Spatial Analysis and Modeling

05 - Spatial Regression

Instructor: Rossano Schifanella

@UNITO

Recap Linear Regression

Linear Regression - Notation

linear relationship between a dependent variable y_i (at location i) and a set of explanatory variables x_{ih}, for h = 1, … , k subject to random error

$$y_i = \Sigma_h \, x_{ih} \, \beta_h + e_i$$

e_i is a random error term, with $E[e_i] = 0$, i.e., no systematic error

Linear Regression - Notation (continued)

in matrix notation, a n x 1 column vector y and a n x k vector X, with a k x 1 coefficient vector β and a n x 1 random error vector e

$$y = X\beta + e$$

$$E[e] = 0$$

Marginal Effect

the effect of a change in X on y

in a linear regression the marginal effect equals the regression coefficient (this is not the case in a non linear regression)

$$E[y|\Delta X] = \Delta X \beta$$

Selected Regularity Conditions

X non-stochastic (or if stochastic, with bounds on second moment) - the only randomness follows from the dependent variable y, any randomness in X is inconsequential

error term independent identically distributed (i.i.d), i.e., $Var[e_i]=\sigma^2$ or $E[ee'] = \sigma^2 I$ = spherical error term

x_i and e_i uncorrelated for all i, i.e., signal (X) and noise (e) are not related

Ordinary Least Squares (OLS) Regression

under set of regularity conditions, yields the best (smallest variance) unbiased estimator = Gauss-Markov theorem

$$b = (X'X)^{-1} X'y$$

$$E[b] = E[(X'X)^{-1}X'(X\beta)] + E[(X'X)^{-1}X'e] = \beta$$

(since $E[X'e] = 0$)

Predicted Value

value of y_i given x_i using the estimates b

$$y_{ip} = \Sigma_h \, x_{ih} \, b_h$$

Residual

difference between observed and predicted

$$u_i = y_i - y_{ip}$$

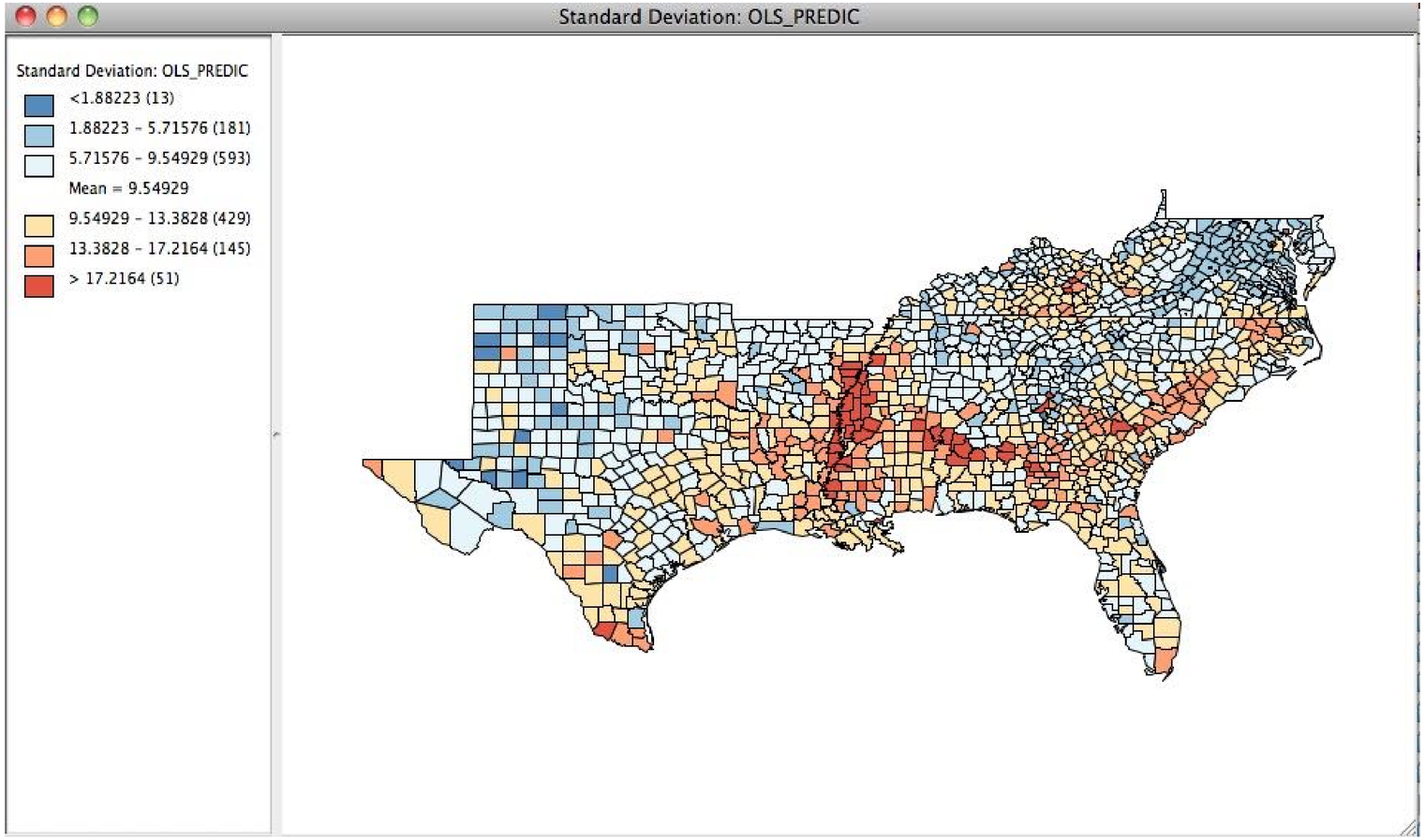
for regression with constant term avg[u_i] = 0
residual is NOT the same as the error term (e_i)

Visual Diagnostics

Predicted Value Map

shows spatial distribution of model prediction

a form of smoothing, i.e., what the model suggests y should be, given the X at each location



Predicted Value map

1990 county homicide rates (standard deviational map)

Residual Map

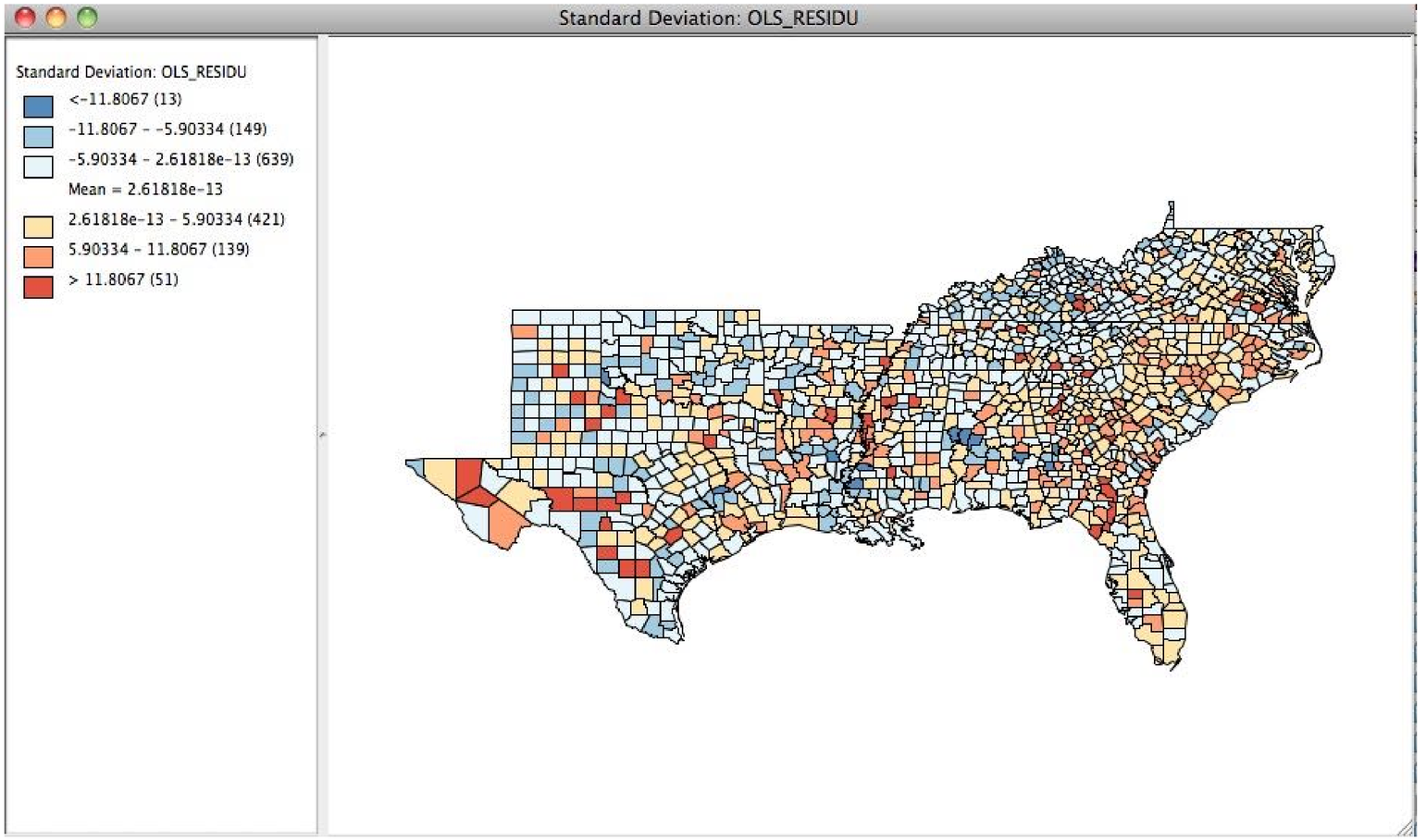
high values (red) = model under-predicts ($y > y_p$)

low values (blue) = model over-predicts ($y < y_p$)

note extremes = poor fit of model

note spatial patterns, but visual inspection can be misleading

need for formal diagnostics



Residual map - 1990 county homicide rates
(standard deviational map)

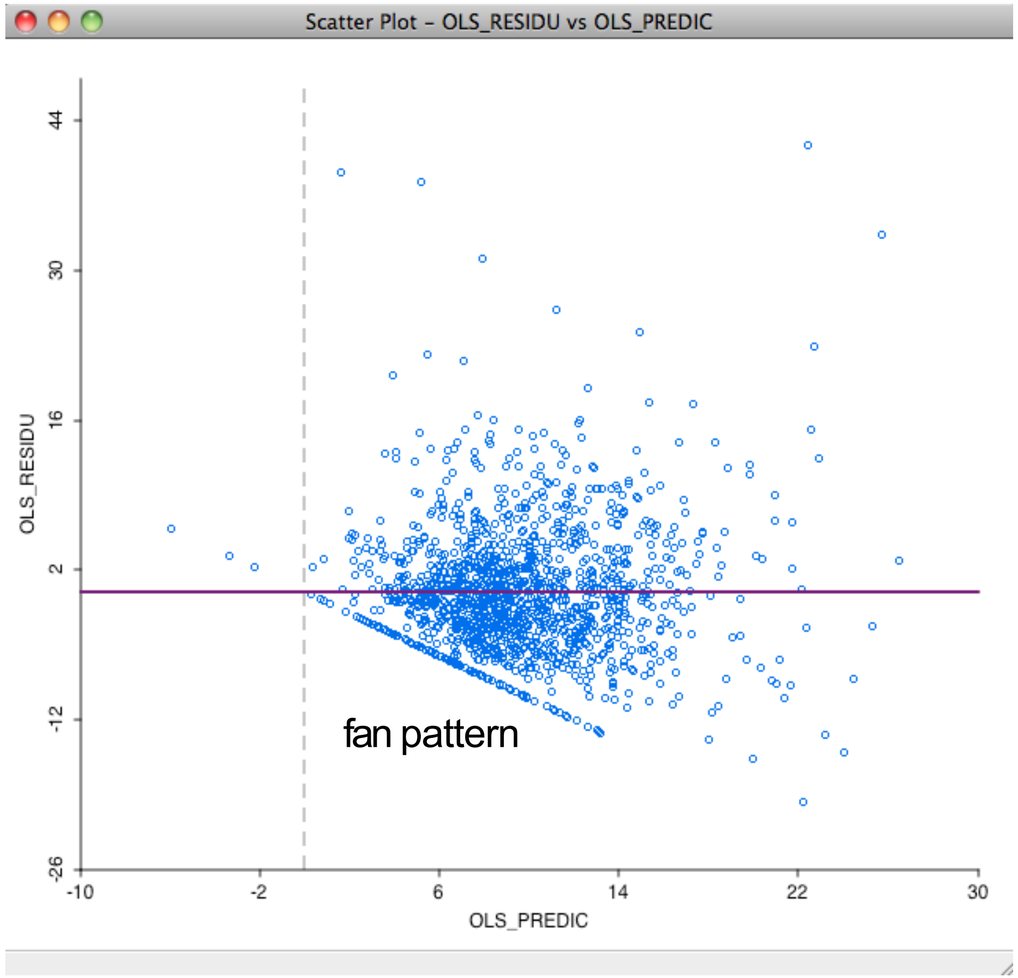
Diagnostic Plots

use scatter plot function

plot residuals (y-axis) vs predicted values (x-axis) as a visual diagnostic for heteroskedasticity

pattern should be more or less within the same range

“fan” or “flares” suggest heteroskedasticity, i.e., non-constant error variance



Residual Plot - Heteroskedasticity

Caution

visual inspection of plots and maps can be misleading

use plots and maps to suggest additional variables for model

no substitute for formal specification diagnostics

Specification Tests

```
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   30.863223
TEST ON NORMALITY OF ERRORS
TEST                    DF          VALUE            PROB
Jarque-Bera              2           2833.409        0.0000000


DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                    DF          VALUE            PROB
Breusch-Pagan test       5           515.0796        0.0000000
Koenker-Bassett test     5           124.2749        0.0000000
SPECIFICATION ROBUST TEST
TEST                    DF          VALUE            PROB
White                   20           242.8053        0.0000000
```

Regression Report - Diagnostics

Spatial Autoregressive Process

Spatial Autoregressive Process (SAR)

analog to time series setup

example, first order autoregressive process $y_t = \rho y_{t-1} + u_t$

by consecutive substitution, this becomes a moving average process in the error terms

$$y_t = \rho(\rho y_{t-2} + u_{t-1}) + u_t = \rho^2 y_{t-2} + \rho u_{t-1} + u_t.$$

$$y_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \dots.$$

Fundamental Difference = Feedback

substitution approach breaks down

y_i is present on the right hand side

$$\begin{aligned} y_i &= \rho[\rho(y_{i-2} + y_i) + u_{i-1}] + \rho[\rho(y_i + y_{i+2}) + u_{i+1}] + u_i \\ &= \rho^2(2y_i + y_{i-2} + y_{i+2}) + \rho u_{i-1} + \rho u_{i+1} + u_i \end{aligned}$$

alternative: write full system of simultaneous equations using a spatial weights matrix

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{u},$$

Spatially Lagged Variables

Specifying Spatial Dependence

In time series analysis, the concept of a shift y_{t-k} is observation shifted by k periods

For regular lattices, shift north, south, east, west: $y_{i-1,j}$, $y_{i+1,\ j}$, $y_{i,j-1}$, $y_{i,\ j+1}$ spatial shift of $y_{i,j}$

No analog for irregular spatial layouts

Instead, the notion of a spatially lagged variable (Anselin 1988)

Spatial Lag

Weighted average of neighboring values

Neighbors defined by spatial weights (w_{ij})

$$y_{iL} = w_{i1}.y_1 + w_{i2}.y_2 + \ldots + w_{iN}.y = \Sigma_j w_{ij}.y_j$$

In practice: very few neighbors (weights are sparse)

Spatial Lag in Matrix Notation

Spatial weights matrix times the vector of observations

$$y_L = Wy$$

Wy as such is often used as symbol for a spatially lagged dependent variable

Spatial Lag vs Window Average

Similar to a window average, the spatial lag is a smoother

Lag **Wy** has smaller variance than original variable y

spatial lag is NOT a window average since $\mathbf{w_{ii} = 0}$
observation at “center” of window is not included

Spatially Lagged Variables in a Regression

spatially lagged dependent variable
Wy spatial (autoregressive) lag model

spatially lagged explanatory variables
WX spatial cross-regressive model or SLX model

spatially lagged error terms
We spatial (autoregressive) error model

Spatial Lag Model

Motivation

Explicit model for spatial interaction = substantive spatial dependence

Substantive autocorrelation is when values of Y are systematically related to values of Y in adjacent areas, generating model bias. This can be corrected by including an explicit spatial lag term as an explanatory variable in the model

Peer-effects, etc.

Equilibrium outcome of spatial interaction process, a spatial reaction function (Brueckner 2003)

Non-behavioral motivation = data issue (scale)

Types of Social Interaction

Interaction effects among individual agents = **endogenous effects**

Exogenous group characteristics = **contextual effects**

Observed or unobserved characteristics that agents have in common = **correlated effects**

Mixed Regressive-Spatial Autoregressive

$$y = \rho Wy + X\beta + u$$

Wy = spatial autoregressive (spatial lag)

X = regressive

ρ = spatial autoregressive coefficient

Spatial Filter

Remove effect of spatial autocorrelation

$$y - \rho W y = X\beta + u$$

$$(I - \rho W) y = X\beta + u$$

$(I - \rho W)$ is **spatial filter**

Effect of Spatial Filter

Similar to detrending

Deals with scale problems, i.e., non-behavioral motivation for including spatial lag term

Spatial filter still requires estimate of ρ

Spatial Multiplier

Derived from reduced form

What is the change in y as a result of the change in X

$$E[y|\Delta X] = (I-\rho W)^{-1}(\Delta X)\beta = [I+\rho W+\rho^2 W^2 +...](\Delta X)\beta$$

effect is more than (ΔX) β ‹ multiplier

Direct and Indirect Effects

Total effect of a change in the exogenous variable

$$(I-\rho W)^{-1}\,(\Delta X)\beta$$

direct effect

$$(\Delta X)\beta$$

indirect effect

$$[(I - \rho W)^{-1} - I]\,(\Delta X)\beta$$

$$[\rho W + \rho^2 W^2 + \ldots]\,(\Delta X)\beta$$

Applications of Spatial Multiplier

Policy analysis

Effect of a change in a policy variable x at i extends beyond i to its neighbors, neighbors of neighbors, etc.

Simulate the spatial imprint of a policy change by solving the reduced form for a change in X

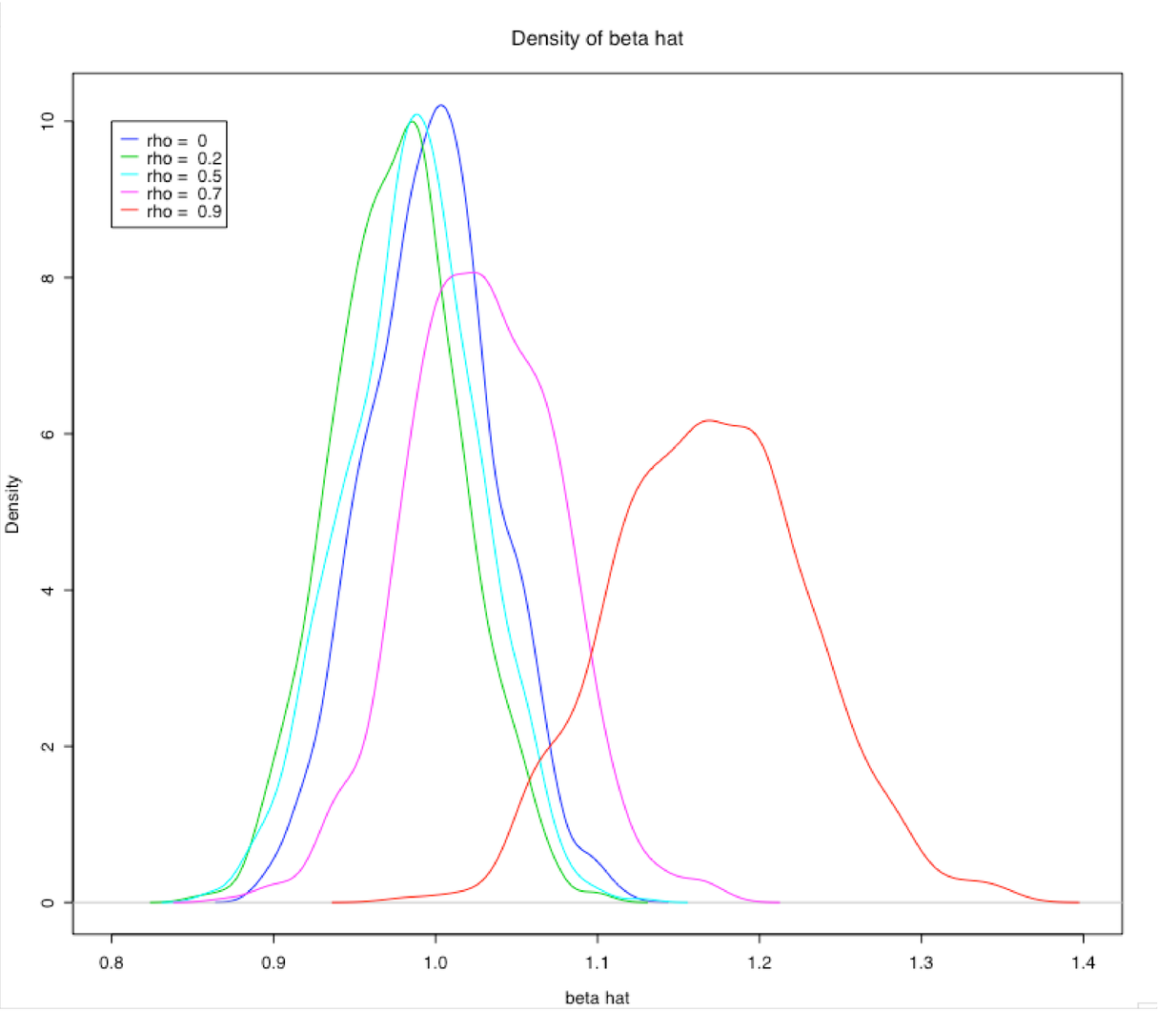
Effects of Ignoring a Spatial Lag

= ignoring substantive spatial interaction

omitted variable problem

OLS biased and inconsistent

Potentially: wrong estimate, wrong sign, wrong standard error, wrong significance, wrong fit



effect of ignoring spatial lag on OLS estimate

OLS vs. SAR

Consider the following linear regression of Obama's margin of victory **(y)** on county-level socio-economic attributes **(X)**:

$\mathbf{y} = \mathbf{X}\beta + \mathbf{s}$.

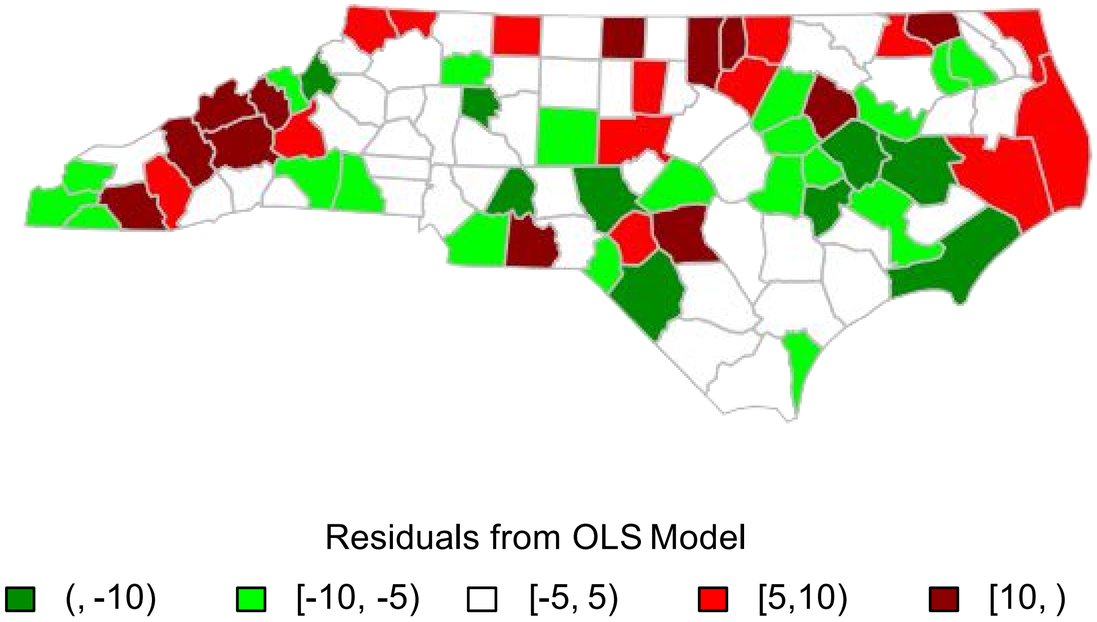
	OLS
(Intercept)	-35.58
	(6.23)***
Percent non-white	1.09
	(0.06)***
Percent college-educated	1.65
	(0.15)***
Veterans	-2.6e-4
	(1.2e-4)*
Median income	-7e-4
	(1.6e-4)***
AIC	729.2
N	100
Moran's *I* Residuals	0.25***

*p ≤ .05, **p ≤ .01, ***p ≤ .001

The Moran's I statistic shows a significant amount of spatial autocorrelation in the residuals.

OLS Residuals

Below is a map of residuals from a linear regression of Obama's margin of victory on county-level socio-economic attributes.



OLS vs. SAR

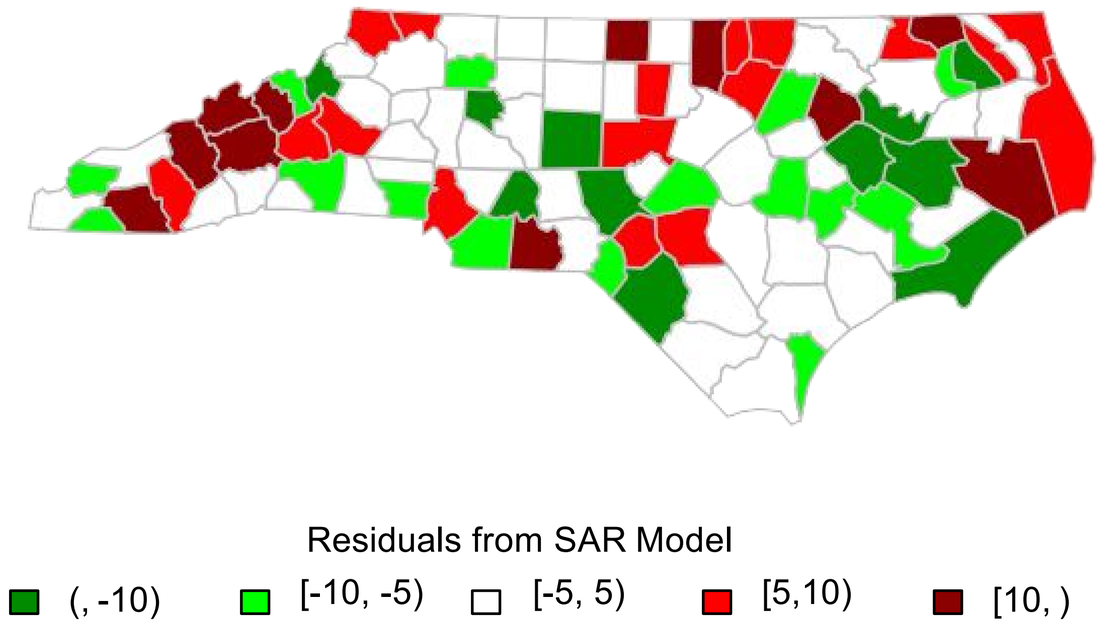
And the same model estimated by SAR: $y = \rho Wy + X\beta + u$.

	OLS	SAR
(Intercept)	-35.58	-28.40
	(6.23)***	(7.05)***
Percent non-white	1.09	0.98
	(0.06)***	(0.08)***
Percent college-educated	1.65	1.62
	(0.15)***	(0.14)***
Veterans	-2.6e-4	-1.8e-4
	(1.2e-4)*	(1e-4)
Median income	-7e-4	-7.8e-4
	(1.6e-4)***	(1.6e-4)***
Lagged Obama margin (ρ)		0.16
		(0.08)*
AIC	729.2	727.09
N	100	100
Moran's *I* Residuals	0.25***	0.15**
*p ≤ .05, **p ≤ .01, ***p ≤ .001		

The ρ coefficient is positive and significant, indicating spatial autocorrelation in the dependent variable. But Moran's I indicates that residuals remain clustered.

SAR Residuals

Below is a map of residuals from the SAR model.

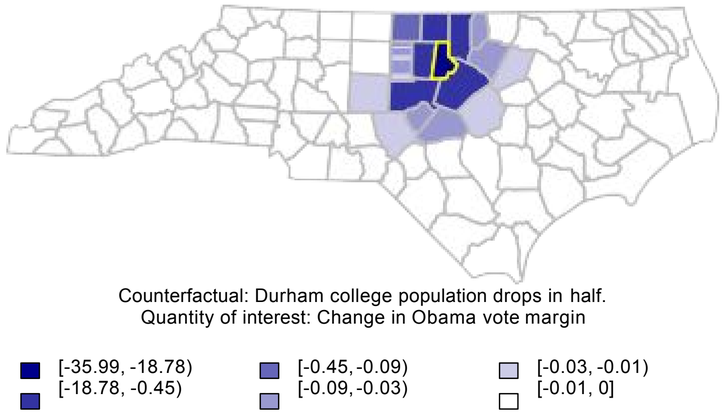


SAR Equilibrium Effects

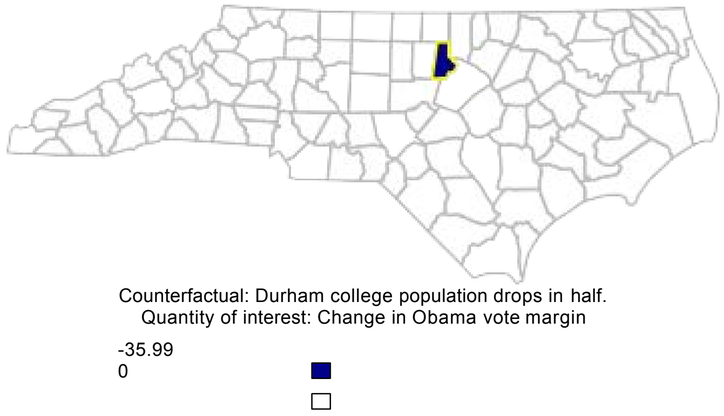
Counterfactual: A 50% decline in Durham's college-educated population.

Below are the equilibrium effects (change in Obama's county vote margin) associated with this counterfactual.

SAR



OLS



Spatial Error Model

Motivation

spatial pattern in error term due to omitted random factors = nuisance spatial dependence

Nuisance involves model residuals only - if this exists it reduces model efficiency and can be corrected including a spatial error specification in the model

mismatch spatial scale process with spatial scale observations (administrative units as “markets”)

no substantive interpretation problem of

efficiency of the estimates

Non-Spherical Error Variance

Due to spatial autocorrelation, error covariances are non-zero

Off-diagonal elements are non-zero

$E[uu'] = \Sigma \neq \sigma^2 I$

Spatial structure in the covariance $E[u_i u_j] \neq 0$, for $i \neq j$

Spatial Autoregressive Error Model

SAR error

$y = X\beta + u \text{ with } u = \lambda Wu + e$

covariance matrix $\Sigma = \sigma^2 [(I - \lambda W)'(I - \lambda W)]^{-1}$

but inverse covariance matrix does not contain inverse terms:

$\Sigma^{-1} = (1/\sigma^2) [(I - \lambda W)'(I - \lambda W)]$

Reduced Form

$$y = X\beta + (I - \lambda W)^{-1}e$$

No substantive spatial multiplier effect

Effect of spatial autocorrelation is on error variance, used in kriging (spatial prediction)

SAR Errors and Heteroskedasticity

Variance $\Sigma = \sigma^2 [(I - \lambda W)'(I - \lambda W)]^{-1}$ has non-constant diagonal terms - depends on number of neighbors

This induces heteroskedasticity in u, even with homoskedastic errors e

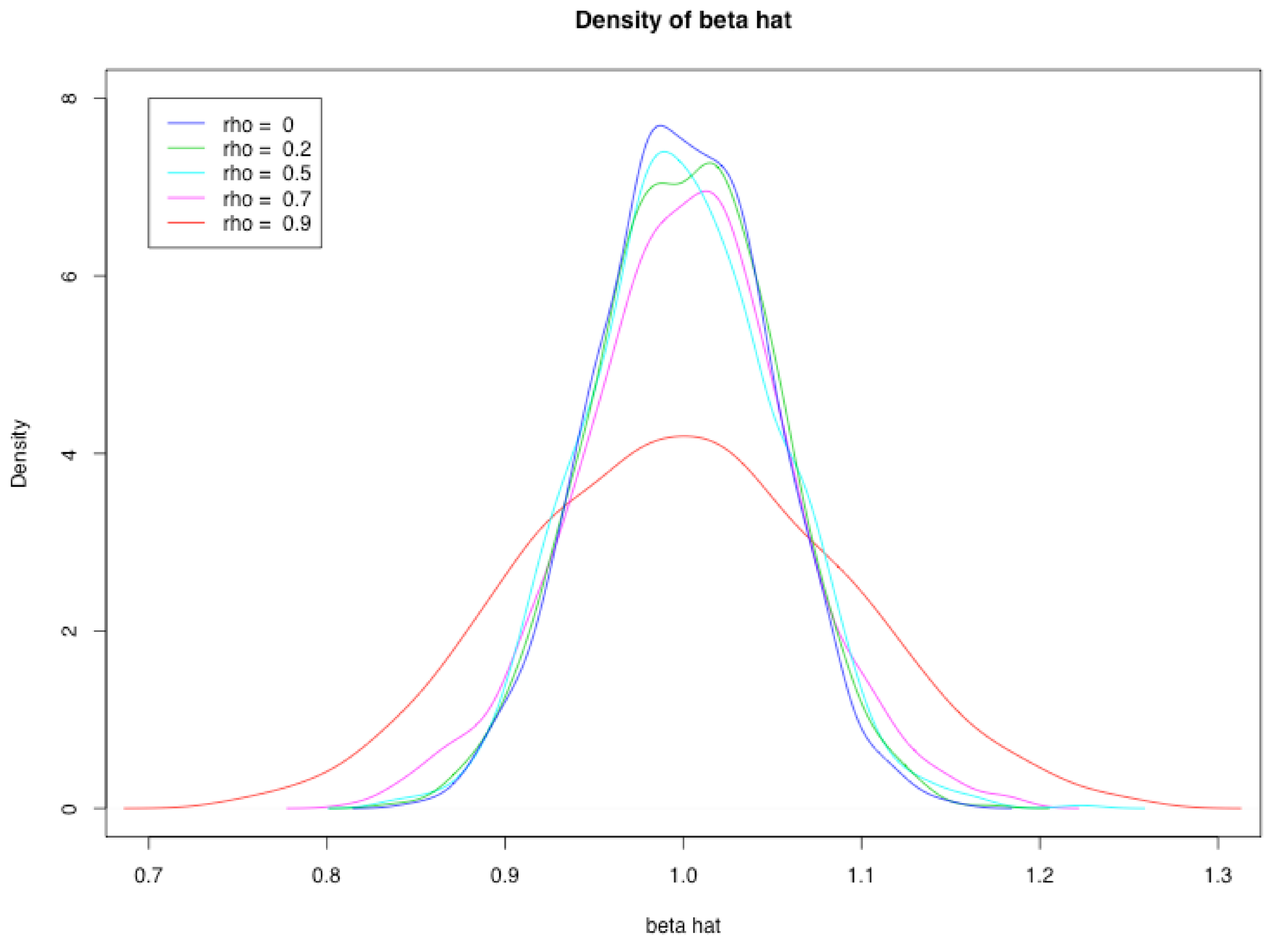
Difficult to disentangle true heteroskedasticity from induced heteroskedasticity

Effects of Ignoring SAR Errors

Problem of efficiency

OLS remains unbiased but inefficient

Potentially: correct estimate, wrong standard error, wrong significance, wrong fit



effect of ignoring SAR errors on OLS estimate

Spatially Lagged Error Model

Use of the spatial error model may be motivated by **omitted variable bias**.

Suppose that y is explained entirely by two explanatory variables x and z , where x, z ~ N(0, $\mathbf{I}_n$) and are independent.

$$y = x\beta + z\theta$$

If z is not observed, the vector $z\theta$ is nested into the error term u.

$$y = x\beta + u$$

Examples of latent variable z : culture, social capital, neighborhood prestige.

SEM Estimates

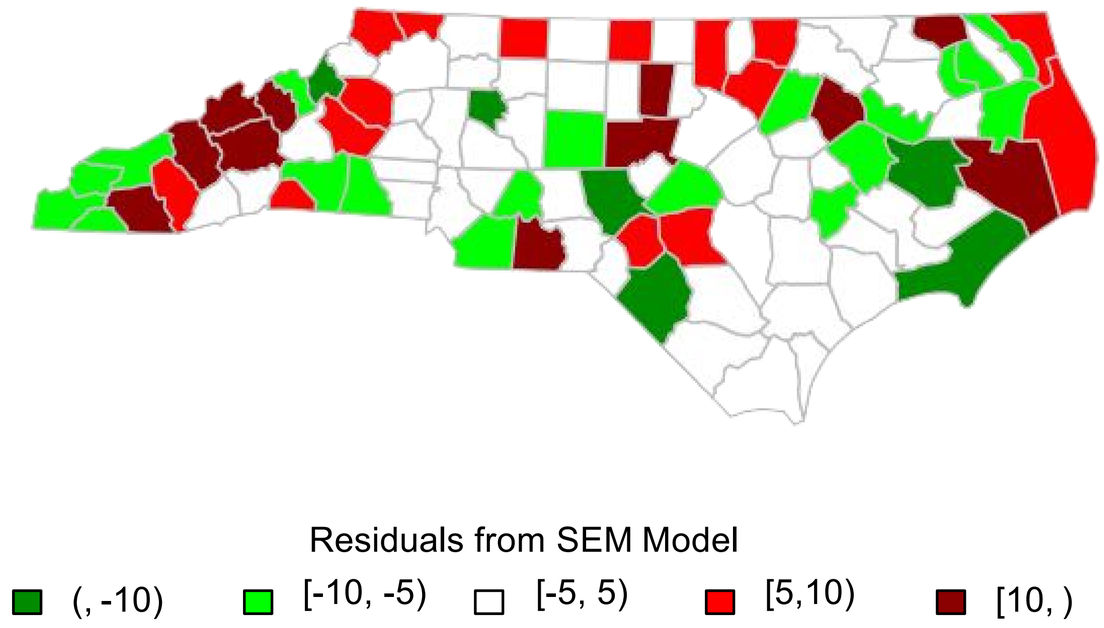
Let's run the model: $\mathbf{y} = \mathbf{X}\beta + \lambda\mathbf{W}\mathbf{u} + e$.

	OLS	SAR	SEM
(Intercept)	-35.58	-28.40	-38.67
	(6.23)***	(7.05)***	(7.34)***
Percent non-white	1.09	0.98	1.16
	(0.06)***	(0.08)***	(0.07)***
Percent college-educated	1.65	1.62	1.44
	(0.15)***	(0.14)***	(0.13)***
Veterans	-2.6e-4	-1.8e-4	-1.5e-4
	(1.2e-4)*	(1e-4)	(1e-4)
Median income	-7e-4	-7.8e-4	-5.9e-4
	(1.6e-4)***	(1.6e-4)***	(1.6e-4)***
Lagged Obama margin (ρ)		0.16	
		(0.08)*	
Lagged error (λ)			0.53
			(0.11)***
AIC	729.2	727.09	715.74
N	100	100	100
Moran's *I* Residuals	0.25***	0.15**	-0.003
$^{*}p \leq .05$, $^{**}p \leq .01$, $^{***}p \leq .001$			

The λ coefficient indicates strong spatial dependence in the errors.

SEM Residuals

Below is a map of residuals from the SEM model.



Spatial Durbin Model

Spatial Durbin Model

Like the SEM, the Spatial Durbin Model can be motivated by concern over **omitted variables**.

Recall the DGP for the SEM:

$$y = X\beta + (I - \lambda W)^{-1}u$$

Now suppose that **X** and **u** are correlated

One way to account for this correlation would be to conceive of **u** as a linear combination of **X** and an error term **v** that is independent of **X**.

$$u = X\gamma + v$$

$$v \sim N(0, \sigma^2 I)$$

where the scalar parameter γ and σ^2 govern the strength of the relationship between **X** and $\mathbf{z = (I - \lambda W)^{-1}}$

Spatial Durbin Model

Substituting this expression for **u**, we have the following DGP:

$$y = X\beta + (I_n - \lambda W)^{-1}(\gamma X + v)$$

$$y = X\beta + (I_n - \lambda W)^{-1}\gamma X + (I_n - \lambda W)^{-1}v$$

$$(I_n - \lambda W)y = (I_n - \lambda W)X\beta + \gamma X + v$$

$$y = \lambda W y + X(\beta + \gamma) + WX(-\lambda\beta) + v$$

This is the **Spatial Durbin Model (SDM)**, which includes a spatial lag of the dependent variable **y**, as well as the explanatory variables **X**

SDM Estimates

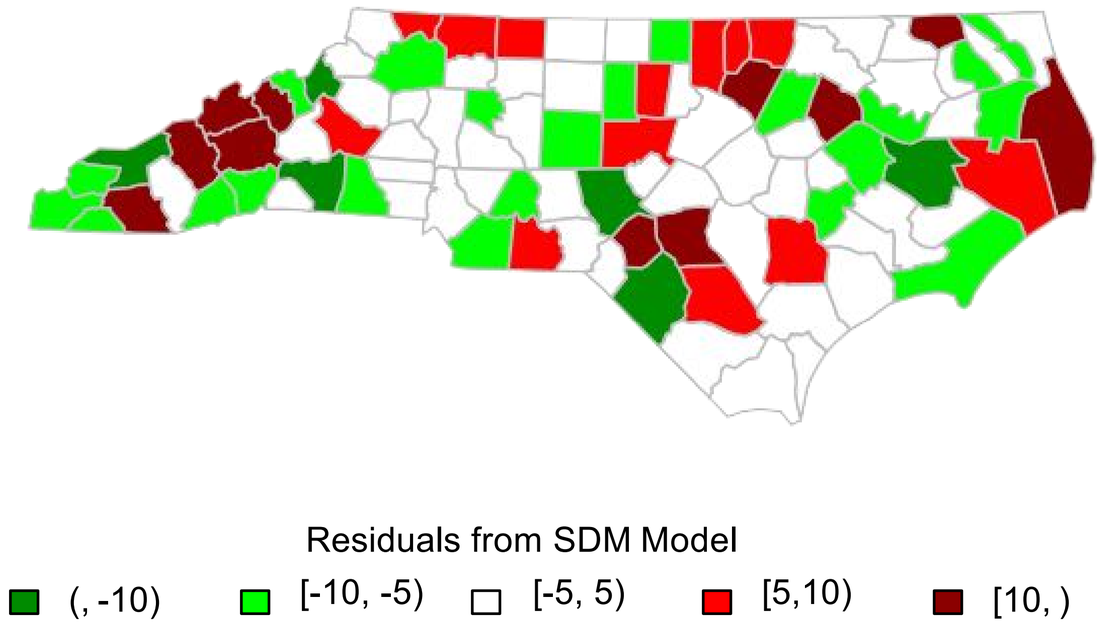
	OLS	SAR	SEM	SDM
(Intercept)	-35.58	-28.40	-38.67	-26.22
	(6.23)***	(7.05)***	(7.34)***	(9.66)***
Percent non-white	1.09	0.98	1.16	1.23
	(0.06)***	(0.08)***	(0.07)***	(0.92)***
.	.	.	.	.
Lagged Obama margin (ρ)		0.16		0.42
		(0.08)*		(0.12)**
Lagged error (λ)			0.53	
			(0.11)***	
Lagged non-white ($\theta_{\text{non-white}}$)				-0.59
				(0.17)***
.				.
AIC	729.2	727.09	715.74	714.22
N	100	100	100	100
Moran's *I* Residuals	0.25***	0.15**	-0.003	0.003

*p ≤ .05, **p ≤ .01, ***p ≤ .001

The SDM results in a slightly better fit

SDM Residuals

Below is a map of residuals from the SDM model.



SLX Model

Motivation for SLX Model

No spatially lagged dependent variable
deal with endogeneity in **WX** if needed

A model for local spillovers

No effect from **X** beyond first order neighbors

No need for spatial econometric estimators

Extensions: Spatial Durbin Error Model (SDEM)

The SDEM model contains spatial dependence in both the explanatory variables and the errors.

$$y = \iota_n\alpha + X\beta + WX\gamma + (I - \rho W)^{-1}s$$

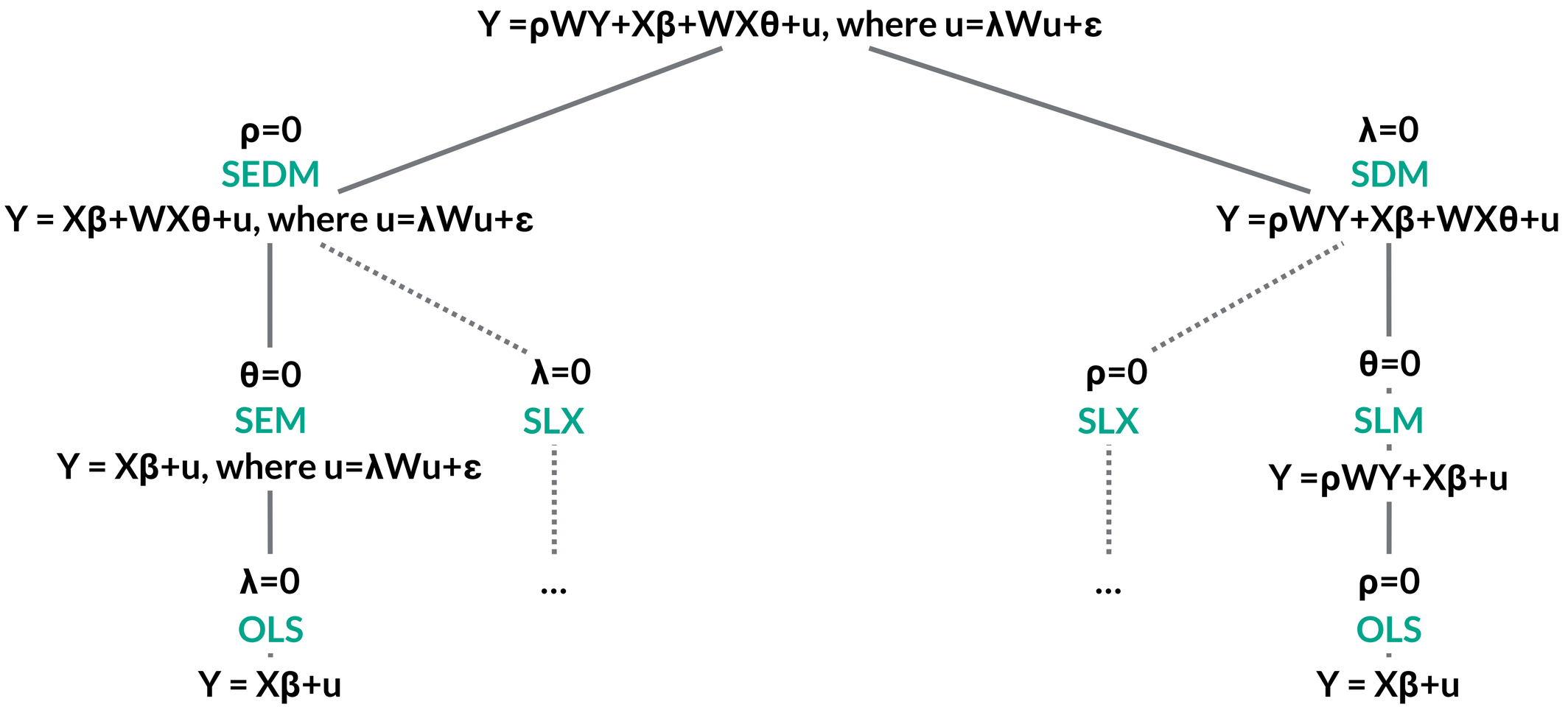
$$s \sim N(0, \sigma^2 I_n)$$

Direct impacts correspond to the β parameters; indirect impacts correspond to the γ parameters

The model can be generalized to incorporate two weights matrices without affecting interpretation of parameters:

$$y = \iota_n\alpha + X\beta + W_1X\gamma + (I_n - \rho W_2)^{-1}s$$

Nested Models



Specification tests

- Reject the null hypothesis against a fully specified alternative model, such as a spatial lag model or a spatial SAR error model
- Based on maximum likelihood (ML) principles
- Examples
  - Wald test (= asymptotic t-test)
  - Likelihood Ratio test (LR)
  - Lagrange Multiplier test (LM)
    - test on significance of slope (gradient) of likelihood function (score)
    - only requires estimation of null model
    - based on OLS residuals

GWR

Geographically Weighted Regression (GWR)

A key assumption that we have made in the models examined thus far is that the structure of the model remains constant over the study area (no local variations in the parameter estimates).

If we are interested in accounting for potential **spatial heterogeneity** in parameter estimates, we can use a Geographically Weighted Regression (GWR) model (Fotheringham et al., 2002).

GWR permits the parameter estimates to vary locally, similar to a parameter drift for a time series model.

GWR has been used primarily for exploratory data analysis, rather than hypothesis testing.

Geographically Weighted Regression (GWR)

GWR rewrites the linear model in a slightly different form:

$$y_i = X\beta_i + s$$

where **i** is the location at which the local parameters are to be estimated.

Parameter estimates are solved using a weighting scheme:

$$\beta_i = (X^j W_i X)^{-1} X^j W_i y$$

where the weight w_{ij} for the **j** observation is calculated with a Gaussian function.

$$w_{ij} = e^{\left(\frac{-d_{ij}}{h}\right)^2}$$

where $d_{i,j}$ is the Euclidean distance between the location of observation **i** and location **j** , and **h** is the bandwidth.

Bandwidth may be user-defined or selected by minimization of root mean square prediction error.

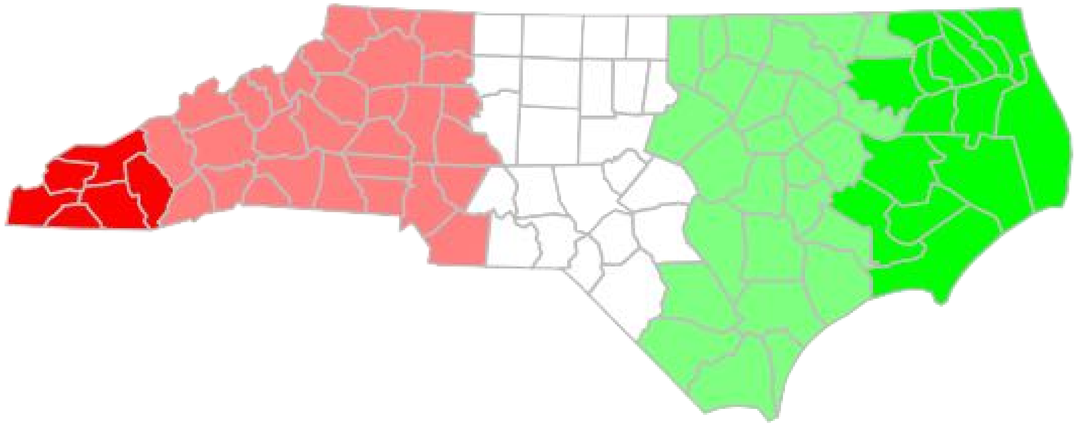
GWR Estimates

Let's try running the same election model as before with GWR:

	Geographically Weighted Regression				
	Global	Min	Median	Max	S.E.
(Intercept)	-35.58	-55.42	-37.65	-24.81	(8.64)
Percent non-white	1.09	0.99	1.12	1.25	(0.06)
Percent college-educated	1.65	1.44	1.63	1.83	(0.11)
Veterans	-3e-4	-3e-4	2.6e-4	-8e-5	(6e-5)
Median income	-7e-4	-1e-3	-9e-4	-3e-4	(2e-4)
Bandwidth	245131.2				
N	100				
Moran's *I* Residuals	0.218				
Moran's *I* Std. Deviate	3.645***				
	'$p \leq .1$, *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$				

GWR Local Coefficient Estimates

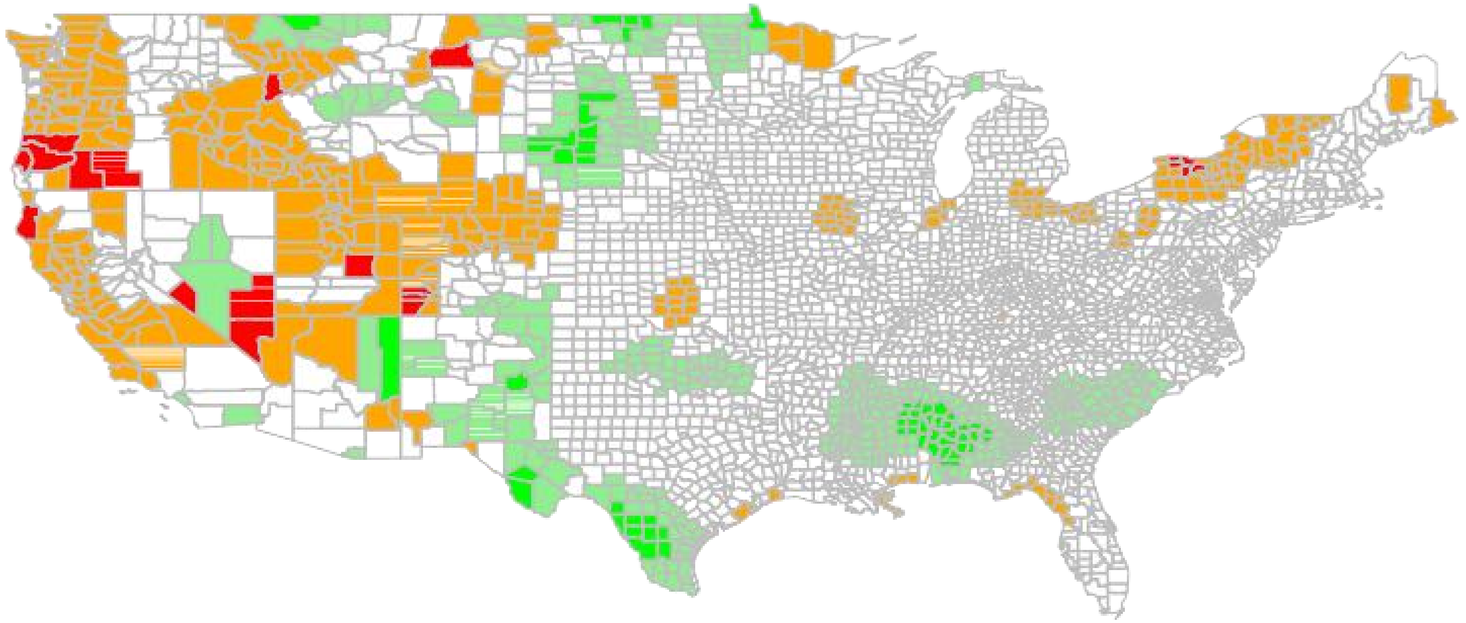
Below is a map of local coefficients. The relationship between college education and Obama's victory margin is largest in **red** areas, and smallest in **green** areas.



GWR Local Coefficient Estimates

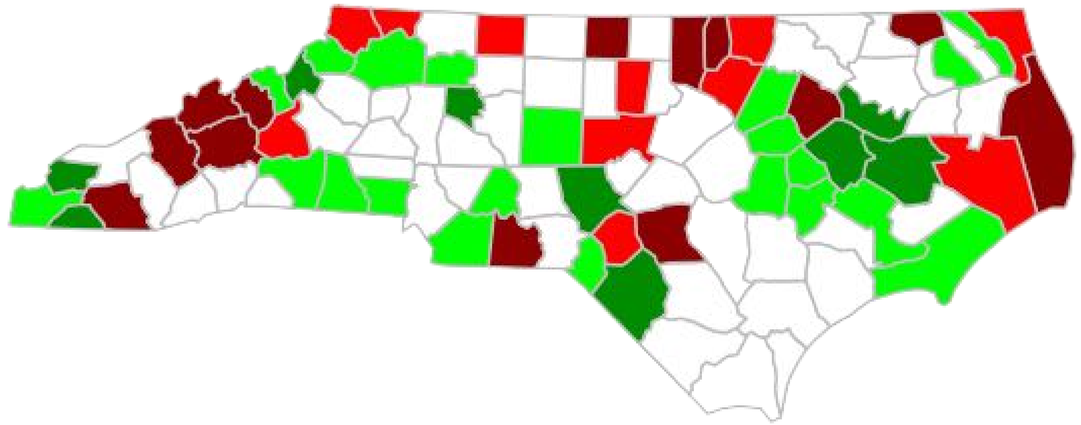
A more interesting example…

The relationship b/w per capita income and Bush's victory margin is negative in **red** areas, and positive in **green** areas.



GWR Residuals

Below is a map of residuals from the GWR model.



Resources

- Spatial Regression (Spring 2017) class
  - Luc Anselin (these slides are heavily based on his material)
  - https://spatial.uchicago.edu/directory/luc-anselin-phd
  - https://www.youtube.com/playlist?list=PLzREt6r1Nenkk7x197-CKPFZ0BuAOCRGT

@rschifan

schifane@di.unito.it

http://www.di.unito.it/~schifane