Time-Series-Cross-Section Data Analysis

Spatial Econometric Models:
Transitioning from Time to Space

Jude C. Hays
jch61@pitt.edu

February 24, 2021

In the beginning, there was...

$$\mathbf{W}$$

Once we have **W**, we can

1. Test for spatial correlation in our outcomes.
2. Estimate Spatial Econometric Models.
3. Identify the source of spatial clustering.
4. Calculate diffusive effect across space.

Spatial and Temporal Dependence

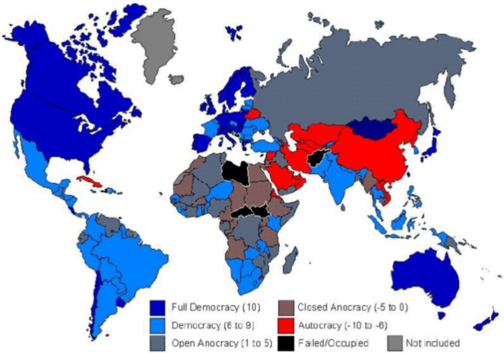
- When we observe outcomes (e.g., presidential approval) across time, they almost always exhibit temporal dependence.
- Knowing the outcome at time $t-1$ will help us to predict the outcome at time t.

Figure: Presidential Approval and the 9/11 Attacks

Spatial and Temporal Dependence

- When we observe outcomes (e.g., democracy scores) across space, they almost always exhibit spatial dependence.
- Knowing the outcome at at one location will help us to predict outcomes at proximate locations.

Figure: The Geography of Democracy



- Here's what we call a first-order spatial autoregressive model. It includes a **spatial weights matrix W** that produces a spatial lag of the outcome variable.

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\beta + \varepsilon$$

- Here's what we call a first-order temporal autoregressive model. It includes a **temporal lag operator** L that produces a temporal lag of the outcome variable.

$$y_t = \phi L^1 y_t + \mathbf{X_t}\beta + \varepsilon_t$$

Temporal Lag Operators and Spatial Weights Matrices

- The temporal lag operator L is defined as a linear operator such that

$$L^i y_t \equiv y_{t-i}$$

- Returning to the first-order temporal lag model

$$y_t = \phi L^1 y_t + \mathbf{X_t}\beta + \varepsilon_t \Leftrightarrow y_t = \phi y_{t-1} + \mathbf{X_t}\beta + \varepsilon_t$$

- Note that we could write this model in matrix notation using a temporal lag matrix

$$\mathbf{y} = \rho \mathbf{L}\mathbf{y} + \mathbf{X}\beta + \varepsilon$$

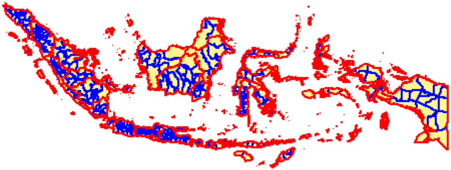
A Spatial-Weights Matrix...

$$\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .5 & .5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

...is an $N \times N$ matrix with elements w_{ij} that reflect the degree of connectivity from unit j to i.

- Shapefiles (ext .shp) contains information regarding the location, shape and attributes of geographical features.
- Shapefiles are available to download from many sites (e.g., `http://www.gadm.org/`).

Figure: GADM Shapefiles for Indonesia

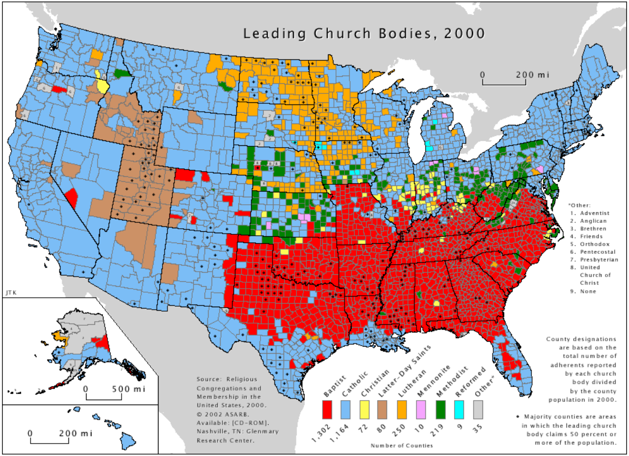


"This we know: All things are connected"

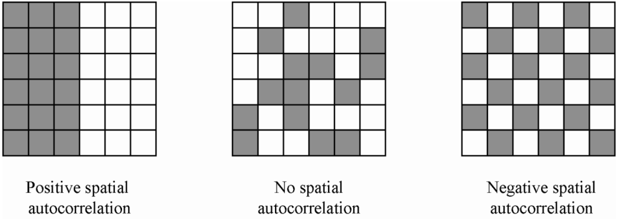
Mantra #1: Tobler's Law

"I invoke the first law of geography: everything is related to everything else, but near things are more related than distant things." - Tobler (1970)

Spatial Association, Correlations, Clustering



Spatial Association, Correlations, Clustering



Spatial Correlation

The first to recognize the inferential problems produced by spatial associatioins was Sir Francis Galton (lesser known for developing the following concepts: standard deviation, properties of the bivariate normal distribution, correlation, the regression line, and regression toward the mean).

Galton's (1889) Problem:

"It was extremely desirable for the sake of those who may wish to study the evidence for Dr. Tylor's conclusions, that full information should be given as to the degree in which the customs of the tribes and races which are compared together are independent. It might be, that some of the tribes had derived them from a common source, so that they were duplicate copies of the same original. Certainly, in such an investigation as this, each of the observations ought, in the language of statisticians, to be carefully weighted."

Spatial Associations

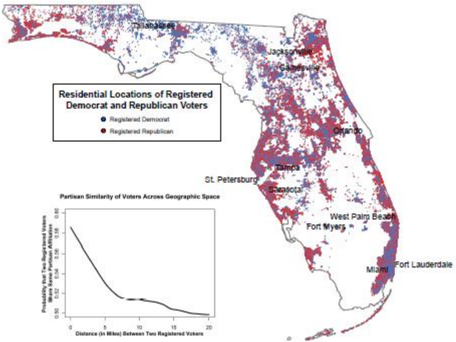
Spatial correlation from interactions arises in several ways:

1. Common Exposure (Observables & Unobservables)
2. Spillovers (Outcomes & Predictors)
3. Selection

Separating these kinds of effects in observation data is difficult because empirically they look very similar

Common Exposure, Spillovers, Selection

Figure 1: Tobler's Law and the Residential Locations of Florida Voters



Sources of Spatial Correlation

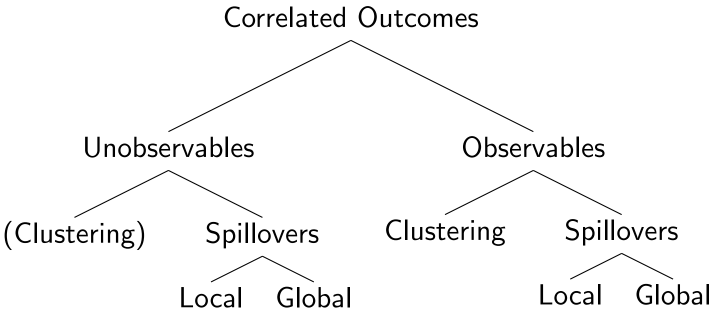
We can think about this more formally. Let $\mathbf{y}$ be an $n \times 1$ vector of outcomes (e.g., democracy). Correlated outcomes implies:

$$\mathrm{cov}(y_i, y_j) \neq 0$$

Suggested this could be for 4 reasons (or any combination):

- $y_i = f(x_i)$ and $\mathrm{cov}(x_i, x_j) \neq 0$: Clustering on Observables
- $y_i = f(x_i, x_j)$: Spillovers in Covariates
- $y_i = f(\epsilon_i)$ and $\mathrm{cov}(\epsilon_i, \epsilon_j) \neq 0$: Clustering on Unobservables
- $y_i = f(y_j)$: Spillovers in Outcomes (Interdependence)

Sources of Spatial Correlation



Spatial Theory/Analysis in Social Sciences

Spatial econometrics is a relatively new field (as compared to time series), yet it's impact across the social sciences is already profound:

1. Many theoretical contexts where spatial dependence indicated
2. Mutlifarious mechanisms by which spatial dependence arises
3. Wide substantive range where we expect spatial dependence
4. Already central in many literatures

General Spatial Theory: Common Shocks

Actors possess similar characteristics (ex. natural endowments which span across units) causing unit outcomes to co-vary. Formally, we could note, as in Andrews (2005), these common-factor residuals and/or predictors as satisfying:

$$u_i = C_g' u_i^*$$
$$x_i = C_g' x_i^*$$

where C_g is a random common (e.g., group) factor with random factor loadings u_i^* & x_i^*. Therefore, if units i and j are each members of group g they are jointly impacted by the respective loading.

E.g., policy or technological innovations which change in the costs of inputs or demand (holding supply fixed) impact the revenues of producers of a good even where there is no direct interaction between them.

General Spatial Theory: Interdependence

Strategic, and therefore spatial, theories of interdependence are ubiquitous and central in the social sciences:

- i's preferences, utilities, actions, choices, outcomes depends on j
- Definition: $y_j = f(y_{j \neq i})$ not simply $corr(y_i, y_{j \neq i}) \neq 0$ correlated

Brueckner's (2003) generic theory of strategic interactions is widely applicable

- negative externalities $\implies$ strategic complementary (respond in same direction) $\implies$ competitive races & early-mover advantages
- positive externalities $\implies$ strategic substitution (respond in opposite directions) $\implies$ free-riding & late-mover advantages

General Spatial Theory: Spatial Spillovers

Interdependence seems natural, and likely, when the unit is individuals (or proxies), but with social aggregates – unemployment rates, crime rates, aggregate demand for cigarettes – this seems unlikely. Instead we expect spillovers:

- knowledge spillovers
- industry spillovers
- growth spillovers

Spatial in Social Science: Multifarious Mechanisms

- In political science:
  - Simmons et al (2005) list is influential in IR & CP
    - Competition
    - Coercion
    - Learning
    - Emulation
  - Other influential works in long history on policy diffusion in American politics literature
- Long discussed in geography:
  - Haegerstrand (1967) canonical
- Manski (2000) summarizes micro & macro-econometric perspective:
  - strategic interdependence arises any time some unit's actions affect the marginal utility of other's actions via interactions, expectations, or preferences

Spatial in Social Science: Substantive Contexts

- Security Policy (e.g., alliances, wars, etc. . .)
  - decisions to initiate, intervene/enter, exist
- Environmental Politics (e.g., air-pollution regulation)
  - Environmental spillovers
  - Costs of regulation
- Regulatory Politics (e.g., standard setting)
  - Attractiveness depends on who & how many others use
- Legislative Behavior: representatives votes depend on others' votes
- Electoral Politics & Voting Behavior
  - candidate selection, campaign strategy, etc. . .
- Coups, riots, revolts. . .
- Peer-effects
- International (or interstate) diffusion (e.g., policies, institutions, regimes)
- CPE & IPE: increasing globalization, monetary policy, etc.

Point Process

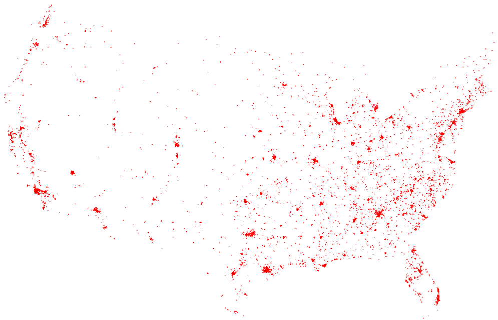


Figure: US Homicides in 2015 (Points)

- Inverse distance
- K-nearest neighbor
- Sphere of influence

Polygons

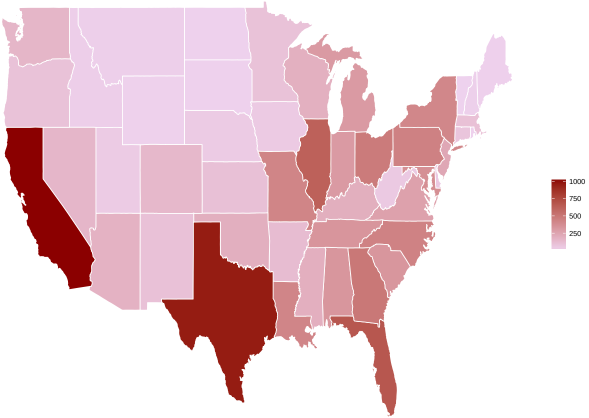


Figure: US Homicides in 2015 (Polygons)

Representing space with polygons

There are two types of weights commonly used:

1. Weights based on shared membership/boundaries
2. Weights based on distance

Weights based on membership/border

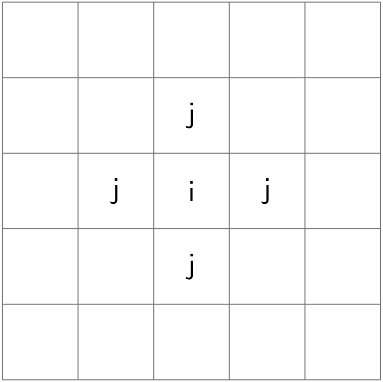
These weights matrices have discrete elements w_{ij} which assume a value of 1 if some condition (e.g., shared border, joint group membership, etc. . .) is met and 0 otherwise. An oft-used example is contiguity:

$$w_{ij} = \begin{cases} 1, & \text{if } border_i \cap border_j \neq \emptyset \\ 0, & \text{otherwise.} \end{cases}$$

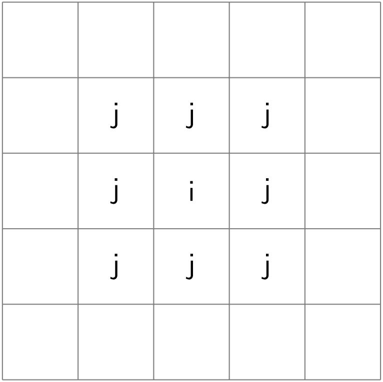
$$W = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

- Matrices tend to be sparse
- Often confront 'islands'
- Can make distance-based weights discrete (e.g., 1 if $d_{ij} < 50$)

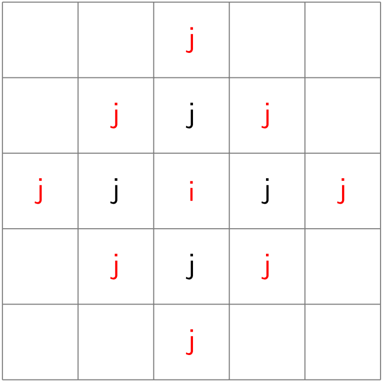
Regular Grids & Polygons: Rook Contiguity



Regular Grids & Polygons: Queen Contiguity



Regular Grids & Polygons: 2nd order (rook)



Weights based on distance

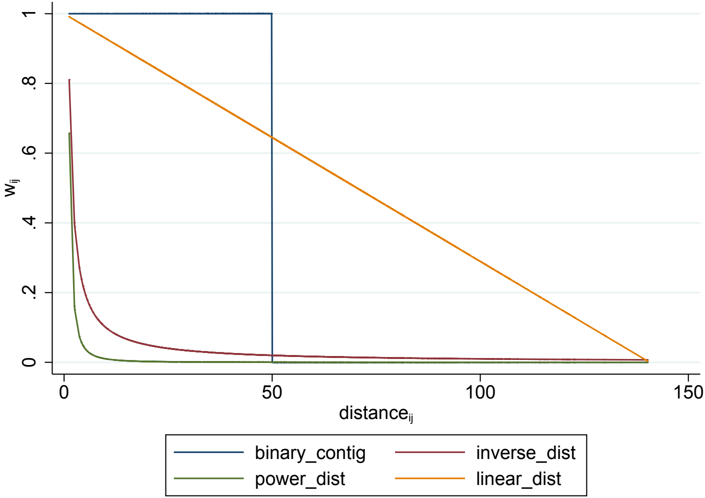
These weights matrices are based on the distance d_{ij} (centroid-centroid, capital-capital, etc...) between each unit i and j. An oft-used example is inverse-distance:

$$w_{ij} = \frac{1}{d_{ij}}$$

$$W = \begin{bmatrix} 0 & {}^1\!/d_{ij} & {}^1\!/d_{ij} \\ {}^1\!/d_{ij} & 0 & {}^1\!/d_{ij} \\ {}^1\!/d_{ij} & {}^1\!/d_{ij} & 0 \end{bmatrix}$$

- Diagonal w_{ii} is always 0 (cannot be a neighbor with yourself)
- Distance-based spatial-weights matrices are dense

Implications of different measures



On the spatial-weights matrix

Important to consider when specifying the spatial-weights matrix:

1. How accurately does $\mathbf{W}$ capture the true network?
2. Is $\mathbf{W}$ exogenous?
3. (How) should I normalize $\mathbf{W}$?

How accurate is **W**?

We never observe the true connectivity in the network. **W** is merely the researchers best guess at the implied network.

If **W** is far off consequences are clear:

- power of diagnostic tests is low
- model parameters are not consistently estimated
- accurate model specification is impaired

How accurate is **W**?

Given the importance of **W**, how should one select this matrix?

- Let theory guide you (Neumayer and Plumper, forthcoming)
- Select **W** empirically using goodness-of-fit criteria
  - AIC/BIC as discussed previously
  - Log-likelihood (Stakhovych and Bijmolt 2009)
  - Bayesian posterior model probability (LeSage and Pace 2009)
- Estimate multiple-**W** models (Franzese, et al. lecture ??)

Critique: LeSage and Pace (2014, 218) argue that the notion that inferences are "sensitive to the use of a particular weight matrix as perhaps the biggest myth about spatial regression models"

- Note, however, they are referring to different transformations of fundamentally similar measures (e.g., inverse distance vs. power distance weights)

Is W exogenous?

As we noted, we treat the spatial weights matrix as an exogenous (and known) network (known to, and pre-supplied by researcher). In non-geographic weight settings this assumption seems increasingly specious.

- Consequences: Inconsistent estimators
- Solutions: Co-evolution models

An area within Spatial Econometrics that still demands more attention.

Should we normalize **W**?

Researchers often normalize their spatial weights matrix to:

- Avoid singularities
- Remove dependence on scale factors
- Facilitate interpretation

Two common approaches arerow normalization and max-eigenvalue normalization.

Other approaches: column normalization (Leenders 2002) and Cliff and Ord (1975) approach discussed in Corrado and Fingleton 2012.

Relevant literature: Kelejian and Prucha 2010; Corrado and Fingleton 2012; Vega and Elhorst 2014; Neumayer and Plumper (forthcoming).

Row Normalization

Row normalization deflates all elements w_{ij} by the total number of connections for i such that rows sum to unity. That is, every non-zero element of the weights matrix W_{RS} is now given as:

$$\frac{w_{ij}}{\sum_j^n w_{ij}}$$

Benefits: This restricts the parameter space (for spatial effects) to the interval $(-1, 1)$, avoiding singularities in the spatial multiplier. Furthermore, it allows us to interpret spatial effects as representing the average effect of ones neighborhood.

Costs: causes lags to lose theoretical meaning (Vega and Elhorst 2014)

- Asymmetric influence as $w_{ij} \neq w_{ji}$
- Mutual proportions between the elements is lost

Max-Eigenvalue Normalization

To avoid the issues of row normalization, various scalar normalization strategies have been proposed. Most widely advanced is to scale the matrix by the maximum eigenvalue:

$$\frac{\mathbf{W}}{\lambda_{max}}$$

As with row normalization max-eigenvalue normalization restricts the parameter space to the interval (-1,1), avoiding singularities, but does not change the mutual proportions between the elements of the weights matrix.

Next steps...

Once we have **W**, we

1. Test for spatial correlation in our outcomes.
2. Estimate Spatial Econometric Models.
3. Identify the source of spatial clustering.
4. Calculate diffusive effect across space.

We will cover these topics next time.