

Basics of Geographic Analysis in R

Spatial Regression

Yuri M. Zhukov

GOV 2525: Political Geography

February 25, 2013

Outline

1. Introduction
2. Spatial Data and Basic Visualization in R
3. Spatial Autocorrelation
4. Spatial Weights
5. Spatial Regression

Inefficiency of OLS estimators

- ▶ In a time-series context, the OLS estimator remains consistent even when a lagged dependent variable is present, as long as the error term does not show serial correlation.
- ▶ While the estimator may be biased in small samples, it can still be used for asymptotic inference.
- ▶ In a spatial context, this rule does not hold, irrespective of the properties of the error term.
- ▶ Consider the first-order SAR model (covariates omitted):

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \epsilon$$

- ▶ The OLS estimate for ρ would be:

$$\hat{\rho} = \left((\mathbf{W}\mathbf{y})'(\mathbf{W}\mathbf{y}) \right)^{-1} (\mathbf{W}\mathbf{y})'\mathbf{y} = \rho + \left((\mathbf{W}\mathbf{y})'(\mathbf{W}\mathbf{y}) \right)^{-1} (\mathbf{W}\mathbf{y})'\epsilon$$

- ▶ Similar to time series, the second term does not equal zero and the estimator will be biased.

Inefficiency of OLS estimators

- ▶ Asymptotically, the OLS estimator will be consistent if two conditions are met:

$$\begin{aligned}\text{plim } N^{-1}(\mathbf{W}\mathbf{y})'(\mathbf{W}\mathbf{y}) &= \mathbf{Q} \quad \text{a finite and nonsingular matrix} \\ \text{plim } N^{-1}(\mathbf{W}\mathbf{y})'\epsilon &= 0\end{aligned}$$

- ▶ While the first condition can be satisfied with proper constraints on ρ and the structure of \mathbf{W} , the second does not hold in the spatial case:

$$\text{plim } N^{-1}(\mathbf{W}\mathbf{y})'\epsilon = \text{plim } N^{-1}\epsilon'(\mathbf{W})(\mathbf{I}_n - \rho\mathbf{W})^{-1}\epsilon \neq 0$$

- ▶ The presence of \mathbf{W} in the expression results in a quadratic form in the error term.
- ▶ Unless $\rho = 0$, the plim will not converge to zero.

Properties of Maximum Likelihood Estimators

By contrast with OLS, maximum likelihood estimators (MLE) have attractive asymptotic properties, which apply in the presence of spatially lagged terms. ML estimates will exhibit consistency, efficiency and asymptotic normality if the following conditions are met:

- ▶ A log-likelihood for parameters of interest must exist (i.e.: non-degenerate $\ln L$)
- ▶ The log-likelihood must be continuously differentiable
- ▶ Boundedness of various partial derivatives
- ▶ The existence, positive definiteness and/or non-singularity of covariance matrices
- ▶ Finiteness of various quadratic forms

The various conditions are typically met when the structure of spatial interaction, expressed jointly by the autoregressive coefficient and the weights matrix, is nonexplosive (Anselin 1988).

Two-stage techniques

Instrumental variable estimation has similar asymptotic properties to MLE, but can be easier to implement numerically.

- ▶ Recall that the failure of OLS in models with spatially lagged DV's is due the correlation between the spatial variable and the error term ($\text{plim } N^{-1}(\mathbf{W}\mathbf{y})'\epsilon \neq 0$)
- ▶ This endogeneity issue can be addressed with two-stage methods based on the existence of a set of instruments \mathbf{Q} , which are strongly correlated with the original variables $\mathbf{Z} = [\mathbf{W}\mathbf{y} \quad \mathbf{X}]$, but asymptotically uncorrelated with the error term.

Two-stage techniques

- ▶ Where \mathbf{Q} is of the same column dimension as \mathbf{Z} , the instrumental variable estimate θ_{IV} is

$$\theta_{IV} = [\mathbf{Q}'\mathbf{Z}]^{-1}\mathbf{Q}'\mathbf{y}$$

- ▶ In the general case where the dimension of \mathbf{Q} is larger than \mathbf{Z} , the problem can be formulated as a minimization of the quadratic distance from zero:

$$\min\Phi(\theta) = (\mathbf{y} - \mathbf{Z}\theta)'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'(\mathbf{y} - \mathbf{Z}\theta)$$

- ▶ The solution to this optimization problem is the IV estimator θ_{IV}

$$\theta_{IV} = [\mathbf{Z}'\mathbf{P}_Q\mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{P}_Q\mathbf{y}$$

with $\mathbf{P}_Q = \mathbf{Q}[\mathbf{Q}'\mathbf{Q}]^{-1}\mathbf{Q}'$ an idempotent projection matrix

Two-stage techniques

- ▶ $\mathbf{P}_Q\mathbf{Z}$ can be seen to correspond to a matrix of predicted values from regressions of each variable in \mathbf{Z} on the instruments in \mathbf{Q}

$$\mathbf{P}_Q\mathbf{Z} = \mathbf{Q}\{[\mathbf{Q}'\mathbf{Q}]^{-1}\mathbf{Q}'\mathbf{Z}\}$$

- ▶ where the bracketed term is the OLS estimate for a regression of \mathbf{Z} on \mathbf{Q} .
- ▶ Let \mathbf{Z}_p be the predicted values of \mathbf{Z} . Then the IV estimator can also be expressed as

$$\theta_{IV} = [\mathbf{Z}'_p\mathbf{Z}]^{-1}\mathbf{Z}'_p\mathbf{y}$$

- ▶ which is the 2SLS estimator.

Two-stage techniques

Instrumental variable approaches are highly sensitive to the choice of instruments. Several options exist:

- ▶ Spatially lagged predicted values from a regression of \mathbf{y} on non-spatial regressors ($\mathbf{W}\mathbf{y}^*$) (Anselin 1980).
- ▶ Spatial lags of exogenous variables ($\mathbf{W}\mathbf{X}$) (Anselin 1980, Kelejian and Robinson 1993).
- ▶ In a spatiotemporal context, a time-wise lagged dependent variable or its spatial lag ($\mathbf{W}\mathbf{y}_{t-1}$) (Haining 1978).

Spatial autoregressive model (SAR): Likelihood function

- ▶ The **full log-likelihood** has the form:

$$\ln L = -\frac{n}{2} \ln(\pi\sigma^2) + \ln|\mathbf{I}_n - \rho\mathbf{W}| - \frac{e'e}{2\sigma^2}$$
$$e = (\mathbf{I}_n - \rho\mathbf{W})y - \mathbf{X}\beta$$

- ▶ It follows that the maximization of the likelihood is equivalent to a minimization of squared errors, corrected by the determinants from the Jacobian (Anselin 1988).
- ▶ This correction – and particularly the spatial term in $|\mathbf{I}_n - \rho\mathbf{W}|$ – will keep the least squares estimate from being equivalent to MLE.

Spatial autoregressive model (SAR): Likelihood function

- ▶ The most demanding part of the functions called to optimize the spatial autoregressive coefficient is the calculation of the Jacobian, the log-determinant of the $n \times n$ matrix $|\mathbf{I}_n - \rho\mathbf{W}|$
- ▶ One option is to express the determinant as a function of the eigenvalues ω of \mathbf{W} (Ord 1975):

$$\ln|\mathbf{I}_n - \rho\mathbf{W}| = \ln \prod_{i=1}^n (1 - \rho\omega_i) = \sum_{i=1}^n \ln(1 - \rho\omega_i)$$

- ▶ An alternative approach is brute-force calculation of the determinant and inverse matrix at each iteration.

OLS vs. SAR

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

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

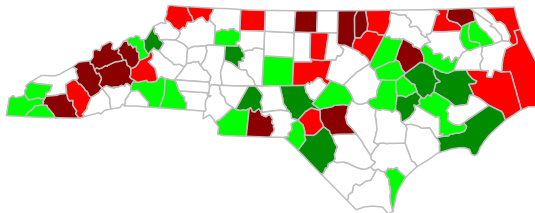
	OLS
(Intercept)	-35.58 (6.23)***
Percent non-white	1.09 (0.06)***
Percent college-educated	1.65 (0.15)***
Veterans	-2.6e-4 (1.2e-4)*
Median income	-7e-4 (1.6e-4)***
AIC	729.2
N	100
Moran's \mathcal{I} Residuals	0.25***

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

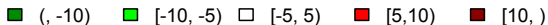
The Moran's \mathcal{I} statistic shows a significant amount of spatial autocorrelation in the residuals.

OLS Residuals

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



Residuals from OLS Model



OLS vs. SAR

And the same model estimated by SAR: $\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\beta + \epsilon$.

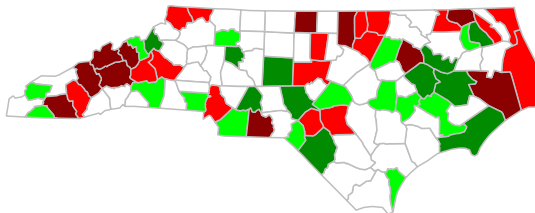
	OLS	SAR
(Intercept)	-35.58 (6.23)***	-28.40 (7.05)***
Percent non-white	1.09 (0.06)***	0.98 (0.08)***
Percent college-educated	1.65 (0.15)***	1.62 (0.14)***
Veterans	-2.6e-4 (1.2e-4)*	-1.8e-4 (1e-4)
Median income	-7e-4 (1.6e-4)***	-7.8e-4 (1.6e-4)***
Lagged Obama margin (ρ)		0.16 (0.08)*
AIC	729.2	727.09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

* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

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

SAR Residuals

Below is a map of residuals from the SAR model.



Residuals from SAR Model

■ $(-, -10)$ ■ $[-10, -5)$ □ $[-5, 5)$ ■ $[5, 10)$ ■ $[10,)$

SAR Equilibrium Effects

- ▶ Because of the dependence structure of the SAR model, coefficient estimates do not have the same interpretation as in OLS.
- ▶ The β parameter reflects the short-run direct impact of x_i on y_i . However, we also need to account for the indirect impact of x_i on y_i , from the influence y_i exerts on its neighbors y_j , which in turn feeds back into y_i .
- ▶ The equilibrium effect of a change in x_i on y_i can be calculated as:

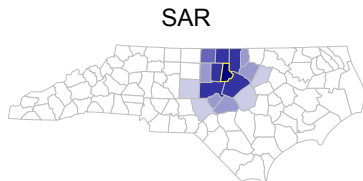
$$\mathbb{E}[\Delta y] = (\mathbf{I}_n - \rho \mathbf{W})^{-1} \Delta \mathbf{X}$$

where $\Delta \mathbf{X}$ is a matrix of changes to the covariates, and Δy is the associated change in the dependent variable.

- ▶ Since each unit will have a different set of connectivities to its neighbors, the impact of a hypothetical change in x_i will depend on which unit is being changed.

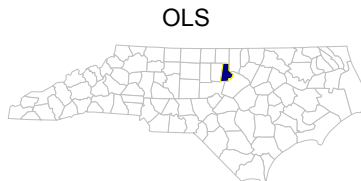
SAR Equilibrium Effects

- ▶ 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.



Counterfactual: Durham college population drops in half.
Quantity of interest: Change in Obama vote margin

■ [-35.99, -18.78)	■ [-0.45, -0.09)	■ [-0.03, -0.01)
■ [-18.78, -0.45)	■ [-0.09, -0.03)	□ [-0.01, 0]



Counterfactual: Durham college population drops in half.
Quantity of interest: Change in Obama vote margin

■ -35.99
□ 0

Spatially lagged error

- ▶ 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 $x, z \sim N(0, \mathbf{I}_n)$ and are independent.

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

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

$$y = x\beta + \epsilon$$

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

Spatially lagged error

- ▶ But we may expect the latent variable z to follow a spatial autoregressive process.

$$z = \lambda \mathbf{W}z + r$$

$$z = (\mathbf{I}_n - \lambda \mathbf{W})^{-1} r$$

- ▶ where $r \sim N(0, \sigma^2 \mathbf{I}_n)$ is a vector of disturbances, \mathbf{W} is the spatial weights matrix, and λ is a scalar parameter.
- ▶ Substituting this back into the previous equation, we have the DGP for the spatial error model (SEM) :

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

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

- ▶ where $u = \theta r$

Spatially lagged error

- ▶ In addition to omitted variable bias, another motivation for the spatial error model might be **spatial heterogeneity**.
- ▶ Suppose we have a panel data set, with multiple observations for each unit.
- ▶ If we want our model to incorporate individual effects, we can include an $n \times 1$ vector \mathbf{a} of individual intercepts for each unit:

$$\mathbf{y} = \mathbf{a} + \mathbf{X}\beta$$

- ▶ But in a cross-sectional setting, with one observation per unit, this approach is not feasible, since we'll have more parameters than observations.

Spatially lagged error

- ▶ Instead, we can treat \mathbf{a} as a vector of spatial random effects.
- ▶ We assume that the vector of intercepts \mathbf{a} follows a spatial autoregressive process:

$$\mathbf{a} = \lambda \mathbf{W} \mathbf{a} + \epsilon$$

$$\mathbf{a} = (\mathbf{I}_n - \lambda \mathbf{W})^{-1} \epsilon$$

- ▶ where $\epsilon \sim N(0, \sigma^2 \mathbf{I}_n)$ is a vector of disturbances
- ▶ Substituting this into the previous model yields the DGP of the SEM:

$$\mathbf{y} = \mathbf{X} \beta + \mathbf{a}$$

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

Spatially lagged error: Likelihood function

- ▶ The **full log-likelihood** has the form:

$$\ln L = -\frac{n}{2} \ln(\pi\sigma^2) + \ln|\mathbf{I}_n - \lambda\mathbf{W}| - \frac{e'e}{2\sigma^2}$$
$$e = (\mathbf{I}_n - \lambda\mathbf{W})(y - \mathbf{X}\beta)$$

Spatially lagged error: Interpretation of coefficients

- ▶ The SEM is essentially a generalized normal linear model with spatially autocorrelated disturbances.
- ▶ Assuming independence between \mathbf{X} and the error term, least squares estimates for β are not efficient, but still unbiased.
- ▶ Because the SEM does not involve spatial lags of the dependent variable, estimated β parameters can be interpreted as partial derivatives:

$$\beta_k = \frac{\delta y_i}{\delta x_{jk}} \quad \forall \quad i, k$$

- ▶ where i indexes the observations and k indexes the explanatory variables.

SEM Estimates

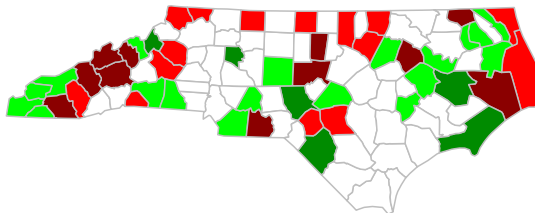
Let's run the model: $\mathbf{y} = \mathbf{X}\beta + \lambda\mathbf{W}\mathbf{u} + \epsilon$.

	OLS	SAR	SEM
(Intercept)	-35.58 (6.23)***	-28.40 (7.05)***	-38.67 (7.34)***
Percent non-white	1.09 (0.06)***	0.98 (0.08)***	1.16 (0.07)***
Percent college-educated	1.65 (0.15)***	1.62 (0.14)***	1.44 (0.13)***
Veterans	-2.6e-4 (1.2e-4)*	-1.8e-4 (1e-4)	-1.5e-4 (1e-4)
Median income	-7e-4 (1.6e-4)***	-7.8e-4 (1.6e-4)***	-5.9e-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 \mathcal{I} Residuals	0.25***	0.15**	-0.003
	* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$		

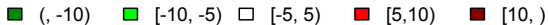
The λ coefficient indicates strong spatial dependence in the errors.

SEM Residuals

Below is a map of residuals from the SEM model.



Residuals from SEM 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:

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

- ▶ Now suppose that \mathbf{X} and \mathbf{u} are correlated.
- ▶ One way to account for this correlation would be to conceive of \mathbf{u} as a linear combination of \mathbf{X} and an error term \mathbf{v} that is independent of \mathbf{X} .

$$\mathbf{u} = \mathbf{X}\gamma + \mathbf{v}$$

$$\mathbf{v} \sim N(0, \sigma^2\mathbf{I}_n)$$

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

Spatial Durbin Model

- ▶ Substituting this expression for \mathbf{u} , we have the following DGP:

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

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

$$(\mathbf{I}_n - \lambda\mathbf{W})\mathbf{y} = (\mathbf{I}_n - \lambda\mathbf{W})\mathbf{X}\beta + \gamma\mathbf{X} + \mathbf{v}$$

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

- ▶ This is the Spatial Durbin Model (SDM), which includes a spatial lag of the dependent variable \mathbf{y} , as well as the explanatory variables \mathbf{X} .

Spatial Durbin Model

- ▶ The Spatial Durbin Model can also be motivated by concern over **spatial heterogeneity**.
- ▶ Recall the vector of intercepts \mathbf{a} :

$$\mathbf{a} = (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\epsilon$$

- ▶ Now suppose that \mathbf{X} and ϵ are correlated.
- ▶ As before, let's restate ϵ as a linear combination of \mathbf{X} and random noise \mathbf{v} .

$$\mathbf{a} = \mathbf{X}\gamma + \mathbf{v}$$

- ▶ Substituting this back into the SEM yields the same expression of SDM as before:

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

Spatial Durbin Model: Likelihood function

- ▶ Let's restate the SDM as follows:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \alpha \iota_n + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\theta + \epsilon$$

- ▶ The **log-likelihood** has a similar form to the SEM:

$$\ln L = -\frac{n}{2} \ln(\pi\sigma^2) + \ln |\mathbf{I}_n - \rho \mathbf{W}| - \frac{\mathbf{e}'\mathbf{e}}{2\sigma^2}$$
$$\mathbf{e} = \mathbf{y} - \rho \mathbf{W}\mathbf{y} - \mathbf{Z}\delta$$

- ▶ where $\mathbf{Z} = [\iota_n \quad \mathbf{X} \quad \mathbf{W}\mathbf{X}]$, $\delta = [\alpha \quad \beta \quad \theta]$, and ρ is bounded by $(\min(\omega))^{-1}, \max(\omega)^{-1}$, where ω is an $n \times 1$ vector of eigenvalues of \mathbf{W} .

SDM Estimates

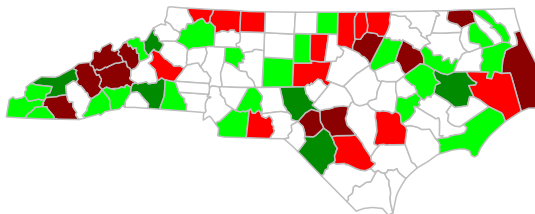
Let's try running the SDM: $\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \alpha \mathbf{1}_n + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\theta + \epsilon$

	OLS	SAR	SEM	SDM
(Intercept)	-35.58 (6.23)***	-28.40 (7.05)***	-38.67 (7.34)***	-26.22 (9.66)***
Percent non-white	1.09 (0.06)***	0.98 (0.08)***	1.16 (0.07)***	1.23 (0.92)***
⋮	⋮	⋮	⋮	⋮
Lagged Obama margin (ρ)		0.16 (0.08)*		0.42 (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 \mathcal{I} Residuals	0.25***	0.15**	-0.003	0.003
	* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$			

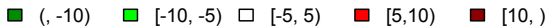
The SDM results in a slightly better fit...

SDM Residuals

Below is a map of residuals from the SDM model.



Residuals from SDM Model



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:

$$\mathbf{y}_i = \mathbf{X}\beta_i + \epsilon$$

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

- ▶ Parameter estimates are solved using a weighting scheme:

$$\beta_i = (\mathbf{X}'\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_i\mathbf{y}$$

- ▶ 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.

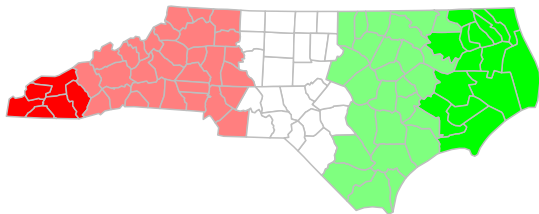
GWR Estimates

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

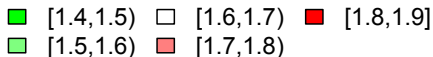
	Geographically Weighted Regression				
	Global	Min	Median	Max	S.E.
(Intercept)	-35.58	-55.42	-37.65	-24.81	(8.64)
Percent non-white	1.09	0.99	1.12	1.25	(0.06)
Percent college-educated	1.65	1.44	1.63	1.83	(0.11)
Veterans	-3e-4	-3e-4	2.6e-4	-8e-5	(6e-5)
Median income	-7e-4	-1e-3	-9e-4	-3e-4	(2e-4)
Bandwidth	245131.2				
<i>N</i>	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.



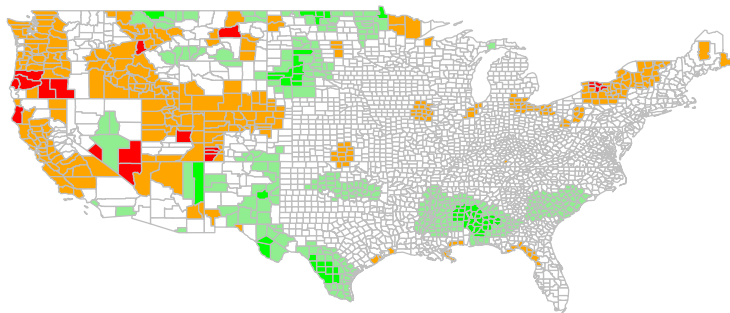
Local Coefficient Estimates (% college educated)



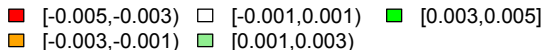
GWR Local Coefficient Estimates

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.

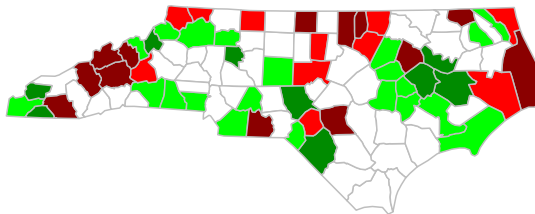


Local Coefficient Estimates (per capita income)



GWR Residuals

Below is a map of residuals from the GWR model.



Residuals from GWR Model

■ (, -10) ■ [-10, -5] □ [-5, 5] ■ [5,10] ■ [10,)

Extensions: Spatial Autocorrelation Model (SAC)

- ▶ The SAC model contains spatial dependence in both the dependent variable and the errors, with (potentially) two different weights matrices.

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X}\beta + \lambda \mathbf{W}_2 \mathbf{u} + \epsilon$$

$$\mathbf{y} = (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} \mathbf{X}\beta + (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} (\mathbf{I}_n - \lambda \mathbf{W}_2)^{-1} \epsilon$$

$$\epsilon \sim N(0, \sigma^2 \mathbf{I}_n)$$

- ▶ The **log-likelihood** has the form:

$$\ln L = -\frac{n}{2} \ln(\pi \sigma^2) + \ln |\mathbf{I}_n - \rho \mathbf{W}_1| + \ln |\mathbf{I}_n - \lambda \mathbf{W}_2| - \frac{\mathbf{e}' \mathbf{e}}{2\sigma^2}$$

$$\mathbf{e} = (\mathbf{I}_n - \lambda \mathbf{W}_2) ((\mathbf{I}_n - \rho \mathbf{W}_1) \mathbf{y} - \mathbf{X}\beta)$$

Extensions: Spatial Autoregressive Moving Average (SARMA)

- ▶ Like the SAC, the SARMA model also contains spatial dependence in the dependent variable and the errors.

$$\mathbf{y} = \iota_n \alpha + \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \beta + (\mathbf{I}_n - \theta \mathbf{W}_2) \epsilon$$

$$\mathbf{y} = (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} (\mathbf{X} \beta + \iota_n \alpha) + (\mathbf{I}_n - \rho \mathbf{W}_1)^{-1} (\mathbf{I}_n - \theta \mathbf{W}_2) \epsilon$$

$$\epsilon \sim N(0, \sigma^2 \mathbf{I}_n)$$

- ▶ The main distinction between the SAC and SARMA is the series representation of the inverse $(\mathbf{I}_n - \theta \mathbf{W}_2)$.
- ▶ As a result, the SAC places more emphasis on higher order neighbors.

Extensions: Spatial Durbin Error Model (SDEM)

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

$$\mathbf{y} = \iota_n \alpha + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\gamma + (\mathbf{I}_n - \rho\mathbf{W})^{-1}\epsilon$$
$$\epsilon \sim N(0, \sigma^2\mathbf{I}_n)$$

- ▶ 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:

$$\mathbf{y} = \iota_n \alpha + \mathbf{X}\beta + \mathbf{W}_1\mathbf{X}\gamma + (\mathbf{I}_n - \rho\mathbf{W}_2)^{-1}\epsilon$$

Examples in R

Switch to R tutorial script.