Modeling the Ecology of Urban Inequality in Space and Time

Corina Graif
PhD Candidate, Department Of Sociology
Harvard University

Presentation for the Workshop on Spatial and Temporal Modeling, Center for Geographic Analysis, Harvard University, April 10, 2009

Overview

I. Space and time in sociological thinking and research

> Focus on urban processes and neighborhood level dynamics
> Research questions
> Sociological theories and research gaps

- Approaches to modeling time and space

> Some limitations and some solutions
> Useful software

I. Less conventional ways to model space and time

II. Key substantive and methodological results

III. Future research

Types of research questions

- How durable are the effects of poverty on neighborhood social capital?
- How do immigration and diversity impact neighborhood development?
  - in-migration of creative class
  - changes in crime rates over time
  - social cohesion and collective efficacy

Why do location and time matter?

Neighborhoods and Change

<=>

Space and Time

The underlying social processes:

- are locally and globally embedded in SPACE
- require TIME to unfold

Theory is saturated with spatial and temporal arguments

- Social Disorganization Theory
  - disadvantage, heterogeneity, and residential instability may disrupt neighborhood institutions and social cohesion
- Segmented Assimilation Theory
  - children of immigrants assimilating into a highly *disadvantaged* neighborhood may adopt oppositional subcultures
- Residential Segregation
  - spatial clusters of disadvantaged neighborhoods are durable traps and catalysts for crime
  - yet, ethnic enclaves may work as social support and buffers against discrimination and crime
- Subcultural Theory of Urbanism
  - Neighborhood diversity may home-breed or attract new, creative and hybrid subcultures (e.g. artists or the creative class)

However…

- Much of the research on neighborhoods
  - uses cross-sectional data and methods of analysis (cost reasons)
  - not accounting for spatial clustering in social and structural processes and outcomes

- Even if accounted for,

  SPACE is treated AS A NUISANCE

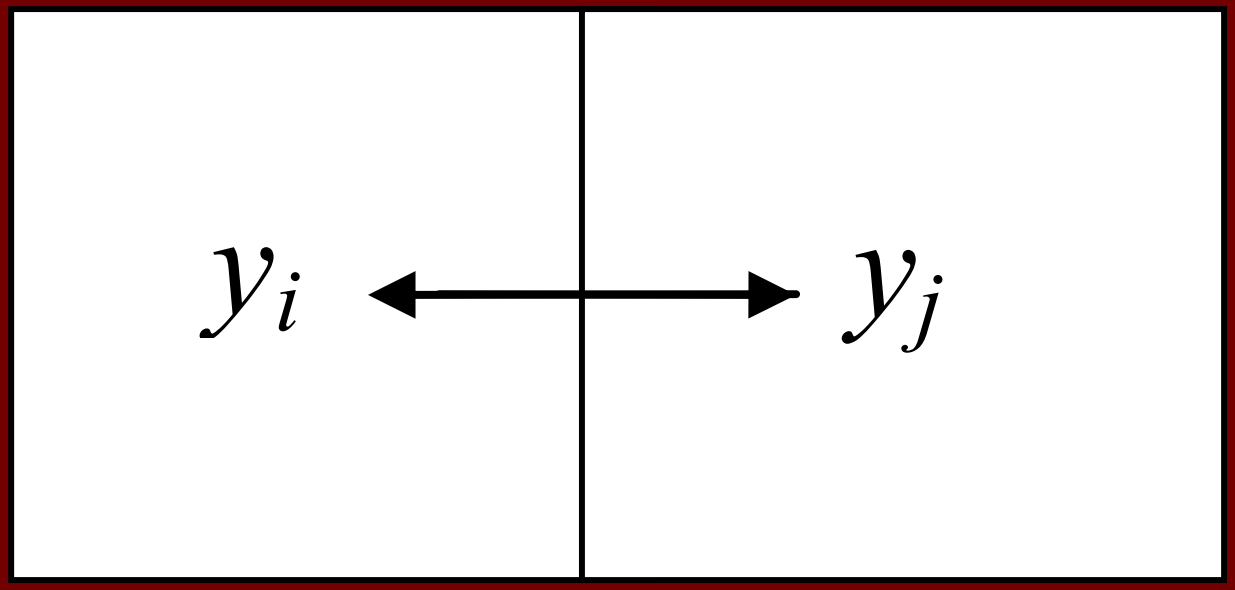
Risks when ignoring space or time

- Ignoring Space:
  - Potential violation of critical assumptions in classical linear models
- Ignoring Time:
  - Failing to understand temporal sequencing of social and spatial patterns means failing to understand causal processes

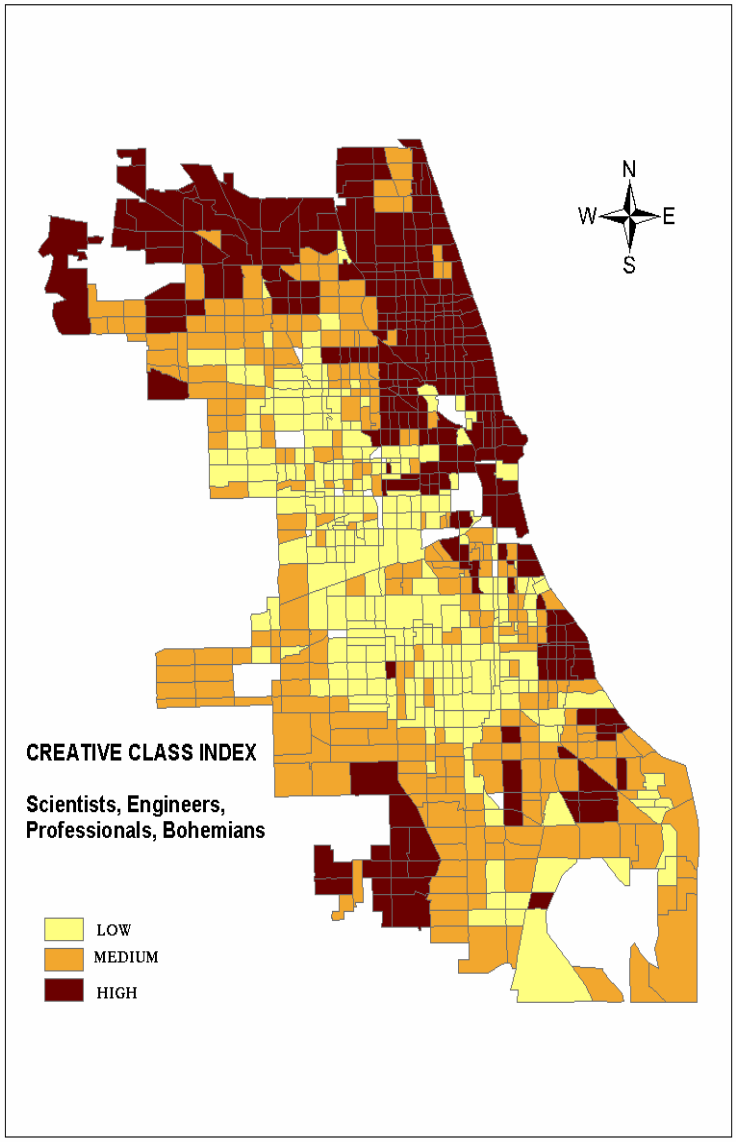
Modeling Space

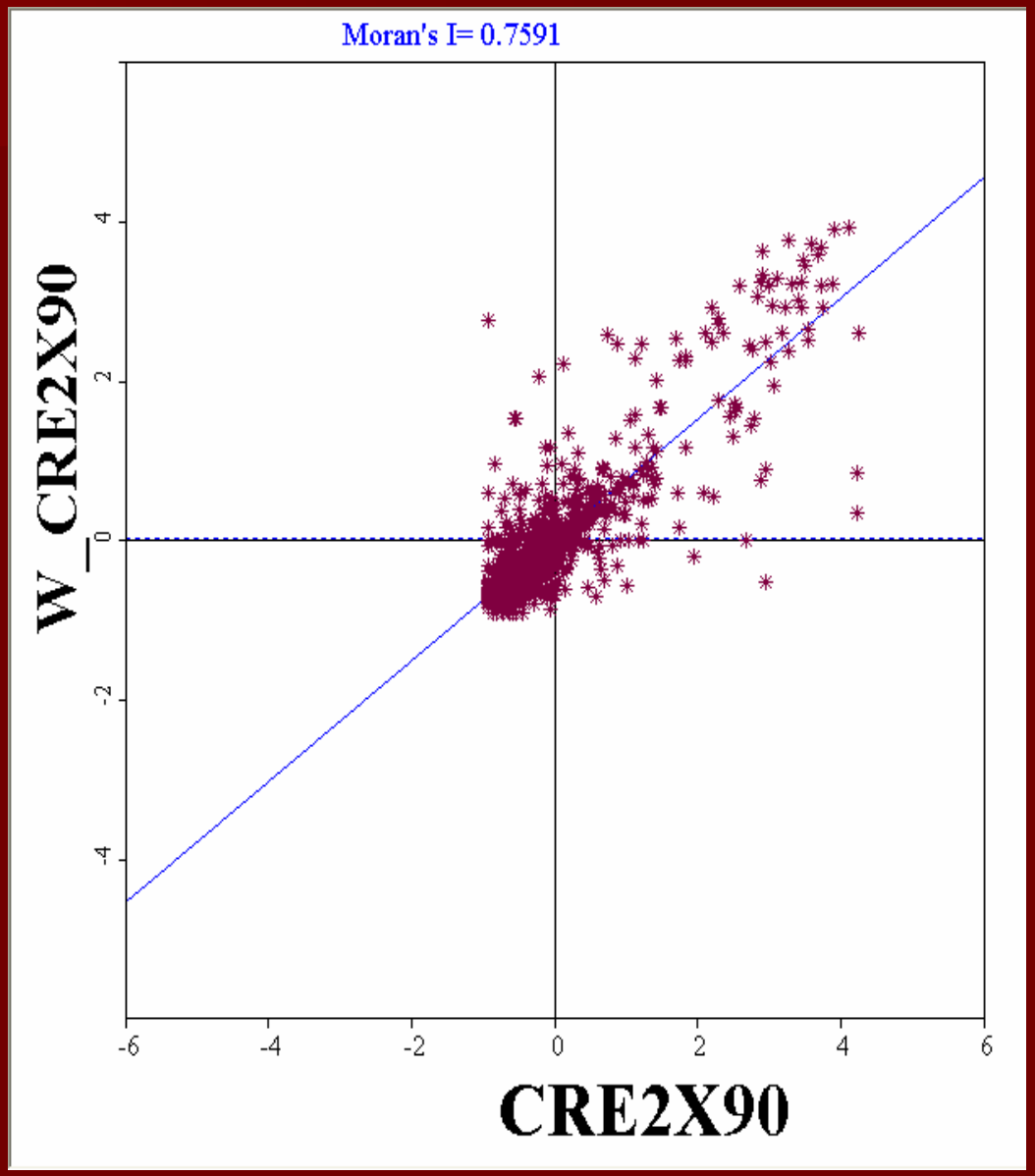
- Requires geographic location information
- Spatial Autocorelation (Univariate)
- Multivariate Global Spatial Models
  - Spatial Error Regression
  - Spatial Lag Regression
- Multivariate Local Spatial Models
  - Geographically Weighted Regression

Spatial Autocorrelation



Moran's I


N
W
E
S
CREATIVE CLASS INDEX
Scientists, Engineers, Professionals, Bohemians
LOW
MEDIUM
HIGH

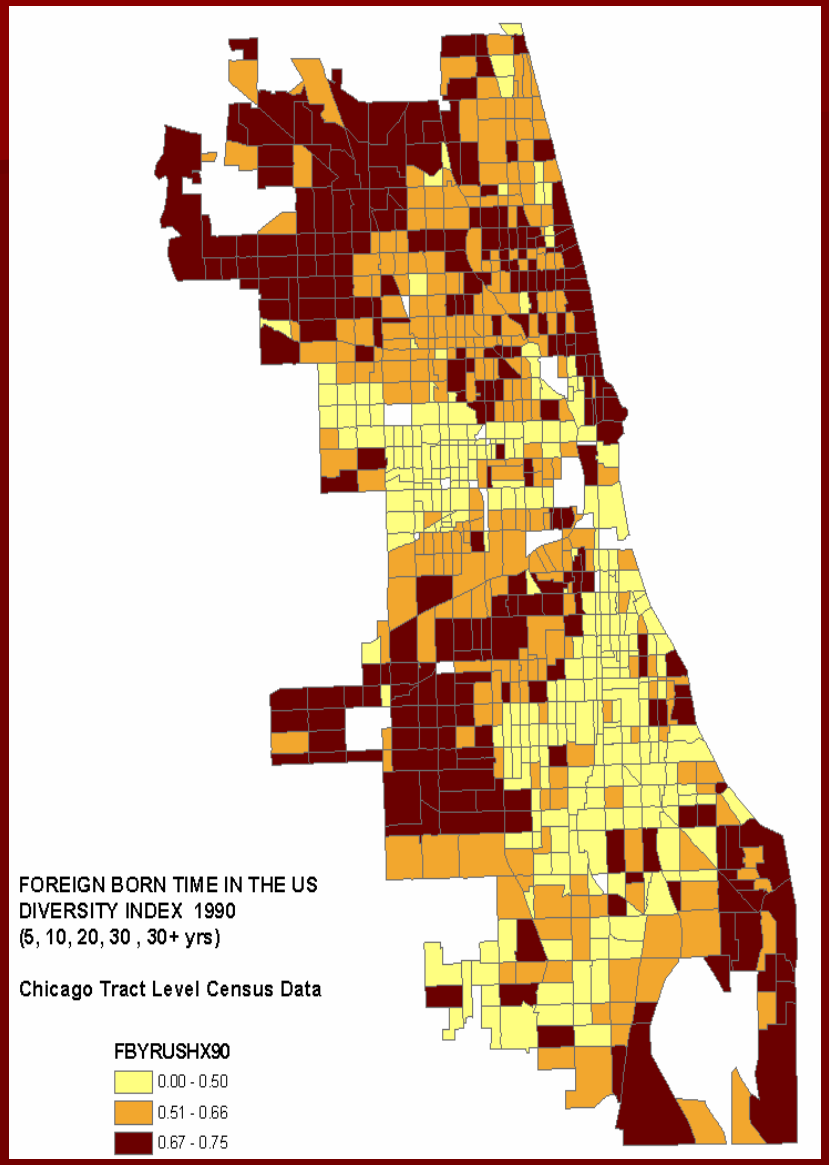

Moran's I= 0.7591
W_CRE2X90
CRE2X90
-6
-4
-2
0
2
4
6

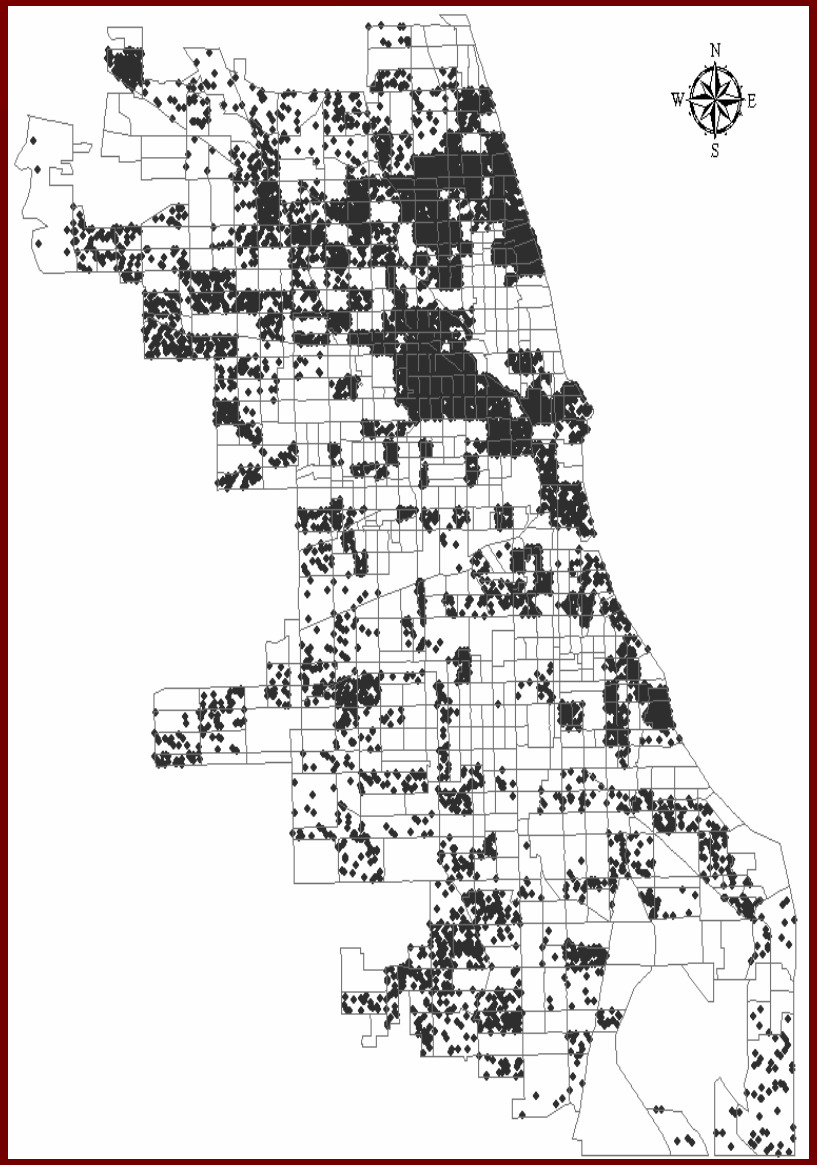
Modeling Spatial Dependence

- Spatial dependence function
  - Distance between units: e.g. rook and queen contiguity criteria, K nearest neighbors, optimal distance bandwidth
- Spatial weight matrix
  - Cell values:
    - binary (if contiguity criterion) or
    - continuous values (if distance decay function)

Spatial Clustering of Foreign Born and Artists

Graif, Corina. "Diversity as a Way of Life" under review (2nd round) at *American Journal of Sociology*.





Modeling Spatial Error

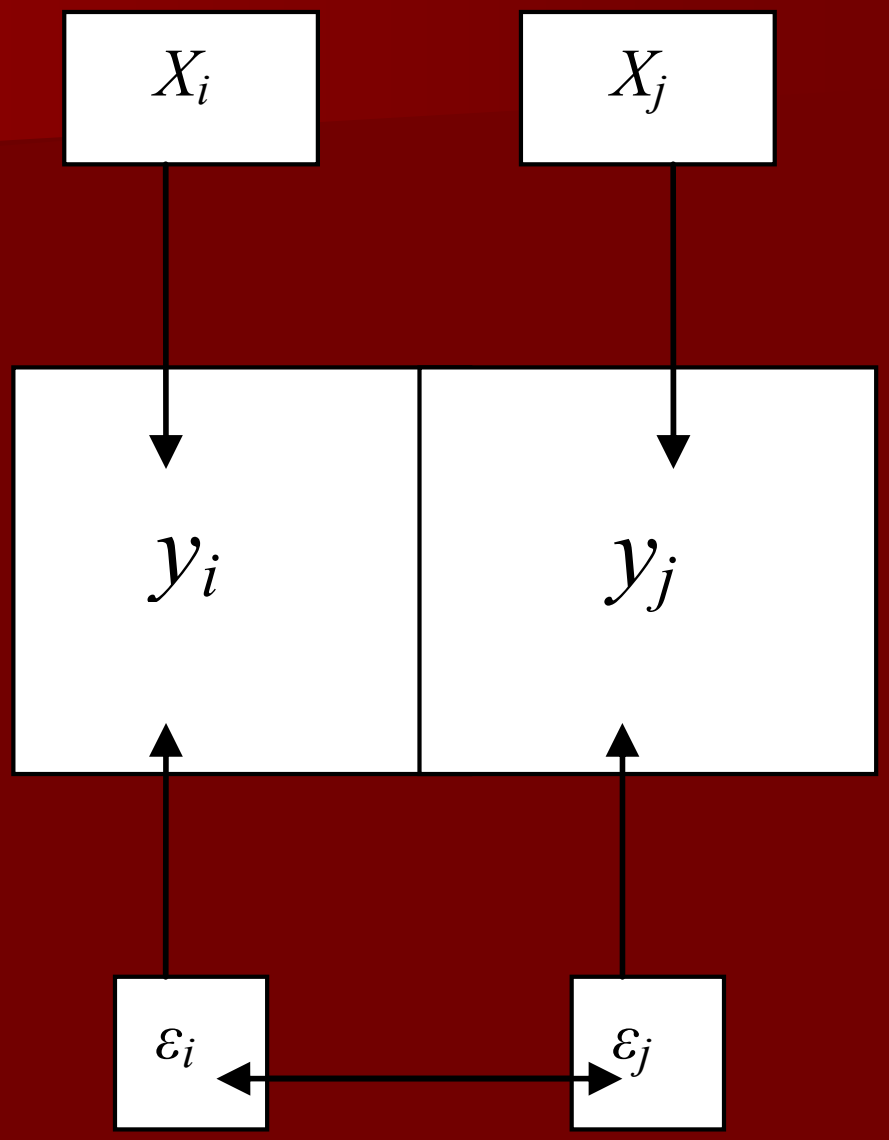
$$\mathbf{y} = \mathbf{X}\,\beta + \varepsilon$$

$$\varepsilon = \rho\mathbf{W}\,\varepsilon + r$$

- Where
  - W is a row standardized spatial weights matrix. The weights are standardized such that $\sum j$ wij =1, for any *i.*
  - β is a vector of regression coefficients and
  - r is a (N × 1) vector of random errors.
  - ρ is the spatial autoregressive parameter

- Consistent with contagion processes
- Risks when ignoring spatial error
  - omitted variables bias

Spatial Error Effects



Modeling Spatial Lag

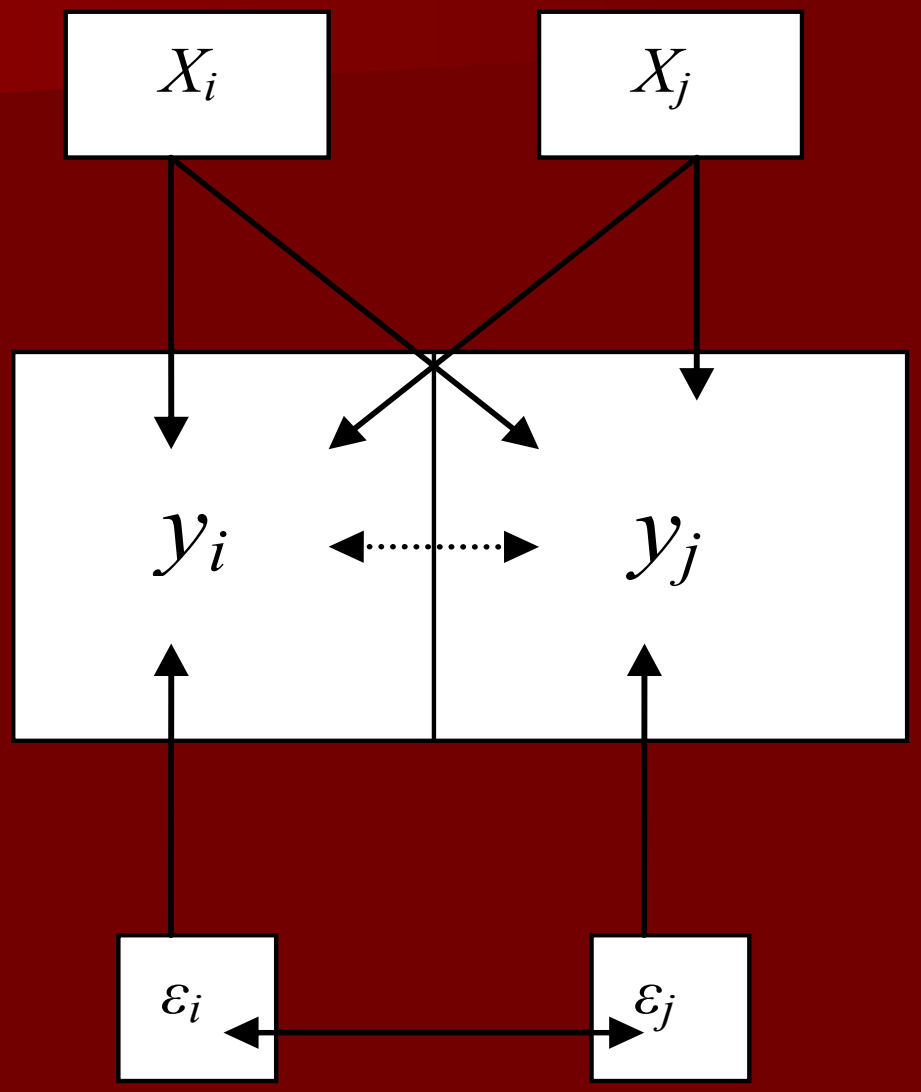
$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\,\beta + r$$

- Where
  - W is a row standardized spatial weights matrix. The weights are standardized such that $\sum_j w_{ij} = 1$, for any *i*.
  - β is a vector of regression coefficients and
  - r is a $(N \times 1)$ vector of random errors.
  - ρ is the spatial autoregression parameter

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\,\mathbf{X}\,\beta + (\mathbf{I} - \rho\mathbf{W})^{-1}\,r$$

- In a temporal context, spatial lag may represent <u>diffusion processes,</u> when events that occur in one place increase the chance of similar events occurring in contiguous places
  - e.g. gang shootings in retaliation for earlier shootings

Spatial Lag Effects



Estimating Resident-Based Social Capital Dimensions on Neighborhood Structural Indices across all Chicago Community Areas (N=77)

	COLLECTIVE EFFICACY		ORGANIZATIONAL INVOLVEMENT		LOCAL NETWORKS		CONDUCT NORMS	
Disadvantage	-.791 ***	-.894 ***	-.298 *	-.315 *	.045	.012	.030	.035
	(.096)	*(.070)*	*(.153)*	*(.133)*	*(.130)*	*(.112)*	*(.151)*	*(.124)*
Residential Stability	.127	-.130	.131	.059	.597 ***	.528 ***	.401 **	.518 ***
	(.103)	*(.080)*	*(.162)*	*(.153)*	*(.137)*	*(.129)*	*(.159)*	*(.142)*
Racial/Ethnic Diversity	-.329		-.724		.042		1.505 m	
	(.562)		*(.917)*		*(.779)*		*(.912)*	
Composite Diversity		-.473 ***		-.317 m		-.141		.574 ***
		(.096)		*(.180)*		*(.153)*		*(.169)*
Spatial Dependence	.535 ***	.270 m	.392 **	.395 **	.389 **	.353 *	.359 **	.309 *
	(.117)	*(.148)*	*(.135)*	*(.135)*	*(.136)*	*(.140)*	*(.139)*	*(.144)*
Constant	.088	-.107	.206	-.036	-.017	-.033	-.392	.097
	(.187)	*(.077)*	*(.278)*	*(.166)*	*(.236)*	*(.134)*	*(.273)*	*(.142)*
R-Squared	*.708*	*.750*	*.204*	*.229*	*.439*	*.441*	*.219*	*.288*
Log-Likelihood	-61.6	*-53.4*	*-100.1*	*-98.9*	*-87.6*	*-87.2*	*-100.1*	*-96.2*
AIC	*137.3*	*114.8*	*208.3*	*205.9*	*183.2*	*182.5*	*208.6*	*200.8*

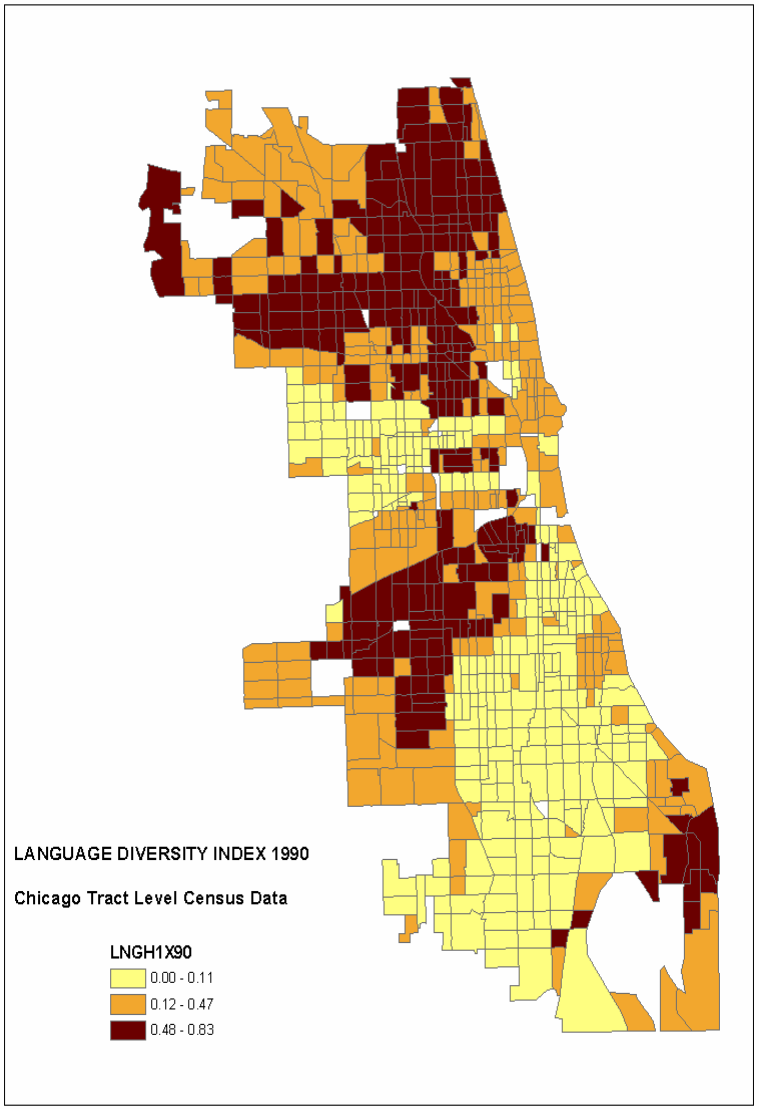
Limitations when Modeling Space

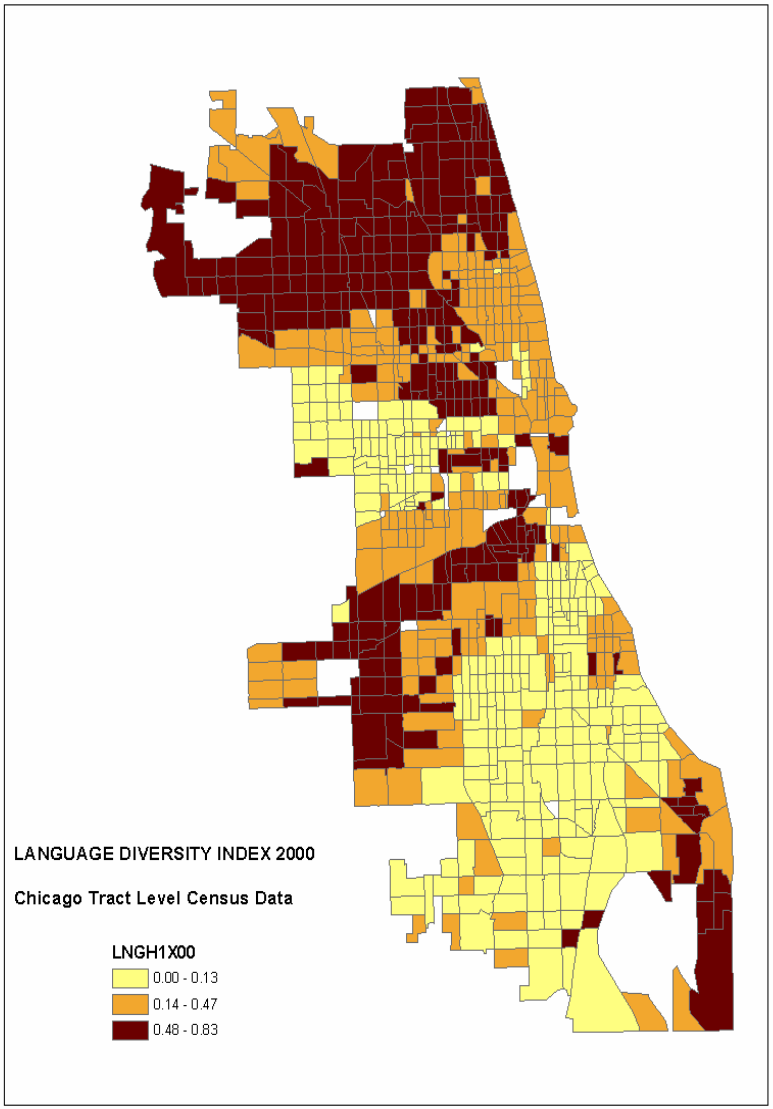
- Estimations require information about spatial proximity of units of analysis
  - (boundary data if working with polygons, or latitude and longitude coordinates if working with points)
  - Understanding projections is no small feat
  - May require serious care about confidentiality (even census data is rounded if too small numbers per data cell)
- Missing data can create important missestimation problems
- Boundary censoring issues (e.g. the city limits)
- Decision about the level of aggregation can be very consequential (need for sensitivity analysis)
- The units of analysis can vary in size and shape

Modeling Time

- Requires panels or time series data
- Serially lagged models
- Change score models
- Residual change models

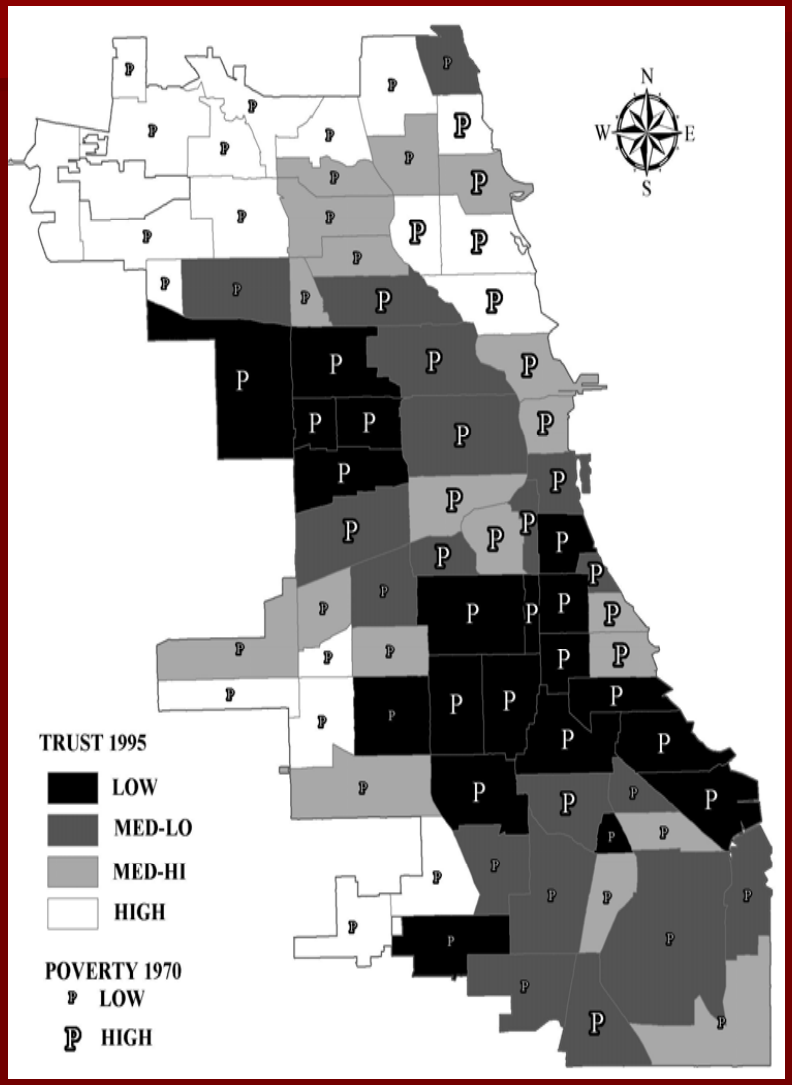
Autocorrelation in Space and Time

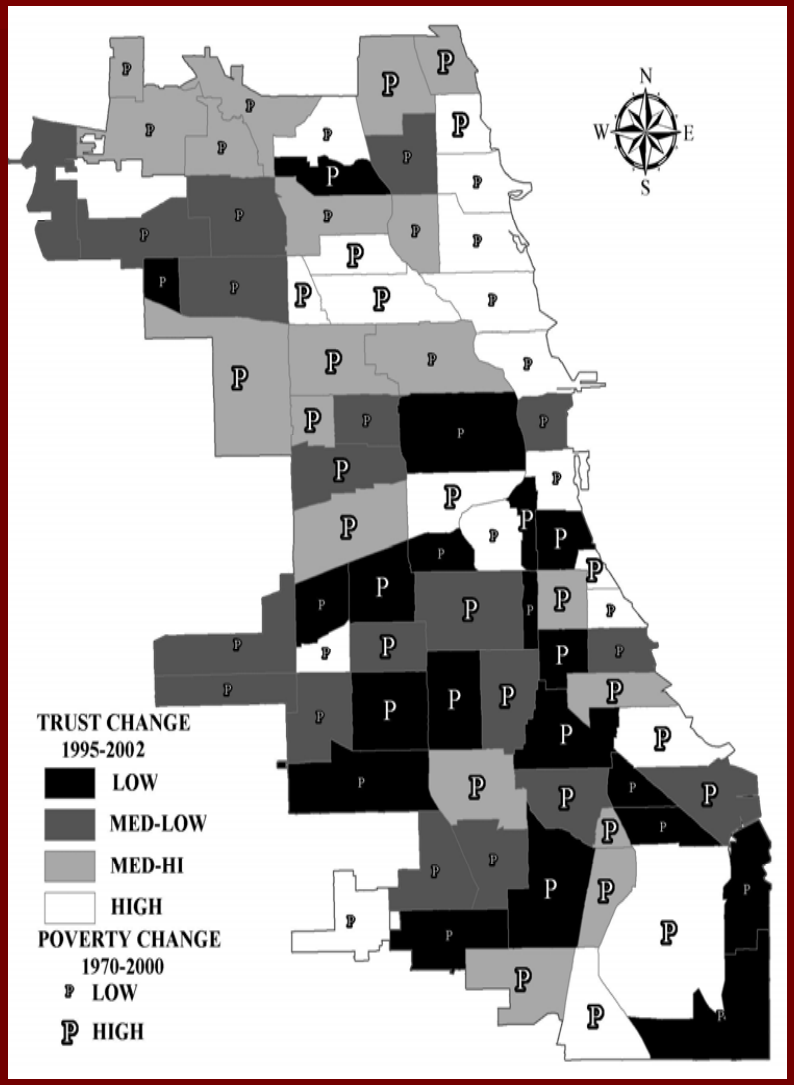




Covariation in Space and Time

Poverty in 1970 by Trust Level in 1995 (left) and
Change in Poverty 1970-2002 by Change in Trust 1995 – 2000 (right)





Modeling Space and Time
Change Models with Spatial and Temporal Lags

$$\Delta y_{90\text{-}00} = \tau y_{90} + \rho W \Delta y_{90\text{-}00} + X_{90}\,\beta_1 + \Delta X_{90\text{-}00}\,\beta_2 + \varepsilon$$

- **Where** $\Delta y_{90\text{-}00}$ represents a Nx1 vector of change scores in y between 1990 and 2000,
- y_{90} represents a Nx1 vector of serially lagged y.
- $W\Delta y_{90\text{-}00}$ represents the spatially lagged change in y from 1990 to 2000.
- ρ is a spatial autoregression coefficient and
- W is a NxN spatial weights matrix, in which the non-diagonal cell values are a function of a neighborhood's first-order spatial contiguity to each of the other neighborhoods, based on commonly shared borders (using the Rook criterion).
- $\Delta X_{90\text{-}00}$ represents a NxK matrix of change scores in the k exogenous predictors between 1990 and 2000.
- X_{90} is a NxK matrix of serially lagged predictors.
- ε is a Nx1 vector of normally distributed, independent, and homoskedastic disturbances.

Modeling Space and Time (continued…)
Residual Change Models

$$\Delta y_{90\text{-}00} = \rho W \Delta y_{90\text{-}00} + X_{90}\,\beta_1 + \Delta X_{90\text{-}00}\,\beta_2 + \varepsilon$$

$$\Delta y_{90\text{-}00} = y_{00} - (\alpha + \beta\, y_{90})$$

- where
  - $\Delta y_{90\text{-}00}$ represents a Nx1 vector of residual change scores in y between 1990 and 2000.
  - $W\Delta y_{90\text{-}00}$ represents the spatially lagged change in y from 1990 to 2000.
  - ρ is a spatial autoregression coefficient and
  - W is a NxN spatial weights matrix in which the non-diagonal cell values are determined as a function of a neighborhood's first-order spatial contiguity to each of the other neighborhoods, based on commonly shared borders (the Rook criterion).
  - $\Delta X_{90\text{-}00}$ represents a NxK matrix of residual change scores in the k exogenous predictors between 1990 and 2000.
  - X_{90} represents a NxK matrix of serially lagged predictors.
  - ε is a Nx1 vector of normally distributed, independent, and homoskedastic disturbances.

Estimating Residents' Collective Efficacy and Change in Collective Efficacy

PHDCN Community and Key Informant Surveys

	Collective Efficacy 2002			Change in Collective Efficacy 1995 - 2002		
	Coeff.	Std.Err		Coeff.	Std.Err	
Constant	1.776	*(.386)*	***	2.383	*(.484)*	***
% Residents in Poverty 1970	-.004	*(.002)*	**	-.005	*(.002)*	**
Change in % Poor 1970-00	-.013	*(.003)*	***	-.012	*(.004)*	***
% Residential Stability	.009	*(.002)*	***	.010	*(.002)*	***
Change % Resid Stability 1970-00	.008	*(.002)*	***	-.001	*(.002)*	
Racial Diversity 1970	-.105	*(.136)*		-.149	*(.143)*	
Change in Racial Diversity 1970-00	.082	*(.091)*		.081	*(.096)*	
1995 Collective Efficacy				-.856	*(.152)*	***
Spatial Lag	.330	*(.117)*	***	.037	*(.145)*	
R-square		*.611*			*.371*	
N		*77*			*77*	

*Notes: *** p<.01, ** p<.05, *p<.10*

Limitations when Modeling Time and Space

- Require repeated measurements at several time points
- Measurement may not be fully equivalent across time
- Higher data collection cost
- Sometimes the time points can be a decade far from each other (e.g. Census Data)
- Unit boundaries may change from a time point to another

Limitations (continued…)

Potential Heterogeneity in Spatial Effects

Geographically Weighted Regression

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

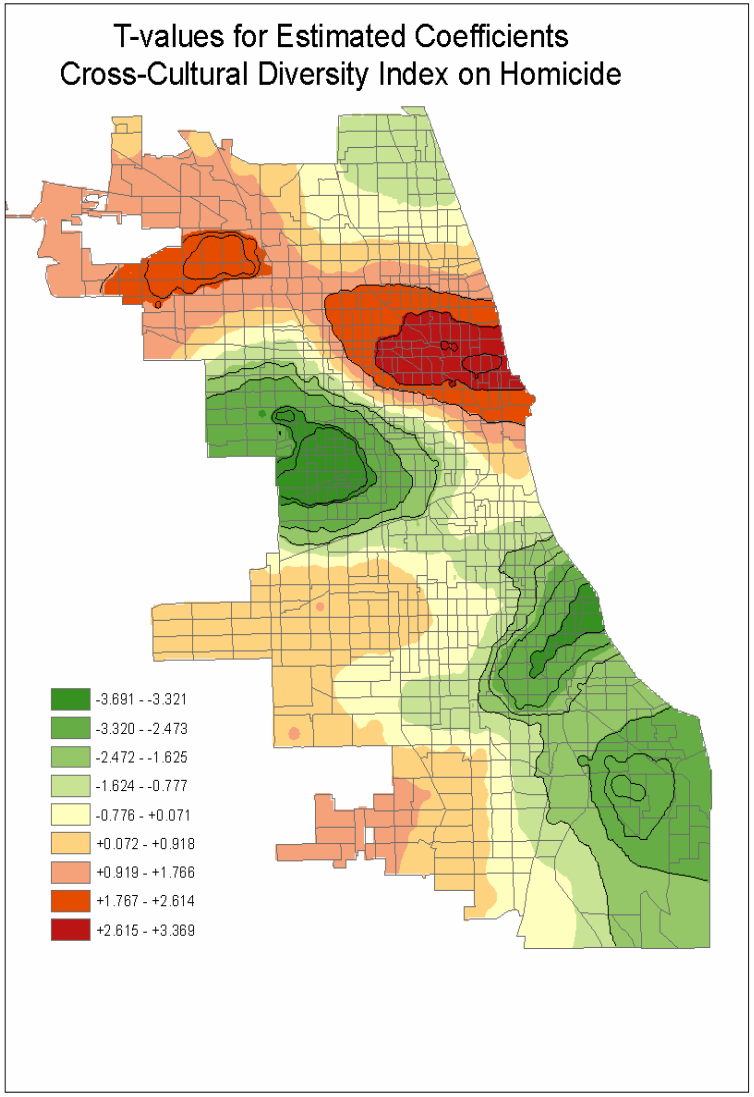
- where
  - (*uivi*) represents the coordinate location of the data point *i* (e.g. in our case, the census tract centroid), and
  - β k(*uivi*) is a realization of the continuous function β k(*uv*) at point i.
- β is estimated as in the following

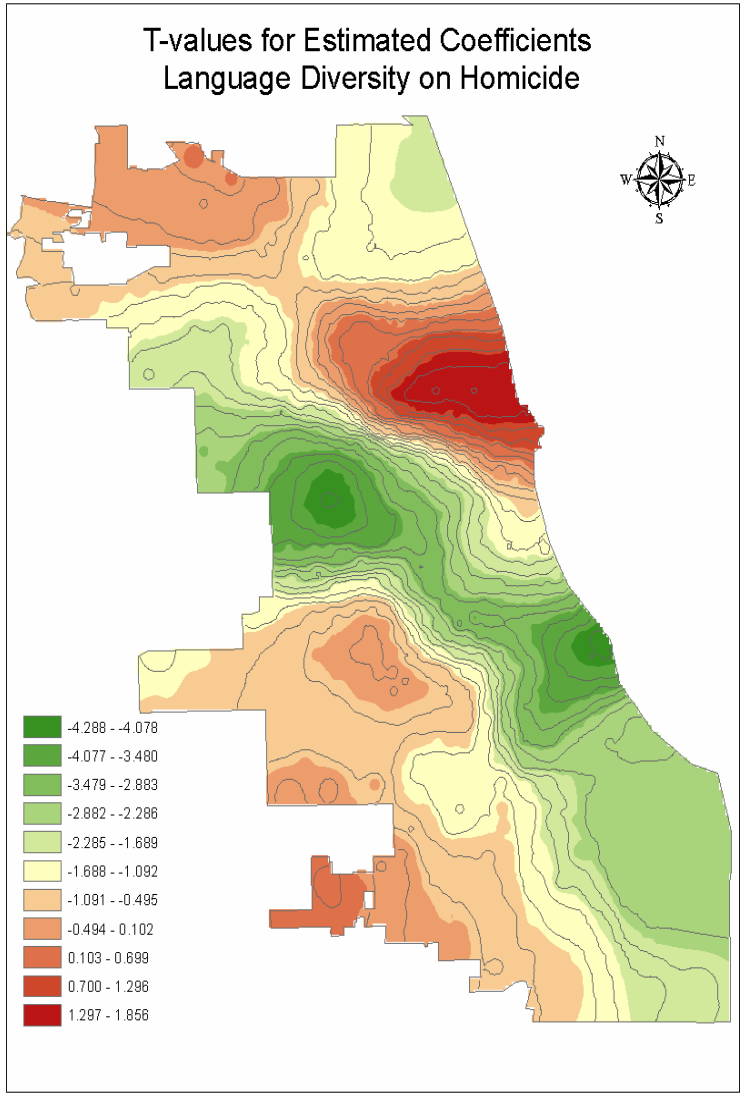
$$\hat{\beta}(u_i, v_i) = \left(\mathrm{X}^T \mathrm{W}(u_i, v_i)\ \mathrm{X}\right)^{-1} \mathrm{X}^T \mathrm{W}(u_i, v_i)\, \mathrm{y}$$

- where
  - W(*ui,vi*) is an n by n matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of observed data for point

Heterogeneity in Spatial Effects

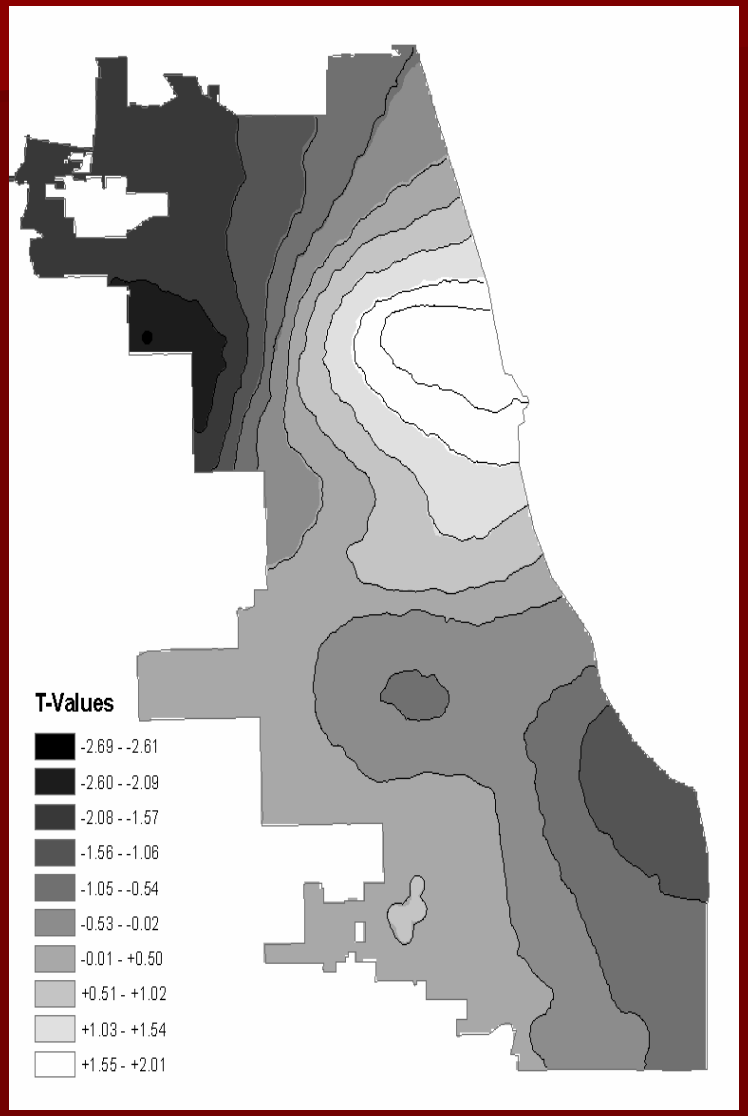
Cross-Cultural Diversity and Language Diversity Predicting Homicide

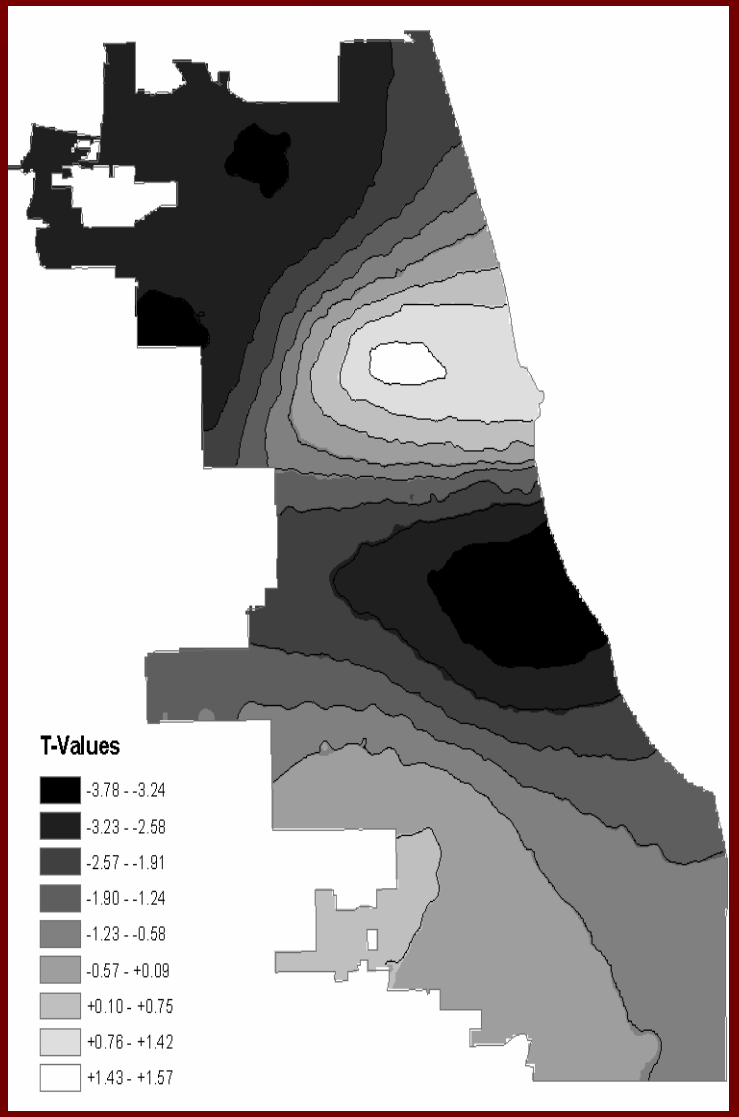




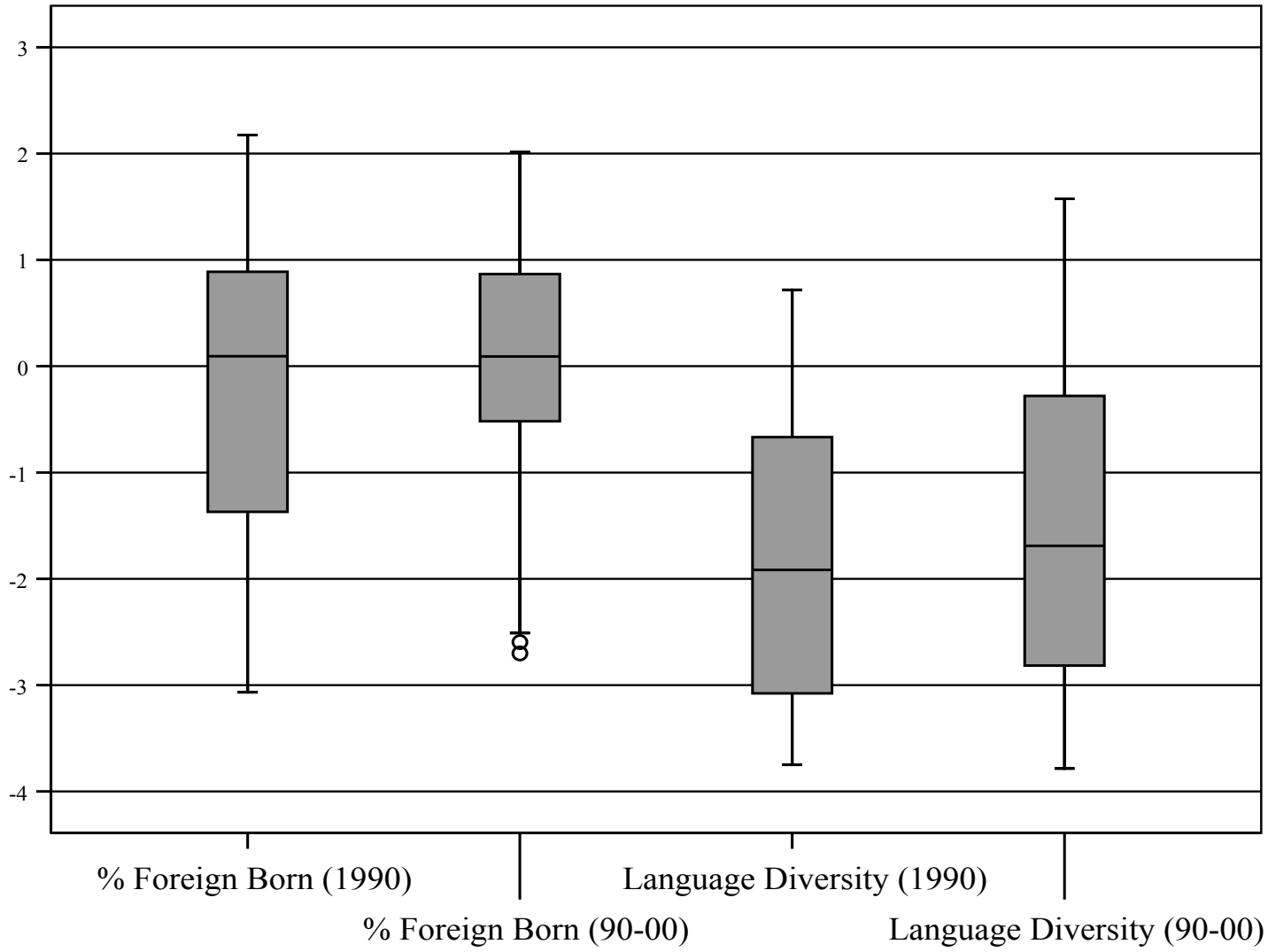
Modeling Spatially Varying Effects in Time

Graif, Corina and Robert J. Sampson. 2009."Spatial Heterogeneity in the Effects of Immigration and Diversity on Neighborhood Homicide Rates." *Homicide Studies* 13(3).

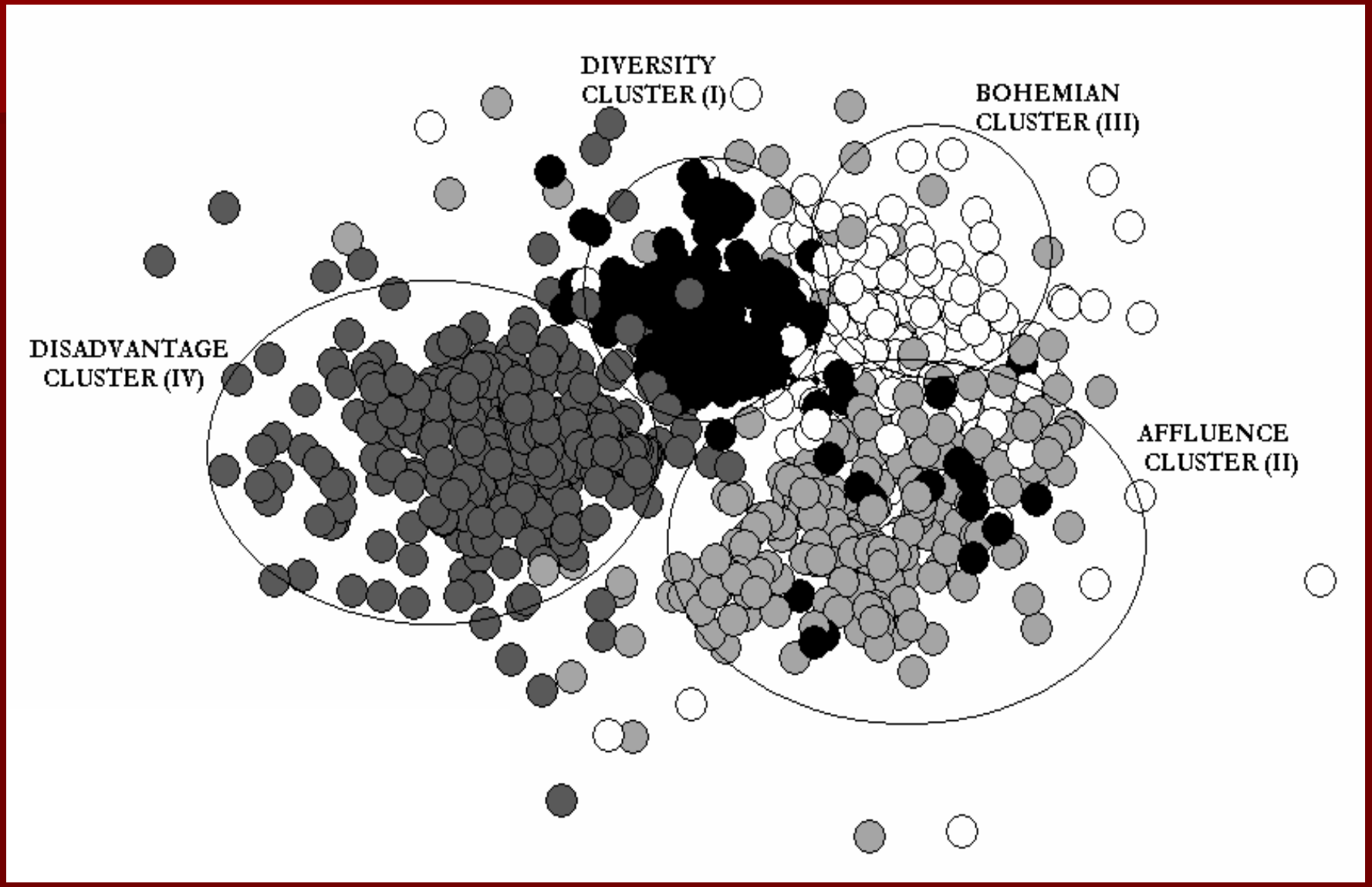




Predicting Increases in Homicide Rates:
Frequency Distribution of T-Values for Estimated Local Parameters

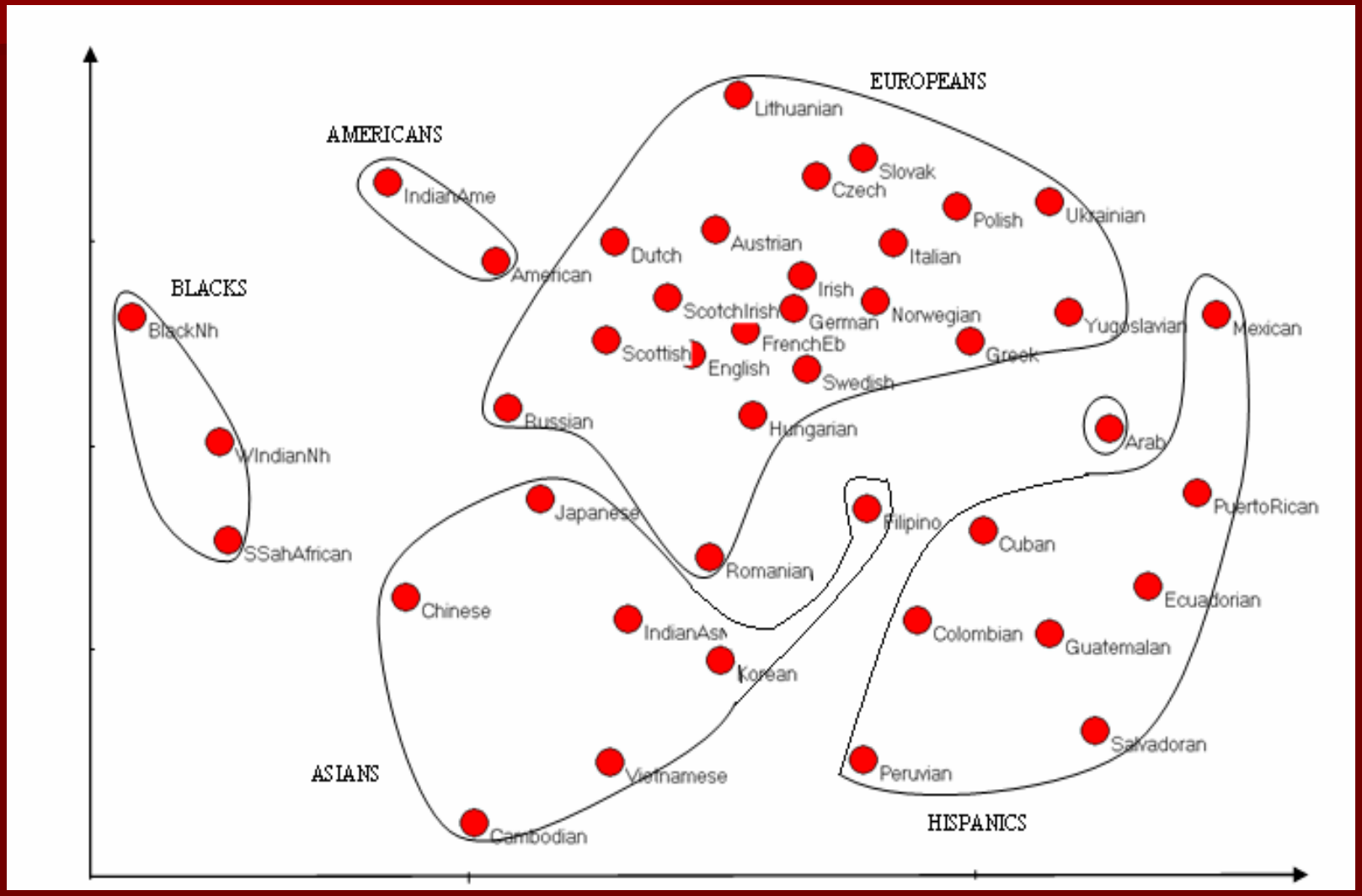


Unconventional Space
Modeling Social Distance as Spatial Distance



"Mapping" Residential Segregation in Non-Geographical Space

Multidimensional Scaling Analysis



Useful Tools

- ARCGIS
- GEODA
- R
- GWR
- STATA
- PAJEK
- UCINET
- MATLAB

Key Substantive Findings

I. Concentrated poverty predicts multiple dimensions of residents' trust some 30 years later

II. Neighborhood diversity attracts an increasing influx of creative class residents (artists, scientists, professionals) over the long run

III. Neighborhood characteristics vary significantly in predicting homicide– with countervailing effects depending on spatial location

IV. Language diversity is consistently related to lower homicide

Methodological Findings Related to Modeling Space and Time

Table A2
DIAGNOSTICS OF SPATIAL DEPENDENCE AND FIT TESTS

		Diagnostics of Spatial Dependence		Regular (OLS) Models			Spatial (ML) Models			Change in Fit (b)	
	Model	MI	LM	R-sq	LL	AIC	R-sq	LL	AIC	ΔLL	ΔAIC
Table 3											
Gays	(1990)	8.46 ***	75.19 ***	.12	-280.29	572.57	.20	-249.50	513.00	-30.79	59.58
	(2000)	3.58 ***	37.25 ***	.19	-240.38	494.76	.24	-223.63	463.26	-16.75	31.50
Bohemians	(1990)	19.97 ***	459.11 ***	.44	-1319.45	2650.90	.69	-1117.95	2249.89	-201.50	401.01
	(2000)	11.27 ***	174.37 ***	.56	-1241.24	2496.47	.65	-1162.21	2340.41	-79.03	156.06
Table 4											
Bohemians	(1)	8.61 ***	47.68 ***	.29	-1166.45	2344.90	.34	-1144.58	2303.16	-21.87	41.74
	(2)	7.73 ***	70.51 ***	.14	-1196.88	2403.77	.23	-1165.57	2343.15	-31.31	60.62
	(3)	6.51 ***	37.30 ***	.37	-1119.76	2269.51	.40	-1102.33	2236.67	-17.43	32.84
	(4)	5.96 ***	51.08 ***	.20	-1165.82	2359.63	.26	-1142.49	2314.98	-23.33	44.65
Creative Class	(1)	10.27 ***	102.79 ***	.42	-291.81	595.62	.50	-245.09	504.18	-46.72	91.44
	(2)	10.10 ***	140.32 ***	.30	-348.90	707.81	.43	-286.05	584.09	-62.86	123.71
	(3)	7.95 ***	67.09 ***	.53	-205.34	440.67	.57	-173.43	378.85	-31.91	61.82
	(4)	7.18 ***	91.35 ***	.44	-256.36	540.72	.51	-213.55	457.09	-42.81	83.63
Table 5											
Scientists	(1)	3.32 ***	4.01 *	.39	-1828.88	3687.77	.39	-1826.86	3685.72	-2.02	2.05
	(2)	2.54 *	1.31	.40	-1818.02	3670.03	.41	-1817.34	3670.68	-.68	-.65
Professionals	(1)	7.53 ***	88.20 ***	.51	-2133.82	4297.63	.57	-2090.47	4212.94	-43.35	84.69
	(2)	5.84 ***	52.76 ***	.62	-2022.12	4078.23	.65	-1996.59	4029.18	-25.53	49.05
Talent Index	(1)	9.71 ***	160.06 ***	.58	-2609.88	5249.76	.67	-2528.10	5088.21	-81.78	161.55
	(2)	6.16 ***	93.28 ***	.66	-2521.51	5077.03	.71	-2473.30	4982.59	-48.21	94.44

NOTE.- (a) MI refers to Moran's I test; LM represents the Lagrange Multiplier for the rho models presented in the Tables 3,4, and 5; LL refers to Log Likelihood, a larger value (closer to zero) indicates better fit; AIC refers to Akaike Information Criterion, a lower value (closer to zero) indicates better fit. (b) ΔLL represents the difference between LL in the OLS model vs. the ML model; a negative value signals an improvement in fit for the ML model; ΔAIC refers to the difference between AIC in the OLS model vs. the ML model; a positive value represents an improvement in fit for the ML model.

Final Thoughts

Models should be as simple as possible … but not simpler

- " the supreme goal of all theory is to make the irreducible basic elements as simple as possible without having to surrender the adequate representation of a single datum of experience"

(Albert Einstein. 1934. "On the Method of Theoretical Physics" *Philosophy of Science* 1(2):163-169, p.165)

Future Directions

- Delving deeper into mechanisms and social processes
  - To what extent the spatial variation in effects of immigration and diversity is mediated by their impact on neighborhood social cohesion
- Modeling spatial contagion and diffusion
- Connecting dynamic spatial analysis with network analysis

References:

- Graif, Corina. "Diversity as a Way of Life*:* From Neighborhood Social Differentiation to Spatial Dynamics of the Creative Class". Revised and resubmitted to *American Journal of Sociology*.

- Graif, Corina and Robert J. Sampson. 2009. "Spatial Heterogeneity in the Effects of Immigration and Diversity on Neighborhood Homicide Rates." *Homicide Studies*. Forthcoming in August 13(3).

- Sampson, Robert J. and Corina Graif. 2009. "Structural and Temporal Contexts of Trust: Durable Social Processes in Chicago Neighborhoods." Forthcoming. *The Complexities and Limits of Trust* edited by Karen Cook, Russell Hardin, and Margaret Levi. New York: Russell Sage Foundation Press.

- Sampson, Robert J. and Corina Graif. 2009. "Neighborhood Social Capital as Differential Social Organization: Resident and Leadership Dimensions." Forthcoming. *American Behavioral Scientist*.