

# Applied Spatial Statistics in R, Section 6

## Spatial Regression

Yuri M. Zhukov

IQSS, Harvard University

January 19, 2010

# Outline

## ① Introduction

- Why use spatial methods?
- The spatial autoregressive data generating process

## ② Spatial Data and Basic Visualization in R

- Points
- Polygons
- Grids

## ③ Spatial Autocorrelation

## ④ Spatial Weights

## ⑤ Point Processes

## ⑥ Geostatistics

## ⑦ Spatial Regression

- Models for continuous dependent variables
- Models for categorical dependent variables
- Spatiotemporal models

# 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):

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

- The OLS estimate for  $\rho$  would be:

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

- 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:

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

$$\text{plim } N^{-1}(\mathbf{W}\mathbf{y})'\boldsymbol{\epsilon} = 0$$

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

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

- 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 ( $\text{plim } N^{-1}(\mathbf{W}\mathbf{y})'\boldsymbol{\epsilon} \neq 0$ )
- 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  $\theta_{IV}$  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  $\theta_{IV}$

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

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:

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

- 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  $\omega$  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 percent of county vote won by President Bush ( $\mathbf{y}$ ) on per capita income in the county ( $\mathbf{X}$ ):  $\mathbf{y} = \mathbf{X}\beta + \epsilon$ .

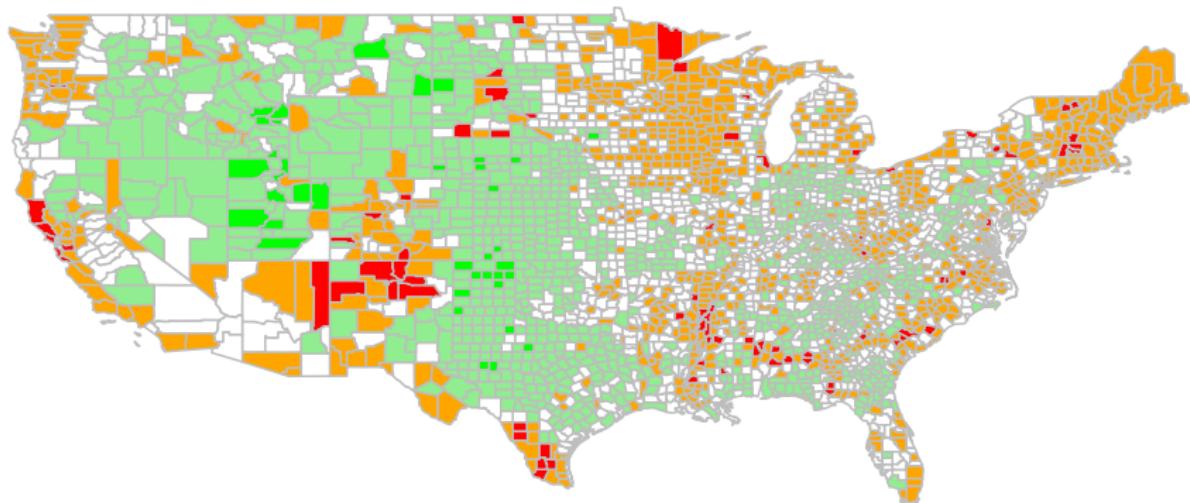
	OLS
(Intercept)	63.4340 (0.8893)***
Per capita income	-0.0002 (0.0000)***
AIC	24,666
N	3,111
Moran's $I$ Residuals	0.550
Moran's $I$ Std. Deviate	51.138***

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

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 percent of country vote received by Bush on per capita income.



Residuals from OLS Model

- [-50,-25]
- [-25,-5]
- [-5,5]
- [5,25]
- [25,50]

# OLS vs. SAR

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

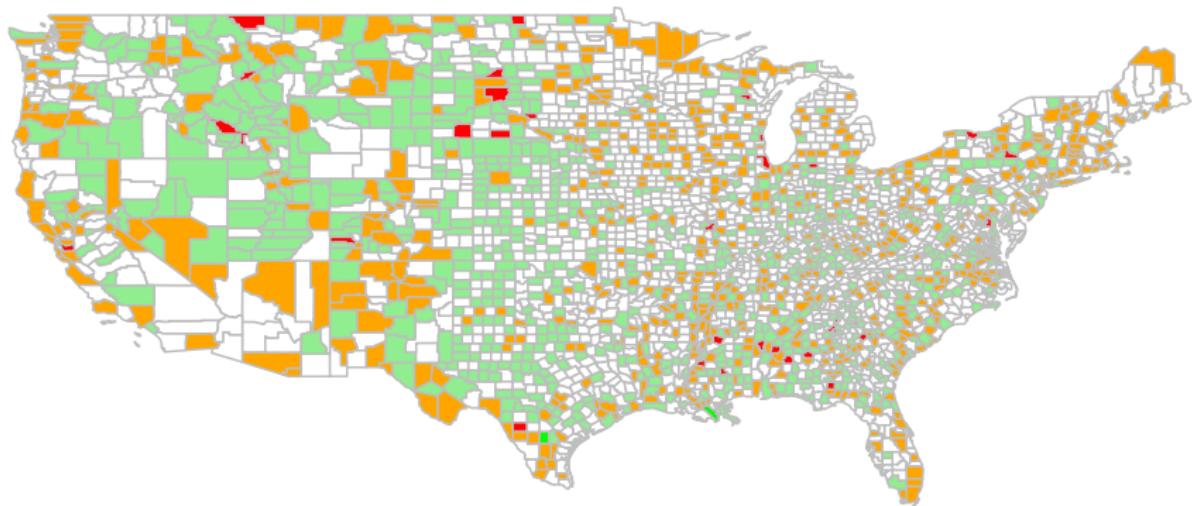
	OLS	SAR
(Intercept)	63.4340 (0.8893)***	14.073 (1.0572)***
Per capita income	-0.0002 (0.0000)***	5.46e-05 (3.38e-05)
$\rho$		0.7510 (0.0143)***
AIC	24,666	22,860
$N$	3,111	3,111
Moran's $I$ Residuals	0.550	-0.0410
Moran's $I$ Std. Deviate	51.138***	-3.7788

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

The  $\rho$  coefficient is positive and highly significant, indicating strong spatial autocorrelation in the dependent variable. The Moran's  $I$  statistic indicates that the residuals are no longer spatially clustered.

# SAR Residuals

Below is a map of residuals from the SAR model.



Residuals from SAR Model

- [-50,-25) ■ [-25,-5) □ [-5,5) ■ [5,25) ■ [25,50]

# SAR Equilibrium Effects

- Because of the dependence structure of the SAR model, coefficient estimates do not have the same interpretation as in OLS.
- The  $\beta$  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  $\Delta y$  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

- Below are the equilibrium effects (increase in percent of county vote for Bush) associated with a doubling of per capita income in Bronx County.

County	OLS	SAR
Currituck	0	4.81
Plymouth	0	4.47
Bronx	-2.86	4.31
Hunterdon	0	3.89
Lebanon	0	3.80
Mercer	0	3.70
Dare	0	3.52
Dauphin	0	3.42
Edgecombe	0	3.37
Barnstable	0	3.36

# 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  $x, z \sim N(0, I_n)$  and are independent.

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

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

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

- 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  $\lambda$  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:

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

# 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  $\beta$  are not efficient, but still unbiased.
- Because the SEM does not involve spatial lags of the dependent variable, estimated  $\beta$  parameters can be interpreted as partial derivatives:

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

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

# SEM Estimates

Let's run the election model from before:  $\mathbf{y} = \mathbf{X}\beta + \lambda\mathbf{Wu} + \epsilon$ .

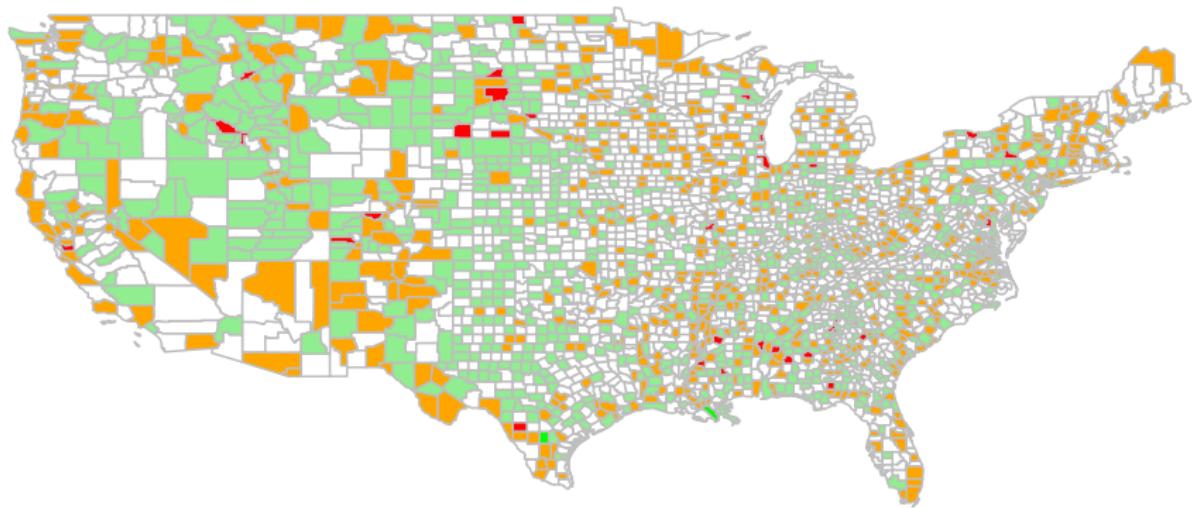
	OLS	SAR	SEM
(Intercept)	63.434 (0.8893)***	14.073 (1.0572)***	58.3470 (.9910)***
Per capita income	-0.0002 (0.0000)***	5.46e-05 (3.38e-05)	8.02e-05 (4.17e-05)'
$\rho$		0.7510 (0.0143)***	
$\lambda$			0.7612 (0.01422)***
AIC	24,666	22,860	22,864
N	3,111	3,111	3,111
Moran's $I$ Residuals	0.550	-0.0410	-0.0511
Moran's $I$ Std. Deviate	51.138***	-3.7788	-4.7192

' $p \leq .1$ , \* $p \leq .05$ , \*\* $p \leq .01$ , \*\*\* $p \leq .001$

The  $\lambda$  coefficient is positive and highly significant, indicating strong spatial dependence in the errors.

# SEM Residuals

Below is a map of residuals from the SEM model.



Residuals from SEM Model

- [-50,-25)
- [-25,-5)
- [-5,5)
- [5,25)
- [25,50]

# 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  $\gamma$  and  $\sigma^2$  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  $\mathbf{a}$ :

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

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

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

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

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

# Spatial Durbin Model: Likelihood function

- Let's restate the SDM as follows:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \alpha \boldsymbol{\iota}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \epsilon$$

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

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

$$\mathbf{e} = \mathbf{y} - \rho \mathbf{W}\mathbf{y} - \mathbf{Z}\boldsymbol{\delta}$$

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

# SDM Estimates

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

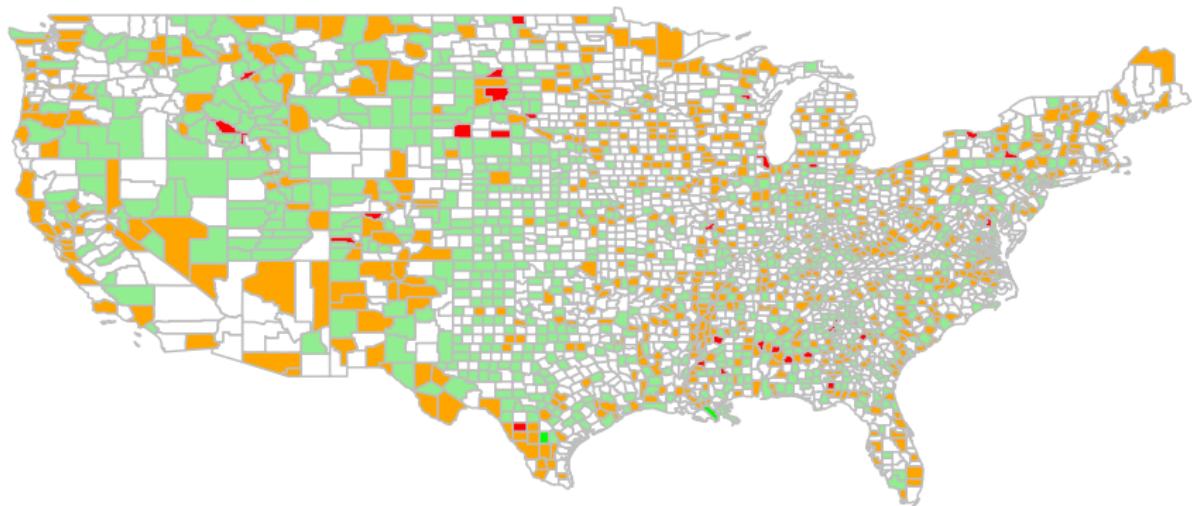
	OLS	SAR	SEM	SDM
(Intercept)	63.434 (0.8893)***	14.073 (1.0572)***	58.3470 (.9910)***	16.848 (1.2588)***
Per capita income	-0.0002 (0.0000)***	5.46e-05 (3.38e-05)	8.02e-05 (4.17e-05)'	0.0002 (0.0000)***
Lagged Bush vote ( $\rho$ )		0.7510 (0.0143)***		0.7501 (0.0144)***
Lagged error ( $\lambda$ )			0.7612 (0.01422)***	
Lagged income ( $\theta$ )				-0.0003 (0.0001)***
AIC	24,666	22,860	22,864	22,843
N	3,111	3,111	3,111	3,111
Moran's $I$ Residuals	0.550	-0.0410	-0.0511	-0.0454
Moran's $I$ Std. Deviate	51.138***	-3.7788	-4.7192	-4.1894

' $p \leq .1$ , \* $p \leq .05$ , \*\* $p \leq .01$ , \*\*\* $p \leq .001$

The SDM results in a slightly better fit...

# SDM Residuals

Below is a map of residuals from the SDM model.



Residuals from SDM Model

- [-50,-25)
- [-25,-5)
- [-5,5)
- [5,25)
- [25,50]

# 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 Model (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.

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

- Direct impacts correspond to the  $\beta$  parameters; indirect impacts correspond to the  $\gamma$  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. Section 6.a.

# 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.

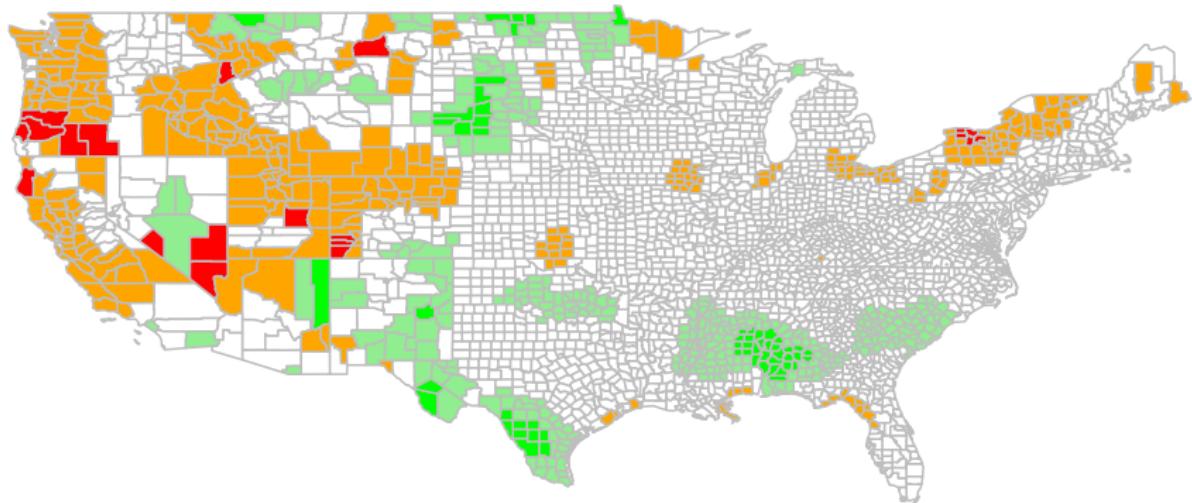
# GWR Estimates

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

	Geographically Weighted Regression				
	Global	Min	Mean	Max	S.E.
(Intercept)	63.4340	-26.02	59.95	185.36	(20.5262)
Per capita income	-0.0002	-0.0061	0.0001	0.0061	(0.0010)
Bandwidth	0.6649				
N	3,111				
Moran's $I$ Residuals	0.0796				
Moran's $I$ Std. Deviate	7.4239***				
	' $p \leq .1$ , * $p \leq .05$ , ** $p \leq .01$ , *** $p \leq .001$				

# GWR Local Coefficient Estimates

Below is a map of local coefficients. The relationship between income and support for Bush is negative in red areas, and positive in green areas.

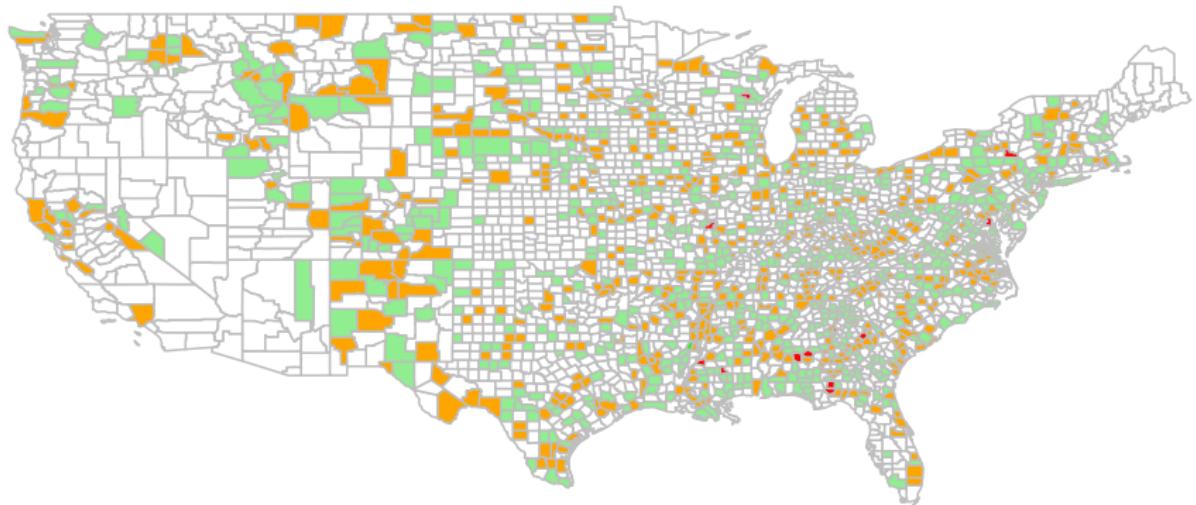


Local Coefficient Estimates (per capita income)

- [-0.005,-0.003] □ [-0.001,0.001] ■ [0.003,0.005]
- [-0.003,-0.001] ■ [0.001,0.003]

# GWR Residuals

Below is a map of residuals from the GWR model.



Residuals from GWR Model

- [-50,-25)
- [-25,-5)
- [-5,5)
- [5,25)
- [25,50]

# Examples in R

Switch to R tutorial script. Section 6.b.

# Spatial autologistic model

- Up to this point, we have only examined models which assume that the dependent variable is continuous and normally distributed.
- But what if we are interested in studying discrete events, measured categorically? (win/lose, war/peace, sick/healthy, Democrat/Republican, etc.)
- We may want to consider spatial dependence between observations with a conditional probability model, where the occurrence of an event  $y = 1$  in neighboring units conditions the likelihood that unit  $i$  will itself experience the event.
- One option for such a task is the spatial autologistic model (Ward and Gleditsch 2002).

# Spatial autologistic model

- The autologistic model states the conditional probability  $p_i$  that  $y_i = 1$ , given values  $y_j$  at units ( $j \neq i$ ):

$$p_i = P(y_i = 1 | \mathbf{W}y_i) = \frac{e^{\alpha + \mathbf{x}'_i \beta + \gamma \mathbf{W}y_i}}{1 + e^{\alpha + \mathbf{x}'_i \beta + \gamma \mathbf{W}y_i}}$$

- where  $\beta$  is a vector of parameters for exogenous variables,  $\gamma$  is a scalar parameter for the spatial lag of  $y$  and  $\mathbf{W}$  is a connectivity matrix.
- When  $\gamma = 0$ , this expression reduces to a standard logistic model and observations are considered independent of each other.
- When  $\beta = 0$ , this expression reduces to a pure autologistic model where unit-level covariates exert no independent influence on  $y$  once spatial dependence is taken into account.

# Spatial autologistic model

- A maximum pseudo-likelihood estimator (MPE) for the unknown parameter vector  $\theta = (\alpha \quad \beta \quad \gamma)$  is defined as the vector  $\hat{\theta}$  which maximizes

$$\prod_{i=1}^n P(y_i = 1 | \mathbf{W}y_i) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}$$

- An analytical form of the full likelihood is intractable because observations  $y_i$  are conditionally dependent on one another (Besag, 1974).
- Two solutions have been proposed:
  - ① Maximum pseuso-likelihood estimation (MPLE).
  - ② MCMC techniques.

# Spatial autologistic model: MPLE approach

- Maximum pseudo-likelihood estimation maximizes the function obtained by multiplying together the logit likelihoods represented by equation on the previous slide (Besag 1977).
- This is equivalent to a maximum likelihood fit for a logit regression model with independent observations  $y_i$ .
- This procedure has been shown to provide consistent estimates of model parameters (Cressie, 1993).
- However, the standard errors of the estimated parameters are not directly applicable because they assume independence of the observations.
- The inefficiency of MPLE increases when the strength of spatial interaction is high (Huffer and Wu 1998).

# Spatial autologistic model: MCMC approach

- Markov Chain Monte Carlo estimation can yield approximations closer to the full likelihood function (Geyer and Thompson 1992, Ward and Gleditsch 2002).
- One approach uses a probabilistic random map generated from the autologistic model, defined by parameters  $\theta$  and sufficient statistics  $s(y)$ .

$$s(y) = \left( \sum_{i=1}^n y_i, \quad \sum_{i=1}^n X_i y_i, \quad \frac{1}{2} \sum_{i=1}^n \mathbf{W} y_i \right)$$

- A statistic  $s(y)$  is sufficient for  $y$  if it contains all the information about  $y$  that is available in the sample.

# Spatial autologistic model: MCMC

- A Gibbs sampler is used to generate a set of  $m$  sampled simulated maps with sufficient statistics ( $y_{l \in \{1, \dots, m\}}$ )
- The samples are conditioned on the vector of parameters  $\psi$ , the initial values for which are typically pseudolikelihood estimates for  $\hat{\theta}$ .
- The idea is to find the values of  $\hat{\theta}$  that yield the sufficient statistics  $s(y)$  for the observed data.
- MCMC maximum likelihood is obtained by solving the score equation

$$s(y) = \frac{\sum_{l=1}^m s(y_m) e^{(\hat{\theta}-\psi)' s(y_m)}}{\sum_{j=1}^m e^{(\hat{\theta}-\psi)' s(y_m)}}$$

# Spatial autologistic model: MPLE vs. MCMC

- Let's try running the autologistic with some real data.
- Ward and Gleditsch (2002) estimate a simple model of war, where the probability of war in country  $i$  is conditioned on the number of neighboring countries experiencing war and the level democracy in those countries:

$$P(War_i = 1 | War_i' \mathbf{W}) = \frac{e^{\alpha + \beta_1 Dem_i + \beta_2 Dem_i' \mathbf{W}^s + \gamma War_i' \mathbf{W}}}{1 + e^{\alpha + \beta_1 Dem_i + \beta_2 Dem_i' \mathbf{W}^s + \gamma War_i' \mathbf{W}}}$$

- where  $Dem_i$  is a country  $i$ 's Polity score (scaled  $-10 : 10$  from least to most democratic),  $\mathbf{W}^s$  is a row-standardized weights matrix and  $\mathbf{W}$  is a binary contiguity matrix.

# Spatial autologistic model: MPLE vs. MCMC

MPLE ( $\hat{\psi}$ ) and MCMC ( $\hat{\theta}$ ) estimates for the model are shown below:

	MPLE		MCMC	
	Coef	S.E.	Coef	S.E.
(Intercept)	-1.87	(0.33)	-1.53	(0.09)
Democracy	-0.02	(0.03)	-0.06	(0.01)
Spatial Lag of Democracy	0.01	(0.05)	-0.02	(0.01)
Spatial Lag of War	0.31	(0.13)	0.21	(0.02)

So, parameter estimates are generally similar, but the standard errors from MCMC are much smaller.

# Examples in R

Switch to R tutorial script. Section 6.c.

# Partial Adjustment Model

- Recall that with cross-sectional data, we often assume that observations represent an equilibrium outcome of a spatiotemporal process working over time.
- Here we will examine how spatiotemporal models relate to models used for CS data (SAR, SEM).
- For simplicity, we assume that:
  - Units are influenced by their own and their neighbors' history (no simultaneous dependence).
  - $\mathbf{W}$  is symmetric.
  - No structural change over time.
  - The matrix  $\mathbf{X}$  – which may include spatial lags of explanatory variables – is constant or deterministically growing with respect to time.

# Partial Adjustment Model

A simple modeling framework for space-time data is the Partial Adjustment Model (PAM).

- Like a conventional temporal model, PAM allows the dependent variable for each unit  $y_t$  to depend on that unit's own past values  $y_{t-1}, \dots, y_0$  and  $X_{t-1}, \dots, X_0$ .
- This framework is extended to allow for spatial dependence on other regions by incorporating spatial lags of temporal lags  $\mathbf{W}y_{t-1}, \dots, \mathbf{W}y_0$  and  $\mathbf{W}X_{t-1}, \dots, \mathbf{W}X_0$ .
- For development of this model, see Greene (1997) and LeSage and Pace (2009).

# Partial Adjustment Model

The spatial PAM is formally defined below:

$$\begin{aligned}\mathbf{y}_t &= (\tau \mathbf{I}_n - \rho \mathbf{W}) \mathbf{y}_{t-1} + \mathbf{X}_t^* \beta + \mathbf{u}_t \\ \mathbf{X}_t^* &= \psi^t \mathbf{X}_0^* = \psi^t [\mathbf{X}_0 \quad \mathbf{W} \mathbf{X}_0 \quad \iota_n] \\ \mathbf{u}_t &= \mathbf{X}_t^* \gamma + r + \epsilon_t\end{aligned}$$

- Where  $\tau$  governs dependence between each region at time  $t$  and  $t - 1$ ,  $\rho$  governs spatial dependence between each region at time  $t$  and neighboring regions at  $t - 1$ ,  $\psi$  is the growth rate parameter for  $\mathbf{X}_0^*$  ( $\psi = 1$  implies no growth,  $\psi > 1$  implies growth; assume  $\psi > \tau$ ).
- As in the SDM, we allow for potential dependence between omitted variables and exogenous variables, such that the error term  $\mathbf{u}_t$  is partitioned into an endogenous component  $\mathbf{X}_t^* \gamma$ , an independent and time-invariant component  $r \sim N(0, \sigma_r^2 \mathbf{I}_n)$ , and independent noise  $\epsilon_t \sim N(0, \sigma_\epsilon^2 \mathbf{I}_n)$  which is allowed to vary with time.

# Partial Adjustment Model

This dynamic process implies a cross-sectional steady state characterized by simultaneous spatial interaction. To demonstrate this, we can use the recursive relation implied in the PAM:

$$\mathbf{y}_{t-1} = (\tau \mathbf{I}_n - \rho \mathbf{W}) \mathbf{y}_{t-2} + \mathbf{X}_{t-1}^* \beta + \mathbf{u}_{t-1}$$

The state of this dynamic system after the passage of  $t$  time periods is:

$$\mathbf{y}_t = (\tau \mathbf{I}_n - \rho \mathbf{W})^t \mathbf{y}_0$$

$$+ \left( \mathbf{I}_n \psi^t + (\tau \mathbf{I}_n - \rho \mathbf{W}) \psi^{t-1} + \dots + (\tau \mathbf{I}_n - \rho \mathbf{W})^{t-1} \psi \right) \mathbf{X}_0^* \beta + \tilde{\mathbf{u}}_t$$

$$\tilde{\mathbf{u}}_t = \widetilde{\mathbf{X}_t^* \gamma} + \tilde{r} + \tilde{\epsilon}_t$$

$$\widetilde{\mathbf{X}_t^* \gamma} = \left( \mathbf{I}_n \psi^t + (\tau \mathbf{I}_n - \rho \mathbf{W}) \psi^{t-1} + \dots + (\tau \mathbf{I}_n - \rho \mathbf{W})^{t-1} \psi \right) \mathbf{X}_0^* \gamma$$

$$\tilde{r} = \left( \mathbf{I}_n + (\tau \mathbf{I}_n - \rho \mathbf{W}) + \dots + (\tau \mathbf{I}_n - \rho \mathbf{W})^{t-1} \right) r$$

$$\tilde{\epsilon}_t = \epsilon_t + (\tau \mathbf{I}_n - \rho \mathbf{W}) \epsilon_{t-1} + (\tau \mathbf{I}_n - \rho \mathbf{W})^2 \epsilon_{t-2} + \dots + (\tau \mathbf{I}_n - \rho \mathbf{W})^{t-1} \epsilon_1$$

# Partial Adjustment Model

Taking the expectation of  $\mathbf{y}_t$  yields:

$$\begin{aligned}\mathbb{E}[\mathbf{y}_t] &\approx \left( \mathbf{I}_n \psi^t + (\tau \mathbf{I}_n - \rho \mathbf{W}) \psi^{t-1} + \dots + (\tau \mathbf{I}_n - \rho \mathbf{W})^{t-1} \psi \right) \mathbf{X}_0^* (\beta + \gamma) \\ &\approx \left( \mathbf{I}_n + (\tau \mathbf{I}_n - \rho \mathbf{W}) \psi^{-1} + \dots + (\tau \mathbf{I}_n - \rho \mathbf{W})^{t-1} \psi^{-(t-1)} \right) \psi^t \mathbf{X}_0^* (\beta + \gamma) \\ &\approx \left( \mathbf{I}_n - \frac{\rho}{\psi - \tau} \mathbf{W} \right)^{-1} \left( \frac{\psi}{\psi - \tau} \right) \mathbf{X}_t^* (\beta + \gamma) \\ &\approx (\mathbf{I}_n - \rho^* \mathbf{W})^{-1} \mathbf{X}_t^* \beta^*\end{aligned}$$

where  $\rho^* = \frac{\rho}{\psi - \tau}$  and  $\beta^* = \frac{\psi(\beta + \gamma)}{\psi - \tau}$ . This implies the familiar expression

$$\mathbf{y}_t = \rho^* \mathbf{W} \mathbf{y}_t + \mathbf{X}_t^* \beta^* + v_t$$

where  $v_t$  are the disturbances.

# Partial Adjustment Model

Let's consider the properties of  $\mathbf{y}_t = \rho^* \mathbf{W} \mathbf{y}_t + \mathbf{X}_t^* \beta^* + \nu_t$ .

- The spatial autoregressive parameter  $\rho$  is amplified by  $\psi - \tau$ , so that values of  $\psi > 1$  (implying growth in  $\mathbf{X}$ ) reduce the estimated spatial dependence of the system as measured by  $\rho^*$ . This gives more weight to the present.
- Lower values of  $\psi < 1$  (similarly, higher values of the temporal parameter  $\tau$ ) increase the role of the past, allowing more time for spatial influences to develop.
- ∴ even correctly-specified cross-sectional and spatiotemporal models could yield very different estimates of spatial dependence:
  - Cross-sectional samples place more emphasis on a long-run equilibrium result of a spatiotemporal process (i.e.: high spatial dependence).
  - Longitudinal samples place more emphasis on the temporal dependence parameters (i.e.: low spatial dependence).
  - But a process with low spatial dependence and high temporal dependence may still imply a long-run equilibrium with high spatial dependence.