

Spatial Tools for Case Selection Using LISA Statistics to Design Mixed-Methods Research

Imke Harbers ¹ Matthew C. Ingram ²

¹University of Amsterdam

²University at Albany, SUNY

American University June 14, 2019







Two Trends:

 growing emphasis on mixed-methods designs
 growing emphasis on interdependence, including geographic interdependence and on spatial analysis as a way to approach this interdependence

Yet, little attention to mixed-methods research designs with spatially dependent data.

Elsewhere, we have offered two strategies for doing this (Harbers & Ingram 2017, in *Poliltical Analysis*)

Here, we offer two case selection strategies to integrate (a) spatial statistics with (b) qualitative analysis.





Motivation: Analytic Issues

- 4 perspectives on spatial dependence:
 - (1) Benign nuisance
 - know interdependence is out there, but not substantively interested in it and assume no meaningful impact
 - (2) Threat to inference
 - know interdependence is out there, don't have a substantive interest in it, but acknowledge that it undermines valid inferences, so account for it
 - (3) Substantive interest
 - interdependence is a key feature of phenomenon of interest and theory, and want to test effects, e.g., diffusion (theory-testing approach in "Geo-Nested Analysis")
 - (4) Substantive interest, but in theory-building mode (approach in this paper)

UNIVERSITY OF AMSTERDAM





Motivation: Spatial Dependence as Given

Interdependence inheres in social phenomena, and most social science data are likely spatial data.

Outcomes we care about are clustered in space (e.g., voting).









Audience

• Primarily mixed-method researchers, but also quantitative researchers working with spatial data

Premises

- Geography or context as placeholder for variables yet to be uncovered
- Agnostic about reasons for spatial dependence

Running example

• County-level homicide rates in the US (Baller et al. 2001)





Motivati

Table 1: Overview of Options

Option Motivation & Logic I Identify clusters to viss (and maps) of outcome of interest Identify clusters to viss (and maps) of outcome of distributed across space Guiding questions: How does the spatial at (clustering) of an outco interest map onto politi boundaries? What is th appropriate level of an in-depth case studies?, cases that stand out for spatial perspective, for because they defy regic patterns? 2 Visualize spatially une LISA statistics (and maps) of residuals from baseline 2 Visualize spatially une variables and scope co	Analytic Proper malize (1) Selection be (2) Selection be	rrites of Cases raced on LISA consists of a set of sthat form part of r ilusters (high-high; based on similar ependent variable nts y clusters (low- its) y clusters (low- its) to ensure within- to ensure within- to in the dependent vith extreme dicating strong vien neighboring	Falte added of Spatial Approach Logic of selection aligns with mixed method designs where case selection is based on the variable (outcome of interest), but a "case" here consists of connected units to account for spatial dependence of the outcome and to provide leverage for understanding the origins of spatial patterns	Steps Calculate LISA statistics, cluster identifiers, and generate Moran scatter plot and map	Prerequisites Geo- referenced data on outcome of interest	
1 Identify clusters to visu LLS. statistics (and maps) of outcome of interest how an outcome of intu maps onto existing pol jurisdictional boundaries? Guiding questions: How does the spatial at (clustering) of an outco interest map onto politi boundaries? What is th appropriate level of an in-depth case studies?, cases that stand out for spatial perspective, for because they defy regis patterns? 2 Visualize spatially ume indepth case studies?, cases that stand out for spatial perspective, for because they defy regis patterns? 2 Visualize spatially une identify high-high and clusters to uncover om variables and scope co	ualizze (1) Selection be clusters: a cuse clusters: a cus	ased on LISA consists of a set of that form part of r 'lusters (high-high; based on similar ependent variable uts y clusters (low-): :: used on LISA stical significance: of two or more with externe dicating strong ween neighboring	Approach Logic of selection aligns with mixed method designs where case selection is based on the variable (outcome of interest), but a"case" here consists of connected units to account for spatial dependence of the outcome and to provide leverage for understanding the origins of spatial patterns	Calculate LISA statistics, cluster identifiers, and generate Moran scatter plot and map	Geo- referenced data on outcome of interest	
LSA statistics how and outcome of init maps onto existing pol- outcome of interest Guideng meetion: Guideng meetions: Guideng meetions: How does the spatial (clussering) of an outco- interest meetions: How does the spatial (clussering) of an outco- interest meetions: How does the spatial (clussering) of an outco- tion outco- tion outco- tion outco- tion outco- spatial perspective. For because they defy regio patterns? 2 Visualize spatially une performance of a mode clusters to uncover of a mode clusters to uncover of a mode variables and scope co	(1) Mixers: a case connected units connected units e For similarity c low-low): For similarity c low-low): For similarity c low-low): For similarity c low-low): For similarity c low-low): For similarity c values on the d in connected units linstance variable (2) Selection he values and stat A case consists connected units LISA values in association between variable	consists of a set of that form part of r r clusters (high-high; based on similar ependent variable uits y clusters (low-); to ensure within- on the dependent vite of <i>USA</i> <i>statical significance</i> : of two or more with externe dicating strong ween neighboring	with mixed method gets designs where case selection is based on the valled of the dependent variable (outcome of interest), but a "case" here consists of connected units to account for spatial dependence of the outcome and to provide leverage for understanding the origins of spatial patterns	LISA statistics, cluster identifiers, and generate Moran scatter plot and map	referenced data on outcome of interest	
2 Visualize spatially uncertify 2.1/2.5 Visualize spatially uncertified 2.1/2.5 Visualize spatially uncertified 1.5 Visualize spatia	titteal or e sea and is e for similarity e for similarity e for similarity e for similarity e for dissentiation alysis for nar there onal (2) Selection be variable (2) Select	s that form part of r slusters (high-high; based on similar ependent variable nts y clusters (low- its) to ensure within- to ensure within- to ensure within- to ensure within- to ensure within- to the state of the state of State of the state of State of the state of the state of the dicating strong ween neighboring	designs where case selection is based on the variable (outcome of interest), but a "case" here consists of connected units to account for spatial dependence of the outcome and to provide leverage for understanding the origins of spatial patterns	statistics, cluster identifiers, and generate Moran scatter plot and map	data on outcome of interest	
outcome of interest jurisfictional boundarini distributed across space Guiding questions: How does the spatial (clustering) of an outco interest map onto politi boundaries? What is th appropriate level of an in-depth case studies? cases that stand out for spatial perspective, for because they defy regis patterms? 2 LISA statistics (and maps) of residuals from baseline Visualize spatially une performance of a mode clusters to uncover om variables and scope co	es and is e ssociation association the same cluste iou of the same cluste values on the divide values on the divide values on the divide values on the divide values on t	T clusters (high-high; based on similar ependent variable uts y clusters (low-): to ensure within- on the dependent steed on LISA statical significance: of two or more with extreme dicating strong ween neighboring	selection is based on the value of the dependent variable (outcome of interest), but a "case" here consists of connected units to account for spatial dependence of the outcome and to provide leverage for understanding the origins of spatial patterns	luster identifiers, and generate Moran scatter plot and map	outcome of interest	
interest distributed across space Guiding guestions: How does the spatial a (clustering) of an outce interest map onto possible studies? What is the appropriate level of am in-deptic cases that stand out for spatial perspective, for spatial perspective, for because they defy region patterns? 2 Visualize spatially une level of a mode identify high-high and clusters to uncover om variables and scope cover stores to uncover or spately region and scope cover stores to uncover or spately region and scope cover stores to uncover or spately region and scope cover stores to uncover or spately region and scope cover stores to uncover or spately to the spateline scope cover stores to uncover or spately here to uncover or spately to the scope cover stores to uncover or spately to the scope cover stores to uncover or spateline scope cover stores to uncover or spately to the scope cover stores to uncover or spately to the scope cover stores to uncover or spately to the scope cover stores to uncover or spately to the scope cover stores to uncover or spately to the scope cover store to the scope cover stores to uncover or spately to the scope cover stores to uncover or spately to the scope cover stores to uncover or spately to the scope cover stores to uncover storesto uncover stores to uncover stores to uncover store	e for similarity c low-low): sociation ome of alysis for and the set of alysis for and the set of alysis for and the set of and the set of an	tusters (high-high; based on similar ependent variable hts y clusters (low- to ensure within- to ensure within- on the dependent <i>used on LISA</i> <i>stitual significance:</i> of two or more with extreme dicating strong ween neighboring	value of the dependent variable (outcome of interest), but a "case" here consists of connected units to account for spatial dependence of the outcome and to provide leverage for understanding the origins of spatial patterns	identifiers, and generate Moran scatter plot and map	interest	
2 Visualize spatially une performance of a motor interest map onto politi boundaries? What is th appropriate level of am in depth cases studies?, cases that stand out for spatial perspective, for because they defy regis patterms? 2 Visualize spatially une performance of a mode identify high-high and clusters to uncover wariables and scope cor wariables and scope cor	For similarity v low-low): Cases selectedl values on the d values on the d instance onal (2) Selection be values and stat A case consists connected units (2) Selection be values and stat A case consists connected units LISA values in association between the selectedl sector of the sector of the sector of the sector of the sector sector of the sector o	lusters (high-high; based on similar ependent variable its y clusters (low-): to ensure within- on the dependent teed on LISA statical significance: of two or more with extreme dicating strong ween neighboring	variable (outcome of interest), but a "case" here consists of connected units to account for spatial dependence of the outcome and to provide leverage for understanding the origins of spatial patterns	and generate Moran scatter plot and map		
2 Visualize spatially une LISA statistics (and maps) of residuals for baseline tables and scope con- variables and s	units					
LISA statistics performance of a mode (and maps) of identify high-high and residuals from baseline variables and scope con	ven Selection based	on LISA clusters:	Logic of selection aligns	Estimate	Geo-	
(and maps) of identify high-high and residuals from clusters to uncover om variables and scope con	el: a case consists	of a set of	with mixed method	baseline	referenced	
residuals from baseline variables and scope con	low-low connected units	that form part of	designs where case	regression	data on	
baseline variables and scope con	itted the same cluste	r .	selection is based on	model;	outcome of	
d-1	nditions		residuals, but a "case"	calculate	interest and	
model			here consists of connected	LISA statistics	predictors for	
Guiding questions: How does the model pe across space? Which re the study area are well predicted? Which are p predicted? Are there cl over- and under-predict these clusters map onto boundaries?	For similarity c low-low): gions of Cases selected l of residuals and orly usters of political	lusters (high-high; based on clustering I model fit	units to account for the spatial structure of the error term; this facilitates access to case knowledge to think through what may be missing from the model	for residuals, cluster identifiers, and generate map	baseline, non- spatial model (e.g. OLS)	
VERSITY OF AMSTERD						

◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ─ □ ─ つへぐ

University of Amsterdam



LISA statistics for outcome of interest.

Identify clusters to visualize how an outcome of interest is distributed geographically, including how it maps onto existing boundaries, e.g., administrative, political, jurisdictional.

Guiding questions:

Is there clustering? How does the spatial association map onto political boundaries? What is the appropriate level of analysis for in-depth case studies? Do these patterns suggest scope conditions? Are there sites that stand out for one reason or another?















LISA Values (DV: HR90)



LISA p-values (DV: HR90)



LISA Clusters (DV: HR90)



Homicide Rates (HR90)



LISA Values, positive (DV: HR90)



LISA Clusters (DV: HR90)





60

40

20

0

10.0 7.5

5.0

2.5

LISA p-values (DV: HR90)



LISA Values, negative (DV: HR90)



LISA Clusters (DV: HR90)



Conclusions

Table 2. Focal units selected based on LISA clusters							
High-high	1						
FIPSNO	County	State	HR90	Wy	LISA	р	
28055	Issaquena	Mississippi	34.92	23.06	10.397	0.002	
22035	East Carroll	Louisiana	37.77	19.22	9.335	0.002	
28151	Washington	Mississippi	30.42	23.12	9.286	0.002	
28125	Sharkey	Mississippi	28.30	22.80	8.334	0.002	
28083	Leflore	Mississippi	27.67	20.62	6.975	0.002	
Low-low							
FIPSNO	County	State	HR90	Wy	LISA	р	
20007	Barber	Kansas	0	0.32	0.867	0.002	
20109	Logan	Kansas	0	0	0.867	0.002	
31101	Keith	Nebraska	0	0.09	0.867	0.002	
31135	Perkins	Nebraska	0	0.08	0.867	0.002	
46045	Edmunds	South Dakota	0	0.43	0.854	0.002	
Low-high							
FIPSNO	County	State	HR90	Wy	LISA	р	
13183	Long	Georgia	0	13.08	-1.054	0.01	
8007	Archuleta	Colorado	0	11.62	-1.013	0.008	
48365	Panola	Texas	0	12.65	-0.947	0.018	
21109	Jackson	Kentucky	0	12.72	-0.917	0.012	
48271	Kinney	Texas	0	12.63	-0.904	0.016	
High-low							
FIPSNO	County	State	HR90	Wy	LISA	р	
8053	Hinsdale	Colorado	71.38	3.58	-6.83	0.004	
46063	Harding	South Dakota	19.97	0.29	-1.832	0.002	
31173	Thurston	Nebraska	19.22	1.15	-1.455	0.004	
17035	Cumberland	Illinois	18.74	1.84	-1.282	0.002	
41023	Grant	Oregon	16.98	0.99	-1.253	0.002	





Assume have baseline, non-spatial model (e.g., OLS) and have extracted component that remains unexplained (residuals) **LISA statistics (and maps) of residuals**.

Visualize spatially uneven performance of a model; identify high-high and low-low clusters to uncover omitted variables and scope conditions.

Guiding questions:

How does the model perform across space? Which regions of the study area are well predicted? Which are poorly predicted? Are there clusters of over- and under-prediction? Do these clusters map onto political boundaries?











LISA Values, positive (residuals)



LISA p-values (residuals)



LISA Clusters (residuals)





Residuals 60

40

20

LISA values

6

4

2

LISA p 0.15 0.10









-4

Option 1 000000 Option 2

Conclusions

High-higł	1					
FIPSNO	County	State	e	We	LISA	р
11001	Washington	D.C.	43.496	4.731	6.799	0.032
49031	Piute	Utah	27.718	3.024	4.334	0.042
24510	Baltimore City	Maryland	13.898	6.426	3.719	0.006
24033	Prince Georges	Maryland	12.395	7.178	3.650	0.004
48435	Sutton	Texas	20.895	2.700	3.501	0.032
Low-low						
FIPSNO	County	State	e	We	LISA	р
1105	Perry	Alabama	-19.102	-4.249	2.840	0.004
28157	Wilkinson	Mississippi	-12.286	-5.344	2.650	0.002
1065	Hale	Alabama	-15.098	-4.185	2.462	0.002
22009	Avoyelles	Louisiana	-8.658	-4.968	1.902	0.002
22029	Concordia	Louisiana	-8.993	-5.418	1.899	0.002
Low-high						
FIPSNO	County	State	e	We	LISA	р
47159	Smith	Tennessee	-0.135	3.257	-0.021	0.038
45047	Greenwood	South Carolina	-0.133	4.118	-0.022	0.016
47111	Macon	Tennessee	-0.295	3.448	-0.046	0.044
13317	Wilkes	Georgia	-0.392	4.681	-0.065	0.028
51700	Newport News	Virginia	-0.497	3.273	-0.068	0.046
High-low						
FIPSNO	County	State	e	We	LISA	р
26051	Gladwin	Michigan	0.021	-2.566	-0.003	0.016
26117	Montcalm	Michigan	0.059	-2.525	-0.006	0.040
39103	Medina	Ohio	0.101	-2.617	-0.011	0.036
29217	Vernon	Missouri	0.126	-3.085	-0.015	0.028
46123	Tripp	South Dakota	0.182	-3.539	-0.027	0.006





Conclusions

Tools from spatial analysis can provide additional leverage for case selection.

- 1 Identify scope conditions
- 2 Clarify bound or unbound nature of phenomena
- 3 Examine causal mechanisms
- 4 Identify new, previously omitted variables, to generate new hypotheses and build theory

Core implication across all proposed strategies:

- need to redefine meaning of "case" as more than 1 unit of observation





М	0	tiv		ti		
	2		-		0	
0	0		O			

Overview o

Option 1 000000 Option 2

Conclusions

Thank You

University of Amsterdam





Spatial Tools for Case Selection Using LISA Statistics to Design Mixed-Methods Research

Imke Harbers ¹ Matthew C. Ingram ²

¹University of Amsterdam

²University at Albany, SUNY

American University June 14, 2019



