Spatial Regression

Lecture #23 | GEOG 510
GIS & Spatial Analysis in Public Health

Varun Goel

Outline

- Spatial regression
  - Spatial Lag and Error models
    - A bit more…
    - Review of pattern vs process
    - Example
  - Geographically Weighted Regression

Comparing Variables

- Correlation
  - Tells us about relationship between two variables
    - Direction and strength
- Regression
  - Tells us about relationship among variables
    - Direction, strength, and magnitude

Residual Autocorrelation

- Regression residuals should not be spatially autocorrelated
  - Residuals should be randomly distributed (independent)
  - To test, join residuals back to spatial data features
    - Run a Moran's *I* analysis using the residual as the variable
      - We hope to get "null" results (high *p* value), meaning a random distribution!

Residual Autocorrelation

- Potential fixes…
  - Find missing independent variable
  - Spatial regression approaches
    - Spatial Lag Model
    - Spatial Error Model
  - Spatial filtering
    - Removes spatial autocorrelation from input variables (prior to regression)

Spatial Dependence

- Nuisance
  - Something to get rid of
  - Involves model residuals only
    - Spatial Error Model
- Substantive
  - Value of interest
  - Involves neighbors of Y
    - Spatial Lag Model

Spatial Regression

- Spatial regression approaches
  - Spatial Lag Models
    - Similar, in thought, to a 2nd order spatial process (spillover)
  - Spatial Error Models
    - Similar, in thought, to a 1st order spatial process (underlying driver)

Pattern and Process

- Simple difference
  - A spatial pattern is generally the result of some spatial process
    - We map/observe patterns
    - We aim to understand processes
- Process is the "real life" action or interaction happening
  - Requires us to think about how real life is expressed via data

Pattern and Process

- Thinking about "why" we observe geographic variation in our data…
  - e.g., why is disease prevalence or incidence higher here than there?
    - The geographic variation that we map is the pattern
    - The real life actions/interactions are the process that result in the geographic variation

Spatial Processes

- First order effects
  - Observed spatial variation is due to an extrinsic (external) factor (or factors)
    - For example
      - Disease cases and population counts
      - Health outcomes and poverty
    - Often, we use correlation/regression based methods to capture first order effects

Spatial Processes

- Second order effects
  - Spatial variation is due to interaction or intrinsic factors
    - For example
      - Directly transmittable disease (cases spawn more cases)
      - Some retail behavior (competition), Dispersion

Spatial Regression

- Linking process with models…
  - 2nd order spatial process
    - Geographic variation in outcome variable driven by intrinsic (internal) factors
  - Understanding or characterizing spatial dependence (in the outcome variable), because it helps to explain process
    - Spatial lag model

$$Y = \beta_0 + \beta X + \rho W y + \epsilon$$

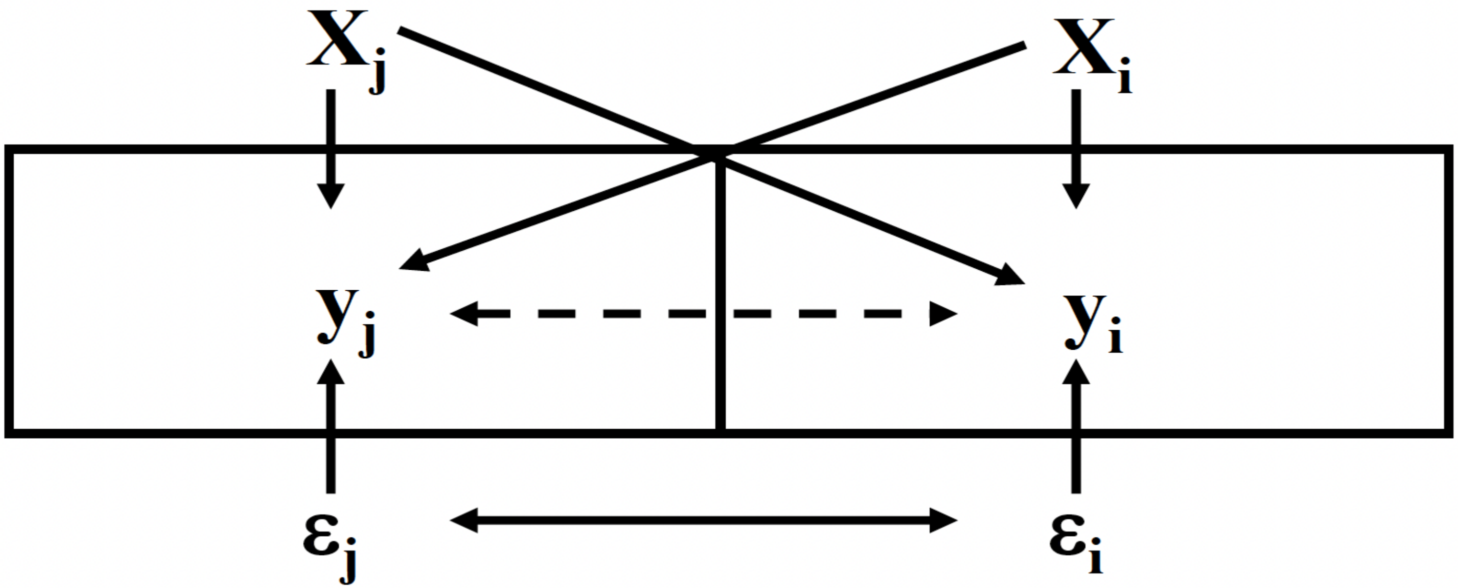
Spatial Lag Model

$$Y = \beta_0 + \beta X + \rho W y + \epsilon$$

- Considers potential autocorrelation in the Y variable
  - Neighbor weight (*W*)
    - *y* is the neighbor's value of *Y*
  - Estimates, ρ, which is very much similar to a regression coefficient

Spatial Lag Model

OLS estimates are ***biased***, and thus inferences based on an OLS model will be incorrect



Spatial Lag Model

$$Y = \beta_0 + \beta X + \rho W y + \epsilon$$

- Maximum Likelihood Estimator
  - Returns similar results to OLS regression
    - Coefficients, SEs, p-values
    - Some software provides a pseudo R^2

Spatial Regression

- Linking process with models…
  - 1st order spatial process
    - Geographic variation in outcome variable driven by extrinsic (external) factors
  - Removing spatial dependence in the residuals, because process not fully captured (by factors)
    - Spatial error model

$$Y = \beta_0 + \beta X + e \quad \text{where,} \quad e = \lambda W \mu + \epsilon$$

Spatial Error Model

$$Y = \beta_0 + \beta X + e \quad \text{where,}$$

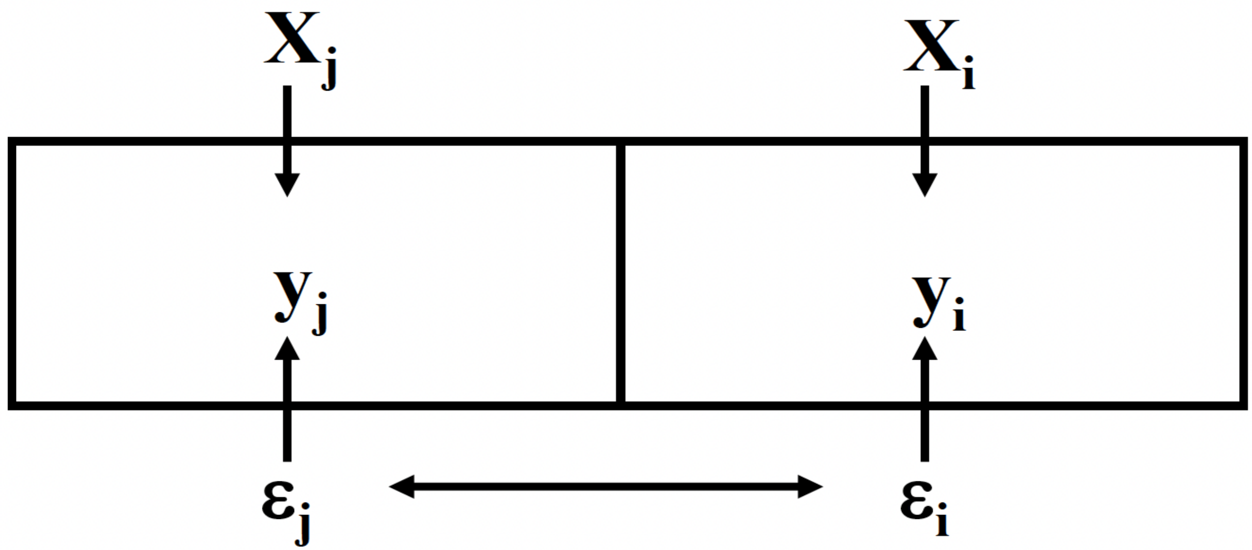
$$e = \lambda W \mu + \epsilon$$

- Considers potential autocorrelation in the error of the model
  - e Ṳ autocorrelated residuals from non-spatial model
  - Neighbor weight (*W*)
    - μ is the “unobserved variable”
  - Estimates, λ, which is very much similar to a regression coefficient

Spatial Error Model

Dependence amongst the errors

OLS estimates become ***inefficient***



Spatial Error Model

$$Y = \beta_0 + \beta X + e \quad \text{where,}$$

$$e = \lambda W \mu + \epsilon$$

- Maximum Likelihood Estimator
  - Returns similar results to OLS regression
    - Coefficients, SEs, p-values
    - Some software provides a pseudo R^2

Spatial Regression

- If residuals from OLS are correlated,
  - How do I know whether to use a Lag or Error model?
    - Theory!
    - Diagnostics (in GeoDa)

Spatial Regression

- Diagnostics
  - Lagrange Multiplier (LM) tests
    - LM error and LM lag
    - Robust LM error and LM lag
  - Start with non-robust tests
    - If one is significant (low *p* value) and the other not, then use one that is significant
    - If both significant, move to Robust versions (interpret similarly to above)
      - If both highly significant, go with lowest *p*

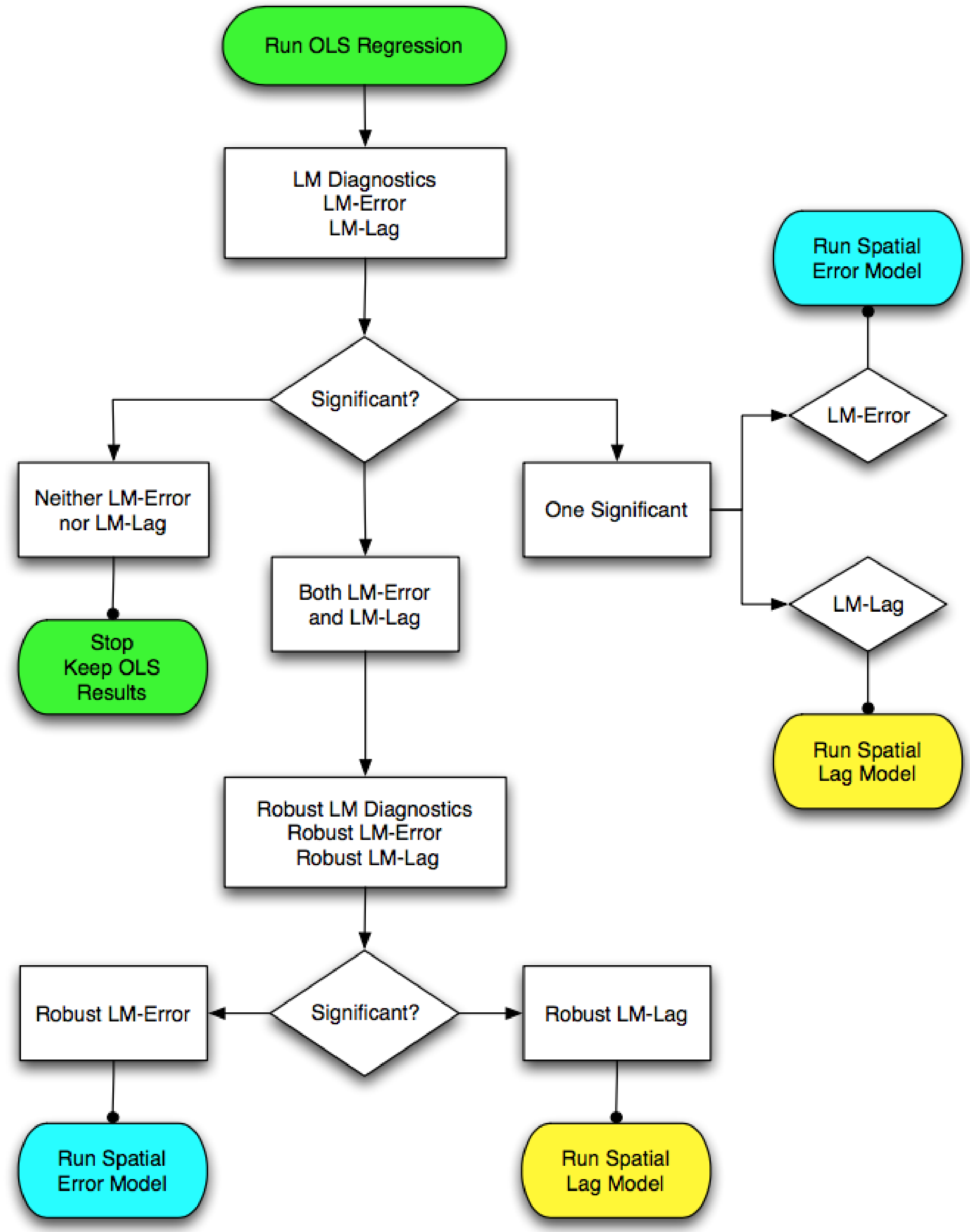


Figure 23.24: Spatial regression decision process.

Spatial Regression

- Neighbor definition and weights
  - May affect the regression model results
    - Similar to autocorrelation tests
    - Good to conduct sensitivity analysis

Spatial Regression

- In practice, using real data…
  - Almost always messy and rarely a straightforward procedure
    - Do your best to ensure regression assumptions are met
    - Interpret with caution

Spatial Processes

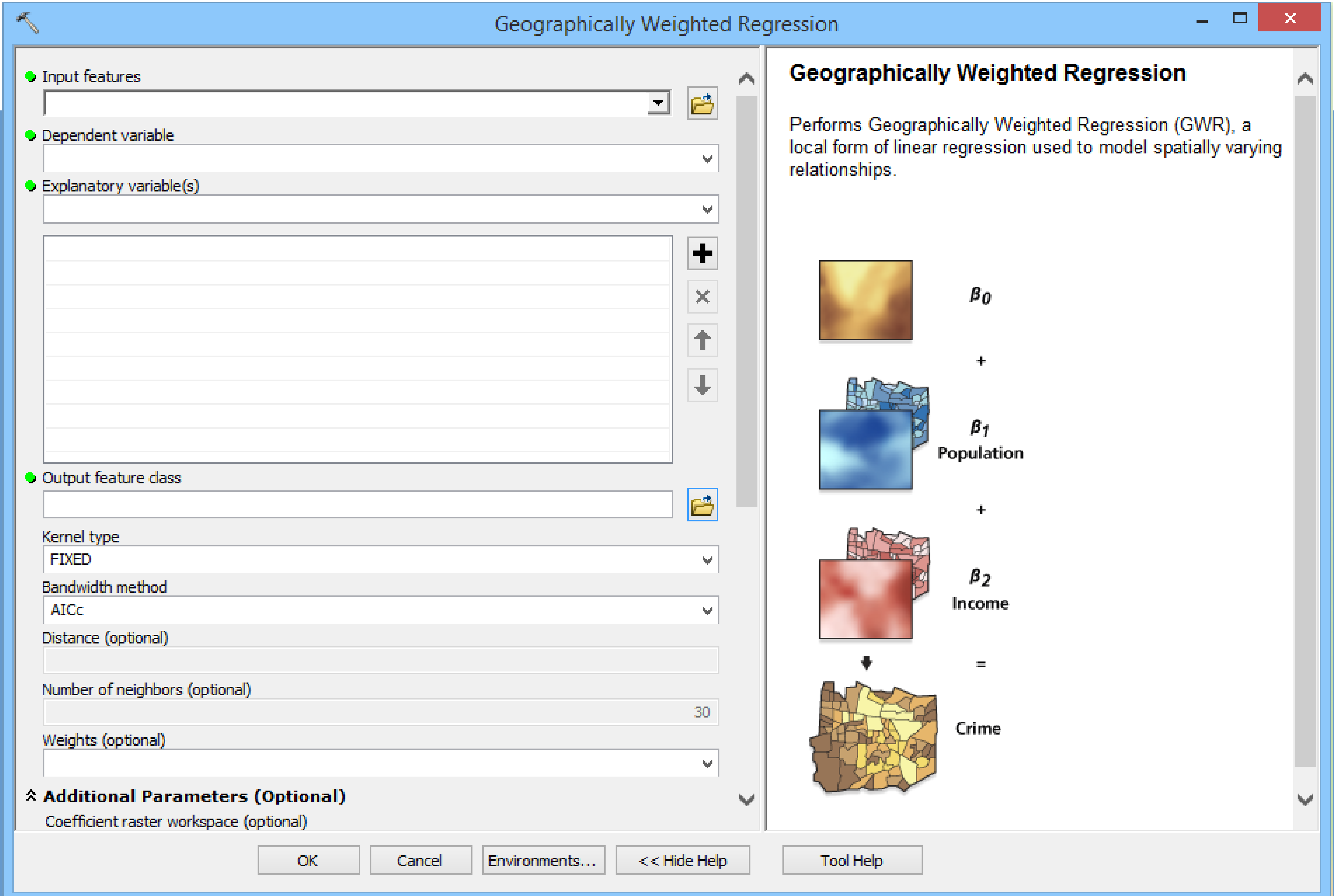
- Relationships among factors often assumed to be static across space
  - e.g., effects of precipitation and temperature on mosquito density
  - Not always the case
    - Relationships themselves can vary across space
      - Spatially varying process

GWR

- Geographic Weighted Regression
- Global regression
  - Assumes that relationships (βs) are stationary across space
    - Invariant from place to place (global)
- GW Regression
  - Assumes that relationships (βs) are <u>non</u>stationary (across space)
    - Vary from place to place (local)

GWR

- Basically, numerous “local” regressions
  - Provides β for every observation
    - Variation in the effect over space
  - Provides R^2 for every observation
- Take care in using for inference
  - Interesting results… but trustworthy?


Geographically Weighted Regression
Input features
Dependent variable
Explanatory variable(s)
Output feature class
Kernel type
FIXED
Bandwidth method
AICc
Distance (optional)
Number of neighbors (optional)
30
Weights (optional)
Additional Parameters (Optional)
Coefficient raster workspace (optional)
OK
Cancel
Environments...
<< Hide Help
Tool Help
Geographically Weighted Regression
Performs Geographically Weighted Regression (GWR), a local form of linear regression used to model spatially varying relationships.
β0
+
β1
Population
+
β2
Income
=
Crime

Regression

- What's available?
  - GeoDa
    - OLS, Spatial Lag, Spatial Error
  - GeoDaSpace
    - OLS, Spatial Lag, Spatial Error, SpLE
      - With standard error corrections (for heteroscedasticity)
  - ArcGIS
    - OLS, GWR

Keywords

- Residual autocorrelation
- Spatial Lag
- Spatial Error
- First and Second order processes
- LM tests
- Spatially varying process
- GWR