \vec{S}_N\right\}$

$\textit{because we will employ } \hat{\delta}'s \textit{ to evaluate spatial dependence through a variogram model.}$

Step3: Plot an experimental variogram for $\hat{\delta}\left(\vec{S}_n\right)'s$

$$i.e., \quad 2\gamma(h) = \frac{1}{|N(h)|}\sum_{i=1}^{N(h)}\left(\hat{\delta}\left(\vec{S}_n\right) - \hat{\delta}\left(\vec{S}_{n+h}\right)\right)^2$$

In the second step, I must evaluate the stationarity of delta hat S_n. Now, I want to evaluate this stationarity of delta hat S_n, because why? Because we are going, we will employ delta hats to evaluate spatial dependence through a variogram model. So, in the next step, step 3, I am going to say that plot an experimental variogram.

And, for a variogram, we are well aware that we need intrinsic stationarity. We are well aware of what is stationarity. We know that if we do not have stationarity, we cannot define a variogram. We also know that if we can get to the variogram, we will have backed out the data-driven spatial dependence structure that we really need to get to omega.

Now so, that is why you know before we jump on to step 3; we should take a step back and evaluate whether my delta hats are stationary. At the minimum, I should worry about you know can there be regimes of different mean values of these error terms given the data I have, can there be some kind of spatial trench that I should filter out? So, if the data you suspect to be non-stationary, you have to first create this filter, and construct this filter which you will deduct from the delta hat values.

And then finally, the residuals that you work with will then estimate a variogram. So, I am going to assume in step 2, that we have assessed that the delta hat values are stationary and we are going to move forward with that knowledge. Again, stationarity is a decision and it is not a hypothesis. Please go back and refer to the lectures where we studied stationarity.

We are going to plot an experimental variogram for delta hat $S_{n's}$, that is to say, I will calculate 2 gamma h which is going to be nothing, but 1 over the count of values separated by the count of pairs, location pairs separated by a lag of h. And, I am going to go from i equals 1 to N of h, this is the set that consists of all the pairs, location pairs separated by a value h lag h delta hat S_n minus delta hat S_n plus h right square.

We have seen this multiple times; we will again see it very soon because you are going to get to the hands-on exercise in a couple of lectures. So, we are coming to the variogram, experimental variogram, the variogram cloud and so on and so forth, right? If you want a little bit more familiarity, well we had plotted the variograms. We had seen the variogram plowed and basically, you are looking at something like this right?

So, you are looking at these variogram values at different levels of lag. So, you vary h you know, h itself is a variable, it is a spatial lag, right? So, with this h you can figure out some very important properties to say that to any value its nearer values are high, you know they are highly correlated or there is a high degree of covariance.

And, as we sort of leave that location the covariance dies out, eventually you know reducing to a large-scale variation in the data; where we have nothing to learn from in the no spatial correlation regime or range.

Step 4: Fit a variogram model (e.g., spherical, experimental, etc.)

And obtain Range: - the extent of spatial dependence

SILL: -large-scale variance

NUGGET: - micro-scale variance

$\Omega = [SILL ? ? \; SILL \quad \ddots \quad 8 ? 0 \quad \ddots \quad \ddots \quad \ddots \quad 0\ 0 \ldots \quad \ddots \quad \ddots$

Step 4 is going to be natural. I am going to fit a variogram model, right? So, we saw variogram models, you can refer to the lecture notes. We looked at the spherical variogram model, we looked at the exponential model, and several others. So, I am just going to leave it at that right? All of this is doable through software. So, variogram we know how to estimate a variogram theoretically.

Now, the hands-on exercise will be able to estimate the variogram model, right? So, when you do that, you obtain what is called the range, you obtain the sill and you obtain the nugget. You know typically we obtain these entities. The range is nothing, but the extent of spatial dependence in data.

Sill is the large-scale variance, and the nugget is the micro-scale variance or variation variance. These parameters now start to ring a bell of how omega will look like. So, omega which is the unknown, what we are after, this is the unknown variance-covariance that we are after is this omega. Now, the sill is going to be a large-scale variance.

So, the sill is something that should fill up my diagonals. The range is where the extent of spatial dependence. So, if for any h greater than 0, let us say there is no spatial dependence. Let us say, we have a model where it is a special case, but you know this the variogram shape is such that it just dies out after leaving it very quickly right? In such a scenario, you will have a case where off-diagonal elements will be 0 because that is what spatial dependence tells us.

If we know that after a certain point, starting from S1, after a certain point after the range, there is no spatial dependence. So, all the facts, and all the covariance factors will be 0. The non-zero ones can also be backed out from the variogram model. So, we are now starting to fill in this omega from what we have learned here in step 4.

Step5: Evaluate the data-driven variance matrix $\hat{\Omega}$

$\hat{\Omega} = \Big[C(0)\ C(\vec{S}_1, \vec{S}_2)\ C(\vec{S}_1, \vec{S}_3)\ C(\vec{S}_2, \vec{S}_1)\ C(0)\ C(\vec{S}_2, \vec{S}_3)\ \vdots\ \vdots\ C(\vec{S}_N, \vec{S}_1)\ \vdots\ \vdots\ C(\vec{S}_N, \vec{S}_2) \ddots \vdots \ \ldots \quad \ldots\ C$

Refer to Redundancy Matric, R in your notes from when we covered Kriging estimation.

$where,\ \gamma(h) = C(0) - C(h)$

$$or\ \gamma\left(\|\vec{S}_i - \vec{S}_j\|\right) = C(0) - C\left(\vec{S}_i, \vec{S}_j\right);$$

$i \epsilon \{1, 2, \ldots, N\}$

$j \epsilon \{1, 2, \ldots, N\}$

So, let us do that as step 5, step 5 just we are saying evaluate the data-driven variance-covariance matrix. And, we are going to call it omega hat because it is a data-driven device. Omega hat is now going to be C0, C0 is the sill the large-scale variation, it is going to be C0s C0s C0s and here I have C S_1 S_2, C S_1 S_3, keep going C S_1 S_n ok. Let me just ok.

Now, C S_2 S_1, C S_2 S_3 all the way C S_2 S_N, similarly I can fill these up C S N S_1, C S N S_2, keep going till C0. And, that is how my variance-covariance matrix will look like, where we know that the value gamma h that we have evaluated from the model in step 4 right, this value is equal to C0 minus Ch.

I can write this as gamma S_1 minus S_2, right? I can say S_i and S_j, just generalize this equals C0 minus C S_i and S_j nothing but the covariogram. The covariance between values at locations S_i and S_j, where we know that i goes from 1 to N and j similarly goes from 1 to N.

So, the variogram provides me with a data-driven analog of omega, right? If you want to sort of, get a refresher on where this matrix is, you can refer to the redundancy matrix in your notes from when we covered the Kriging estimator, the spatial interpolation estimator. We explicitly wrote this matrix; it is called a redundancy matrix and we even called it R.

So, you can go back and look at R, it is the same matrix that we have used for special interpolation. The beauty here is that now we are integrating the same idea that we use for spatial interpolation into spatial regression. And, we are getting away with the inefficiency of our least squares estimators by doing so.

Step6: Evaluate Feasible GLS estimation as

$$\hat{\beta}_{FGLS} = \left(X^1 \hat{\Omega}^{-1} X\right)^{-1} X^1 \hat{\Omega}^{-1} y$$

Unbiased and Efficient in the presence of spatial dependence in the model errors.

$$V\left(\hat{\beta}_{FGLS}\right) = \left(X^1 \hat{\Omega}^{-1} X\right)^{-1}$$

$$\left[\hat{\beta}_{1,FGLS}\ \hat{\beta}_{2,FGLS}\ \vdots\ \hat{\beta}_{k,FGLS}\right] \equiv \hat{\beta}_{FGLS}$$

So, as step 6, I am going to evaluate what is called the feasible, feasible GLS which is what was feasible given the data. We did not know the exact omega structure. Estimator as beta hat FGLS equals x prime omega hat inverse x inverse x prime omega hat inverse y.

And, the variance of beta hat FGLS is given as x in transpose omega hat inverse x, the whole thing inversed, right? So, these are called the feasible generalized least squares estimators. They are unbiased and they are efficient in the presence of spatial dependence in the model errors.

So, beta hat FGLS is unbiased and efficient in the presence of spatial dependence in the model errors. This is very interesting. As a last bit of this particular lecture, we can evaluate the size of this beta hat matrix. So, we have a K by N, K is the number of regresses, here it was K equals 3, for special case K equals 3, for our special example special case example, right?

I am going to say that this is N by N, this is N by 3, sorry N by K. So, I am looking at a K by K, here I have again a K by N, an N by N, and an N by 1. So, I am looking at a K by 1; so, overall, I am looking at a K by K, K by 1. So, I have a vector K by 1 which is nothing, but an estimator for beta 1 hat FGLS, beta 2 hat FGLS right, all the way till beta K hat FGLS. This is a very general form.

And you can show now, you should show that this is going to be a K by K matrix on your own. So, that is all I wanted to cover in this lecture. But I guess it is very clear that the variograms that we have spent so much time on and the idea of stationarity are completely integrable with the regression modelling, right? And, this is very powerful because we have studied variograms at quite a length.

So, now we are also empowered to estimate efficient regression models in the presence of spatial dependence in data. So, I hope you enjoyed this lecture and I will see you next time, where we will then relax another assumption of our classical regression model.

Thank you very much for your attention.