Homework 2: Using Spatial Lag, Spatial Error and Geographically Weighted Regression to Predict Median House Values in Philadelphia Block Groups

CPLN 671/MUSA 500

In the previous assignment, you were asked to use OLS regression to examine the relationship between median house values and several neighborhood characteristics, using Philadelphia data at the Census block group level. For the current assignment, you will use GeoDa and ArcGIS to run spatial lag, spatial error and geographically weighted regression to see whether these methods can account for the spatial autocorrelation that might remain in the OLS residuals.

Remember that this report needs to be written as your previous submission – with an introduction, methods/results, and discussion. Do not simply copy the questions and answer them. Below, you will find an outline which you're asked to follow when writing your report.

Data Description

The attribute table of the Philadelphia Census block group level dataset ***Regression Data.shp*** contains the following variables:

1) **AREAKEY:** Census Block Group ID
2) **MEDHVAL:** Median value of all owner occupied housing units
3) **PCBACHMORE:** Proportion of residents in Block Group with at least a bachelor's degree
4) **PCTVACANT:** Proportion of housing units that are vacant
5) **PCTSINGLES:** Percent of housing units that are detached single family houses
6) **NBELPOV100:** Number of households with incomes below 100% poverty level (i.e., number of households living in poverty)
7) **MEDHHINC:** Median household income

Note that the original Philadelphia block group dataset has 1816 observations. We clean the data by removing the following block groups:

1) Block groups where population < 40
2) Block groups where there are no housing units
3) Block groups where the median house value is lower than $10,000
4) One North Philadelphia block group which had a very high median house value (over $800,000) and a very low median household income (less than $8,000)

The final dataset which you are given contains 1720 block groups.

INSTRUCTIONS

SUGGESTION: READ THE ENTIRE SET OF INSTRUCTIONS BEFORE STARTING TO WORK ON THE ASSIGNMENT

IMPORTANT:

1. **When working in GeoDa, be sure to save your work often. Go to *File -> Save As*, and save everything as a new shapefile to do this. Saving as a new shapefile is the only way to ensure that the new variables that you create are saved in the table and will be retained there even once you close GeoDa. New fields are not saved automatically like in ArcGIS.**
2. **Students may do this assignment in R for extra credit.**
3. **Note that you have done some of the steps previously in R/ArcGIS. Here, I take you through these steps in GeoDa.**

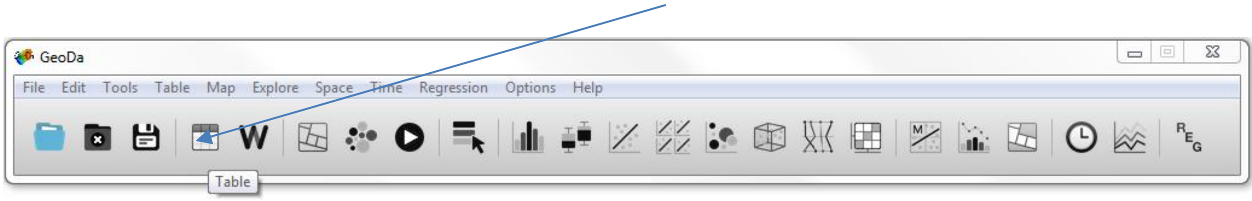
1) In GeoDa, open the file ***Regression Data.shp***.

   a. Recreate the variable **LNNBELPOV100** in GeoDa. (This is to give you a bit of practice with using GeoDa for new variable calculation.)

      i. Recall that you first need to add 1 to **NBELPOV100** prior to taking the natural log, because otherwise you may have a situation where you are taking logarithms of 0's in block groups where *NBELPOV100 = 0* (and as you may recall from algebra or the first lecture, logarithms of 0's are undefined). Unfortunately, new variable creation in GeoDa is a bit tedious, and needs to be done in two separate steps. That is, you cannot simply input the formula *LN(NBELPOV100 + 1)* into GeoDa. Instead, you need to first create a variable **NBELPOV100 + 1** and only then take the natural log of that sum.

      ii. Let's first create the variable **PLUS1**, defined as *NBELPOV100 + 1*

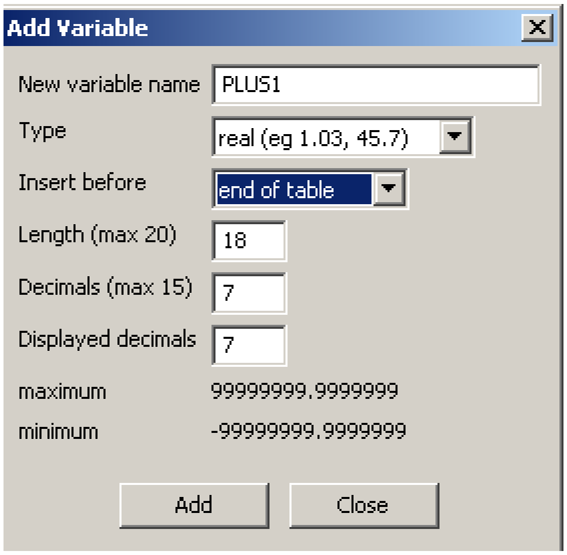
         1. To do this, first open the attribute table



         and right click anywhere on the table that opens up. Then select *Add Variable*.

         2. In the box that pops up, select the following settings. Basically, you're creating a real (continuous) variable called **PLUS1** that will

be placed at the end of the table (i.e., the last column in the table).



New variable name PLUS1

Type real (eg 1.03, 45.7)

Insert before end of table

Length (max 20) 18

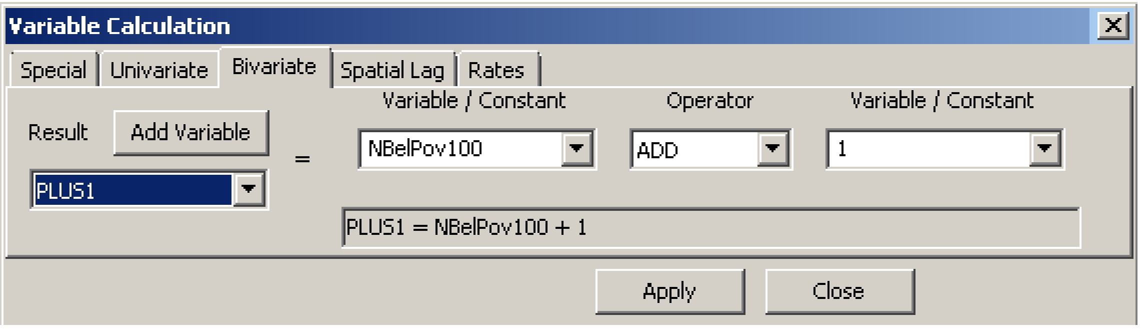
Decimals (max 15) 7

Displayed decimals 7

maximum 99999999.9999999

minimum -99999999.9999999

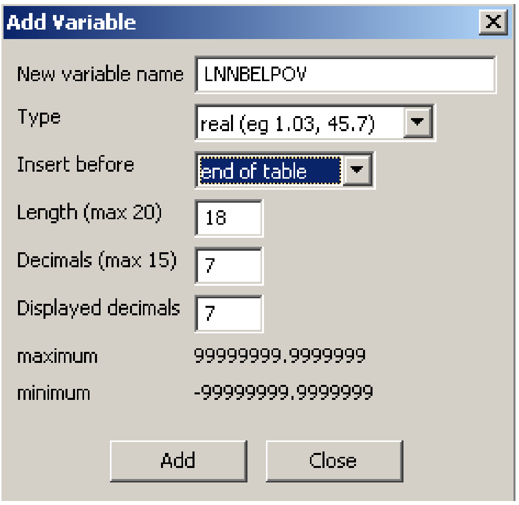
3. In the table, right click on **PLUS1** (which contains all 0's), select *Variable Calculation*. Then compute the variable as below:



Result		Variable / Constant	Operator	Variable / Constant
PLUS1	=	NBelPov100	ADD	1

PLUS1 = NBelPov100 + 1

iii. Using the steps outlined in 1.a.ii.2 above, create another new variable, called **LNNBELPOV**. Again, this variable will be defined as follows: *LNNBELPOV = LN(NBELPOV100+1) = LN(PLUS1)*.



New variable name: LNNBELPOV

Type: real (eg 1.03, 45.7)

Insert before: end of table

Length (max 20): 18

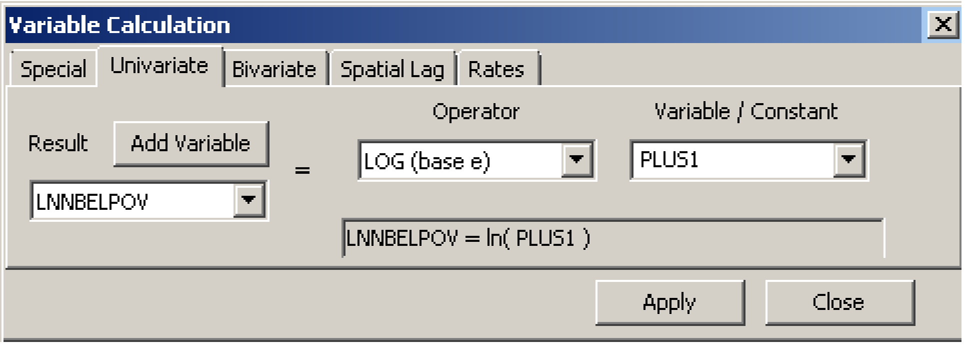
Decimals (max 15): 7

Displayed decimals: 7

maximum: 99999999.9999999

minimum: -99999999.9999999

iv. Your table should now contain the variable called **LNNBELPOV**. Again, it will be the field on the very right of the table. Right click on **LNNBELPOV** and select *Variable Calculation*. The variable should be calculated like this:



Result: LNNBELPOV

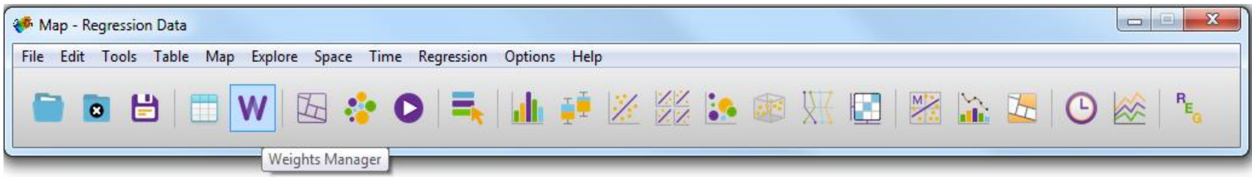
Operator: LOG (base e)

Variable / Constant: PLUS1

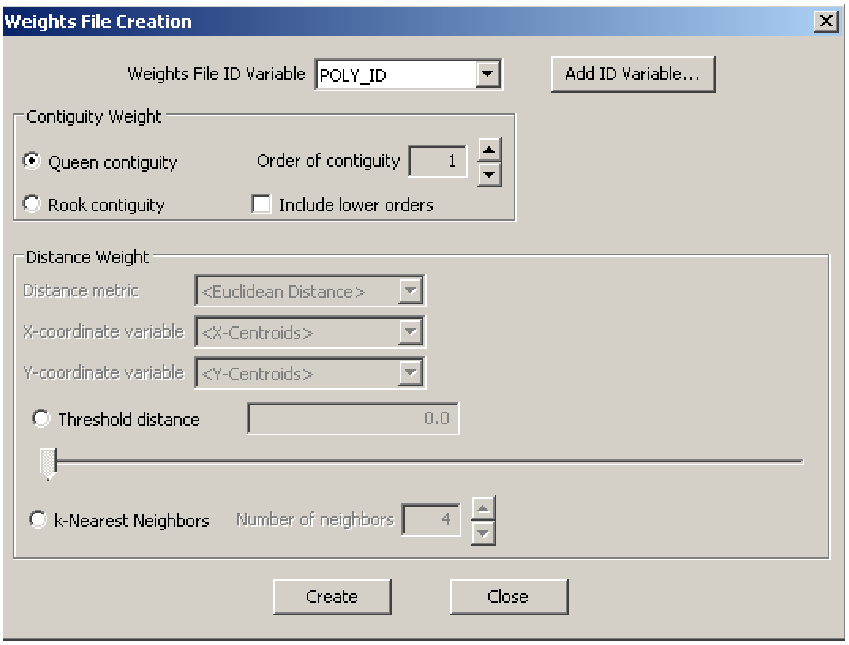
LNNBELPOV = ln(PLUS1)

v. Note: for your convenience, the log of the dependent variable, (log of (median house value + 1)), called **LNMEDHVAL**, is already in the dataset.

b. Create a Queen weight file (like this):



In the window that pops up, click Create and then select the options specified below:



Weights File Creation

Weights File ID Variable: POLY_ID

Contiguity Weight

● Queen contiguity — Order of contiguity: 1

○ Rook contiguity — □ Include lower orders

Distance Weight

Distance metric: <Euclidean Distance>

X-coordinate variable: <X-Centroids>

Y-coordinate variable: <Y-Centroids>

○ Threshold distance: 0.0

○ k-Nearest Neighbors — Number of neighbors: 4

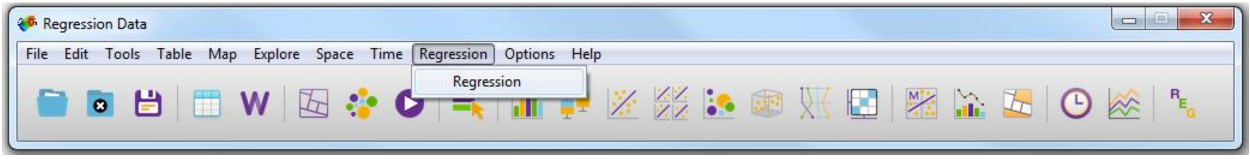
c. Now, we are ready for some analysis. Using the instructions on the slides, for the variable **LNMEDHVAL**, compute the global Moran's I using the Queen weight matrix created above. Then check to see whether the Moran's I value is significant (using 999 permutations). Take a screenshot of your results to present in your report (Moran's I value for the sample, histogram of Moran's I values for the permutations, and the p-value that you obtain will need to be included).

   i. Here and throughout, be sure to crop the screenshots so that only relevant parts are included in the report.

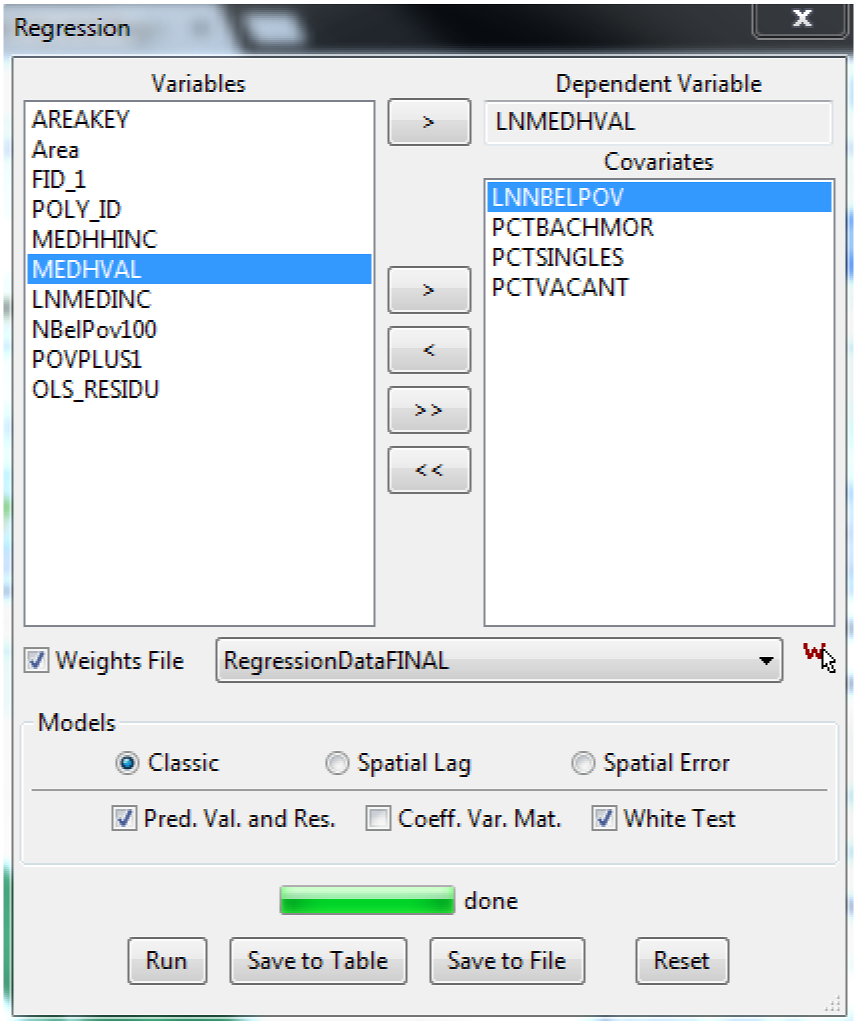
d. Run the *local* Moran's I (LISA) analysis using the Queen weight matrix. Take a screenshot of your results, which will need to be included in the final report.

e. Now, we're ready to run some regression analysis! First, let's rerun the OLS regression in GeoDa.

i. To do this, on the main menu, select *Regression*, as is done below.



   ii. We start out with OLS (Classic) regression – the very same regression we ran in R for the previous assignment. To do this, select the settings as below, (be sure to navigate to the queen weights file that you created in step 1.b), and click *Run*. Depending on the version of GeoDa that you use, the variable selection box may look slightly different from the one below, but all the steps should be the same.



Regression

Variables

AREAKEY
Area
FID_1
POLY_ID
MEDHHINC
MEDHVAL
LNMEDINC
NBelPov100
POVPLUS1
OLS_RESIDU

Dependent Variable: LNMEDHVAL

Covariates: LNNBELPOV, PCTBACHMOR, PCTSINGLES, PCTVACANT

☑ Weights File: RegressionDataFINAL

Models

● Classic ○ Spatial Lag ○ Spatial Error

☑ Pred. Val. and Res. □ Coeff. Var. Mat. ☑ White Test

done

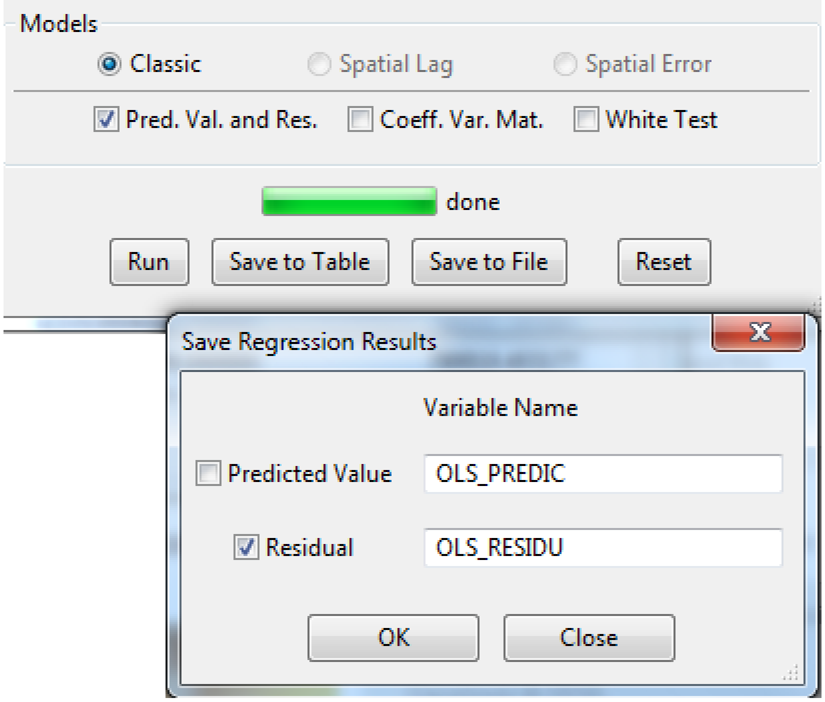
iii. Once the regression finishes running, you will get the output. It should be the output you obtained with R, but will contain a few additional diagnostics.

1. Copy the regression output into Word or a text file. You will be expected to present it in your report.

    a. For best visualization, present this output using the Courier New font, Size 8, single spaced.

iv. Go back to the dialog box above, and click *Save to Table*. Clicking *Save to Table* enables you to save OLS residuals $\hat{\varepsilon}$ and OLS predicted values $\hat{y}$ to the table.

1. So, in the dialog box that pops up (*Save Regression Results*), check *Residual*. This new variable will be given a name along the lines of **OLS_RESIDU**, as shown below. Click OK.



Models

● Classic ○ Spatial Lag ○ Spatial Error

☑ Pred. Val. and Res. □ Coeff. Var. Mat. □ White Test

done

Save Regression Results

Variable Name

□ Predicted Value OLS_PREDIC

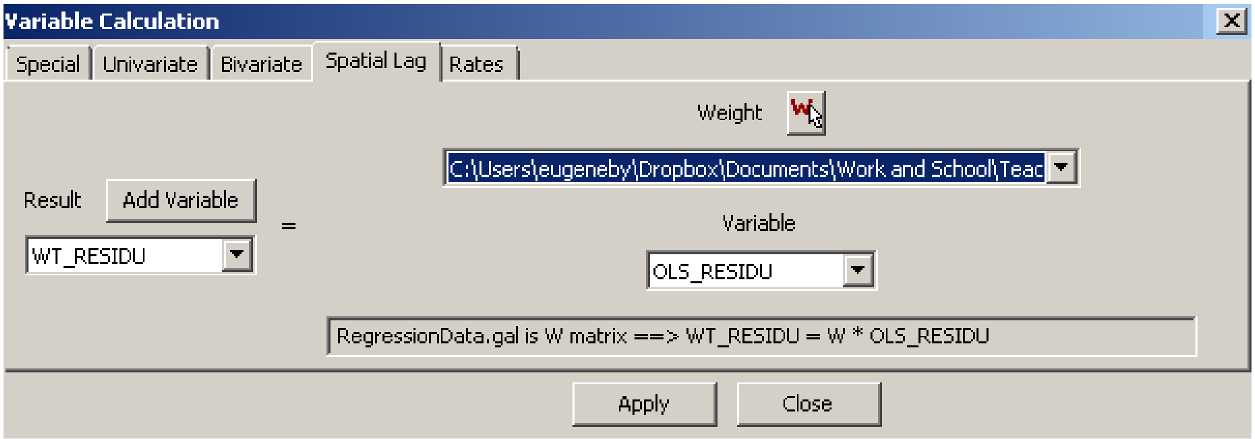
☑ Residual OLS_RESIDU

v. Take a look at the table: it now contains a new field called **OLS_RESIDU** with values of the OLS regression residuals at the very end of the table.

f. Now, let’s use GeoDa to create the weighted residuals. That is, for each block group, we will compute an average of the OLS residuals of the block group’s queen neighbors. For instance, if block group 1’s queen neighbors are block groups 3, 5 and 8, then the value of the weighted residual for block group 1 will be average of the residuals of block groups 3, 5 and 8.

i. In order to do that, first create a new variable called **WT_RESIDU**.

ii. Then, calculate the value of the variable as shown below. Again, for the weight, select the queen weight matrix that you created earlier.



Weight: C:\Users\eugeneby\Dropbox\Documents\Work and School\Teac

Result: WT_RESIDU

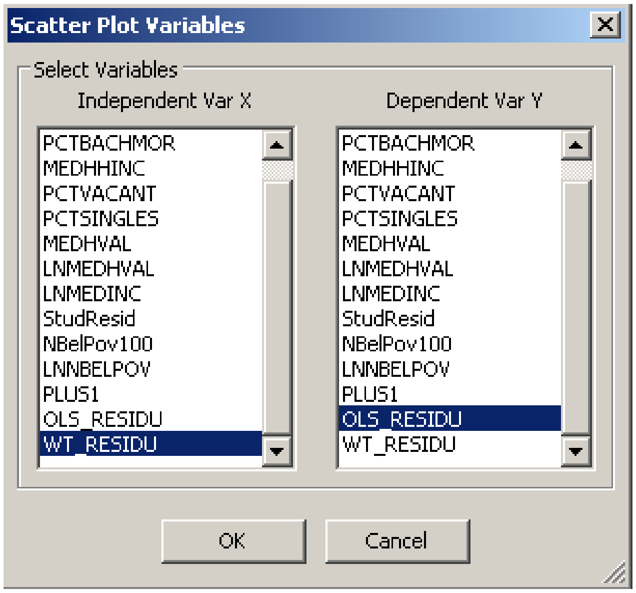
Variable: OLS_RESIDU

RegressionData.gal is W matrix ==> WT_RESIDU = W * OLS_RESIDU

g. Now, let's look a scatterplot that shows OLS residuals plotted against their queen neighbors. Of course, because one of the assumptions of OLS regression is independence of observations, if this assumption holds, there will be no relationship between OLS residuals and their neighbors. However, this assumption is likely to be violated here.

i. On the main menu, go to *Explore -> Scatter Plot*.

ii. Select **WT_RESIDU** as the independent variable and **OLS_RESIDU** as the dependent variable (as shown below), and click OK.



Independent Var X: WT_RESIDU

Dependent Var Y: OLS_RESIDU

iii. Right-click on the scatterplot that pops up, and check *Display Statistics*. Some statistics will be displayed at the bottom of the plot, including ***Slope b*** (and corresponding significance results) – this is the coefficient of **WT_RESIDU** when you regress **OLS_RESIDU** on **WT_RESIDU**.

1. Note that this this is the same thing as running a simple regression with **OLS_RESIDU** as the dependent variable and **WT_RESIDU** as the predictor. For instance, the Beta coefficient of **WT_RESIDU** in that regression will be the same as ***Slope b***.

    a. From the slides, you should see that this beta coefficient is what is commonly known as '*rho*' (*ρ*) – and this *rho* is just another measure of spatial autocorrelation (see slides ~10 – ~16). And in GeoDa, it's referred to as *lambda* (*λ*).

iv. Take a screenshot of that scatterplot and the statistics that appear at the bottom of it to present in your report.

h. Using the same steps as in (1.c) above, Look at the Moran's I of the OLS regression residuals to see whether there is spatial autocorrelation.

i. Again, use the queen matrix that you calculated here.

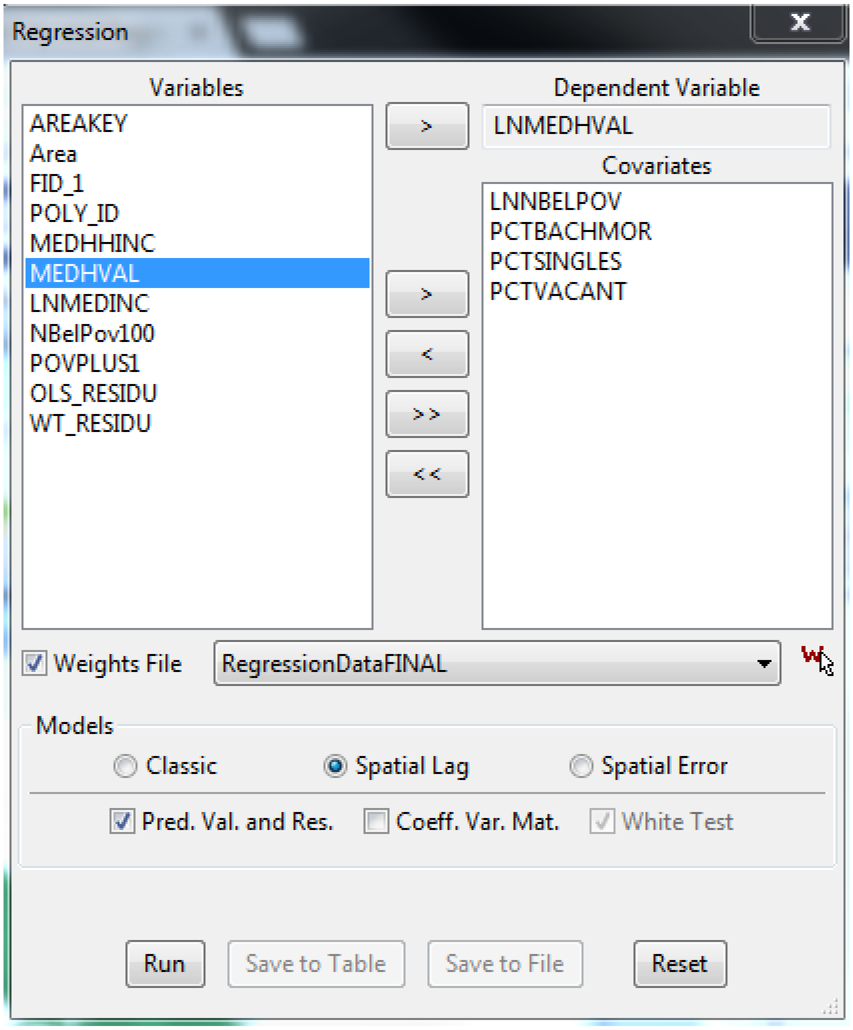
ii. Test whether the Moran's I value is significant by running 999 permutations.

iii. Take a screenshot of the Moran's I results (both the Moran scatterplot and the significance test). You will be expected to present this in your report.

i. Now, let's run the **spatial lag** regression model in GeoDa.

i. On the main menu, go to *Regression*.

ii. In the regression dialog box that pops up, select the following settings:



Regression

Variables

AREAKEY
Area
FID_1
POLY_ID
MEDHHINC
MEDHVAL
LNMEDINC
NBelPov100
POVPLUS1
OLS_RESIDU
WT_RESIDU

Dependent Variable: LNMEDHVAL

Covariates

LNNBELPOV
PCTBACHMOR
PCTSINGLES
PCTVACANT

☑ Weights File: RegressionDataFINAL

Models

○ Classic ● Spatial Lag ○ Spatial Error

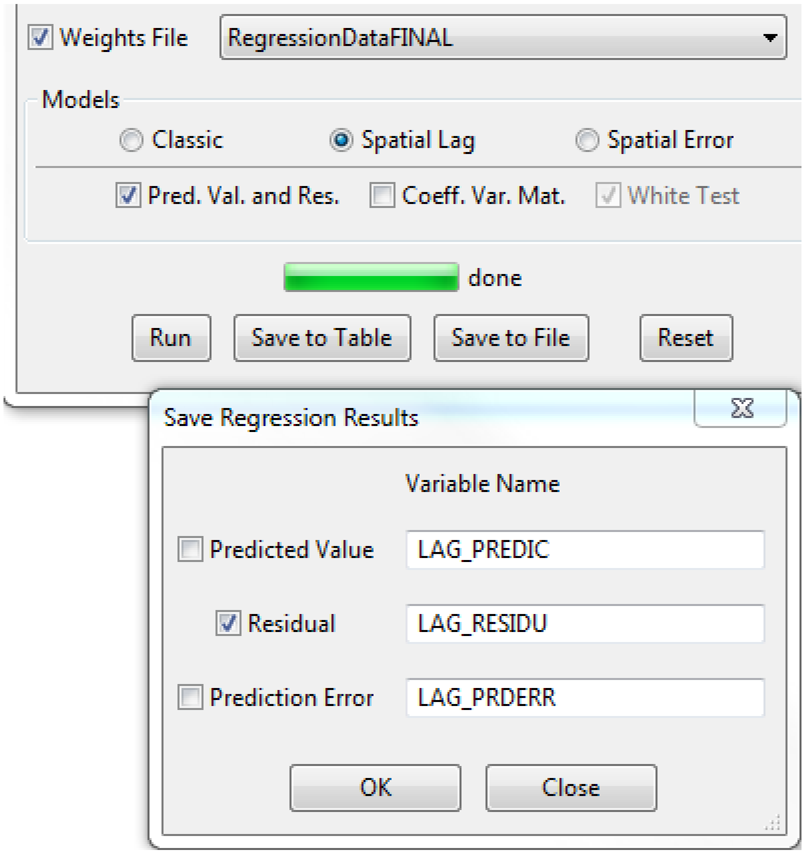
☑ Pred. Val. and Res. □ Coeff. Var. Mat. ☑ White Test

iii. Above, use the same queen weights file that we have created earlier.

iv. Once you click Run, you will get the output. Copy the output into Word or a text file. You will be expected to present it in your report.

1. For best visualization, present this output using the Courier New font, Size 8, single spaced.

v. After the regression is done running, you will also be able to go back to the regression dialog box, and click on *Save to Table*, as shown below. You will be asked to save the spatial lag regression residuals as you did for OLS residuals.



☑ Weights File RegressionDataFINAL

Models

○ Classic ● Spatial Lag ○ Spatial Error

☑ Pred. Val. and Res. □ Coeff. Var. Mat. ☑ White Test

done

Save Regression Results

	Variable Name
□ Predicted Value	LAG_PREDIC
☑ Residual	LAG_RESIDU
□ Prediction Error	LAG_PRDERR

vi. Now, using the same steps as in (1.h) above, look at the Moran's I value of the Spatial Lag (SL) residuals, and run 999 permutations to see whether the spatial autocorrelation in the SL residuals is statistically significant. Once again, be sure to take a screenshot of the Moran's I results (both the Moran scatterplot and the significance test). You will be expected to present this in your report.

j. Now, repeat steps 1.i.i – 1.i.vi for **spatial error** regression. That is, keep everything the same except choose **spatial error** instead of **spatial lag**.

k. Before proceeding, make sure that you have the following outputs saved somewhere:

i. Global and local Moran's I results for the variable **LNMEDHVAL**

ii. OLS Regression Results

iii. Spatial Lag Regression Results

iv. Spatial Error Regression Results

v. A scatterplot of **OLS_RESIDU** and **WT_RESIDU**, with statistics displayed

vi. Moran's I scatterplot (and results of 999 permutations) for OLS Regression

vii. Moran's I scatterplot (and results of 999 permutations) for Spatial Lag Regression

viii. Moran's I scatterplot (and results of 999 permutations) for Spatial Error Regression

l. Be sure to save your file (go to *File -> Save As*, and save as *RegressionFinal.shp*, or something of the sort). Now, you may close GeoDa.

2) Now, open the file *RegressionFinal.shp* in ArcGIS. You will use ArcGIS to run Geographically Weighted Regression.

a. Navigate to: *ArcToolbox -> Spatial Statistics Tools -> Modeling Spatial Relationships -> Geographically Weighted Regression*

b. Select the following settings in the dialog box:

i. Under *Input Features*, select *RegressionFinal.shp*

ii. Under *Dependent Variable*, select **LNMEDHVAL**

iii. Under *Explanatory Variables*, select **LNNBELPOV**, **PCTBACHMOR**, **PCTSINGLES**, **PCTVACANT** (same four variables as you included in the OLS, Spatial Lag and Spatial Error regression models).

iv. Under *Output Feature Class*, select a shapefile where you want your results to be saved.

v. Under *Kernel Type*, select *Adaptive*

vi. Under *Bandwidth Method*, select *AICc*

vii. Click *OK*.

c. In your report, you will need to present the following output:

i. The supplemental table (output table with suffix *_supp*)

1. Open the table and take a screenshot of it.

ii. Present a choropeth map of the local R-squared values in the slides.

d. Open the shapefile from 2.b.iv in GeoDa (this is described in the last 6-7 slides). As is done in step 1.h, look at the Moran's I of the GWR residuals to see whether there is spatial autocorrelation (use *un*standardized residuals when doing the test – i.e., use the *Residual* field and not the *StdResid* Field).

    i. You may need to recalculate the weight matrix (again, use the queen matrix).

    ii. Test whether the Moran's I value is significant by running 999 permutations.

    iii. Take a screenshot of the Moran's I results (both the Moran scatterplot and the significance test). You will be expected to present this in your report.

e. Follow instructions on the slides to obtain local regression results. Specifically, present maps of the ratio of the beta coefficients and the standard error estimates.

    i. Use dark red when the ratio is < - 2, pink when the ratio is between 0 and -2, light blue when the ratio is between 0 and 2, and dark blue when the ratio is > 2.

Now, you are finally ready to start writing your report!

REPORT OUTLINE

A successful report will address *all* the points presented in this outline. You are strongly encouraged to use the outline as a backbone for your report.

The outline here is structured as an outline for a journal article. That is, in the Methods section, only talk about the techniques that you use, present the formulas, etc. Do not present any results in the methods section. In the Results section, actually present the output from R and ArcGIS, any figures, etc, and describe your output.

1) **Introduction (~2 paragraphs)** *Section Title*
    a) State the problem and the setting of the analysis (Philadelphia).
    b) Indicate that in the previous report, you carried out OLS regression to examine the relationship between your dependent variable and predictors (state what the DV and predictors are).
    c) State that OLS analysis is often inappropriate when dealing with datasets that have a spatial component
    d) Mention that the purpose of this report is to use spatial lag, spatial error and geographically weighted regression to see whether these methods perform better than OLS.

2) **Methods (~5 pages)** *Section Title*
    a) **A Description of the Concept of Spatial Autocorrelation** *Subsection Title*
        i. Mention the 1st Law of Geography
        ii. Talk about Moran's I
            1. Present and explain formula for Moran's I
        iii. Mention and explain the weight matrix that you're using.
            1. Indicate that throughout this report, you will be using this weight matrix.
            2. Specify why statisticians sometimes like to use more than one spatial weight matrix in their analyses.
        iv. In your own words, talk about how you test whether the spatial autocorrelation (Moran's I) is significant. State what hypotheses you're testing (present the null and alternative hypotheses) and describe the random permutation process.
        v. Briefly describe the concept of local spatial autocorrelation, without going into any of the mathematical detail.

    b) **A Review of OLS Regression and Assumptions** *Subsection Title*
        i. Begin by giving a *brief* (3-5 sentence) overview of OLS regression. Specifically, list the assumptions of OLS
            1. Refer the reader to your HW 1 for more information on OLS.

        a. [FYI: Referring the reader to a previous HW assignment is often done in ESE 502 in order to avoid rewriting a lot of the same things over again].

  ii. State that when the data has a spatial component, the assumption that your errors are random/independent often doesn't hold

    1. Indicate that you can test the assumption in (ii) above by examining the spatial autocorrelation of the residuals using Moran's I.
    2. Indicate that another way to test OLS residuals for spatial autocorrelation is to regress them on nearby residuals (here, these nearby residuals are residuals at neighboring block groups, as defined by the Queen matrix).
        a. Mention *rho* (ρ) and how it is calculated. [It's that term that's known as *lambda* (*λ*) in GeoDa, and is referred to as *Slope b* in the statistics at the bottom of the scatterplot of **OLS_RESIDU** and **WT_RESIDU**]

  iii. State that GeoDa, the tool that you're using to run your OLS regression, also has a way of testing other regression assumptions.

    1. The first is the assumption of *homoscedasticity*, which is tied to the assumption of independence of errors.
        a. State which test(s) is/are used to examine data for heteroscedasticity in GeoDa, and state the null and alternative hypotheses.
    2. Another assumption is that of *normality of errors*.
        a. State which test is used to test for normality of errors in GeoDa, and state the null and alternative hypotheses.

c) **Spatial Lag and Spatial Error Regression** *Subsection Title*

  i. State that you will be using GeoDa for running spatial lag and spatial error regressions.

  ii. Describe the method of spatial lag regression in several sentences.

    1. Present the model equation for the spatial lag model.
        a. Instead of writing X1…X4, write the names of the actual predictors that you're using in this assignment (e.g., **PCTVACANT**)
        b. Explain what each term is (the β coefficients, ρ, ε, etc)

  iii. Describe the method of spatial error regression in several sentences.

    1. Present the model equation for the spatial error model.
        a. Instead of writing X1…X4, write the names of the actual predictors that you're using in this assignment (e.g., **PCTVACANT**)
        b. Explain what each term is (the β coefficients, λ, ε, u, etc)

iv. Indicate that the assumptions that are needed for OLS are still needed for both spatial lag and spatial error regression models (except that of spatial independence of observations).

v. State the goal of spatial lag and spatial error regression (i.e., what you hope will happen with regression residuals as a result of using these methods).

vi. Mention that you will compare the results of spatial lag regression with OLS and the results of spatial error regression with OLS, and will decide whether the spatial models perform better than OLS based a number of criteria.

1. These criteria include
   a. Akaike Information Criterion/Schwarz Criterion;
   b. Log Likelihood;
   c. Likelihood Ratio Test
2. Be sure to describe what each of the above criteria is, and how you decide which model is better based on this criterion (state any null/alternative hypotheses, if applicable).
3. State that another way of comparing OLS results with spatial lag and spatial error results is by looking at the Moran's I of regression residuals.
   a. Indicate how you would decide which model is better based on this criterion.

d) **Geographically Weighted Regression** *Subsection Title*

i. State that you will do your GWR analyses in ArcGIS.

ii. Introduce GWR by talking about the concepts of Simpson's paradox and local regression.

iii. Present the GWR equations and explain them in your own words

iv. Talk about how local regression is run

v. Discuss the concept of bandwidth, and talk about adaptive vs. fixed bandwidth.

1. State that here, you will be using adaptive bandwidth
   a. Explain why adaptive bandwidth is more appropriate in this problem than the fixed bandwidth

vi. Mention that the OLS assumptions still hold in GWR.

1. When mentioning multicollinearity, talk about the Condition Number, and the issues of multicollinearity/clustering in GWR.

vii. Indicate why p-values are not part of the GWR output.

3) **Results (~3-5 pages, excluding maps, figures & tables)** *Section Title*

a) **Spatial Autocorrelation** *Subsection Title*

i. Present and describe the global Moran's I value and the random permutations test results.

1. Is **LNMEDHVAL** significantly spatially autocorrelated?

ii. For Local Moran’s I results, present the Significance Map and Cluster Map obtained by running the Local Morans’ I.

   1. Discuss the results: what are the not significant, high-high, high-low, low-high and low-low areas on the Cluster Map? Where in the city are these areas?

b) **A Review of OLS Regression and Assumptions: Results** *Subsection Title*

   i. Present the OLS output from GeoDa (call this Table 1)

      1. Give a brief 2 sentence overview of the OLS results (feel free to paste this from your description in HW 1). That is, simply indicate which predictors are significant and what % of variance in **LNMEDHVAL** has been explained by the model.
      2. Comment on the results of the tests on heteroscedasticity
         a. Are the results from the 3 tests consistent with each other?
         b. Do they indicate a problem with heteroscedasticity?
      3. Comment on the results of the test on normality of errors
         a. Do test results indicate a problem with normality?

   ii. Present the scatterplot of **OLS_RESIDU** by **WT_RESIDU** and describe the results.

      1. Are the results (based on the value and significance level of ρ – that’s referred to as ***Slope b*** in the results) indicative of significant spatial autocorrelation?

   iii. Present the Moran’s I scatterplot and results from the 999 permutations for OLS regression residuals.

      1. Are you seeing significant spatial autocorrelation in your OLS residuals, and is this problematic?

c) **Spatial Lag and Spatial Error Regression Results** *Subsection Title*

   i. Present results of Spatial Lag regression (call this Table 2)

      1. Talk about the **W_LNMEDHVAL** term in the spatial lag regression output. State whether it is significant, and how the results can be interpreted.
      2. Are the remaining terms (i.e., the predictors **LNNBELPOV**, **PCTBACHMOR**, **PCTSINGLES**, and **PCTVACANT**) in the model significant?
         a. Compare these results to OLS results.
      3. State whether, based on the Breusch-Pagan test, the spatial lag regression residuals are still heteroscedastic.
      4. Compare the Spatial Lag regression and OLS regression models based on the Akaike Information Criterion/Schwarz Criterion, the Log Likelihood, and the Likelihood Ratio Test.

5. Present the Moran’s I scatterplot of spatial lag regression residuals. Does there seem to be less spatial autocorrelation in these residuals than in OLS residuals?
6. Overall, which model is doing better based on all of these criteria?

ii. Present results of Spatial Error regression (call this Table 3)

1. Talk about the **LAMBDA** term in the spatial lag regression output. State whether it is significant, and how the results can be interpreted.
2. Are the remaining terms (i.e., the predictors **LNNBELPOV**, **PCTBACHMOR**, **PCTSINGLES**, and **PCTVACANT**) in the model significant?
   a. Compare these results to OLS results.
3. State whether, based on the Breusch-Pagan test, the spatial lag regression residuals are still heteroscedastic?
4. Compare the Spatial Error regression and OLS regression based on the Akaike Information Criterion/Schwarz Criterion, the Log Likelihood, and the Likelihood Ratio Test.
5. Present the Moran’s I scatterplot of spatial error regression residuals. Does there seem to be less spatial autocorrelation in these residuals than in OLS residuals?
6. Overall, which model is doing better based on all of these criteria?

iii. Compare the Spatial Lag and Spatial Error results with each other

1. Recall that you should not be using the likelihood-ratio test for this because the models are not nested (i.e., neither method is a special subtype of each other). However, it is OK to compare the two non-nested models, such as spatial lag and spatial error, based on Akaike Information Criterion and the Schwarz Information Criterion.
   a. Which model has better (lower) Akaike Information Criterion and Schwarz Information Criterion values?

d) **Geographically Weighted Regression Results** *Subsection Title*

i. Present GWR results from the _*supp* table (call this Table 4)

1. Compare the (overall) R-squared of the GWR regression with the R-squared of the OLS regression. State which regression method seems to be doing a better job of explaining the variance in the dependent variable.
2. Compare the Akaike Information Criteria of GWR with those of OLS, Spatial Lag and Spatial Error models. Which model seems to be doing a better job based on that (remember, the lower the Akaike Information Criterion, the better the fit).

ii. Present the Moran’s I scatterplot of GWR residuals. Does there seem to be less spatial autocorrelation in these residuals than in OLS residuals? What about the Spatial Lag and Spatial Error Residuals.

iii. Be sure to discuss local regression results, as is done on the slides. Are there locations in the city where the relationships between each of the predictors and the dependent variable possibly significant?

iv. Present and discuss a choropleth map of local R-squared results.

4) <u>Discussion (~1 page)</u> *<u>Section Title</u>*

a) In a couple sentences, recap what you did in the paper and your findings. Discuss what conclusions you can draw, and which of the four regression methods (OLS, Spatial Lag, Spatial Error, GWR) was the best, based on the results.

b) Give a brief description of the limitations (i.e., which assumptions were not met).