Spatial Economics – Assignment 2

Gustav Pirich (h11910449) Peter Prlleshi () Filip Lukijanovic ()

April 2, 2024

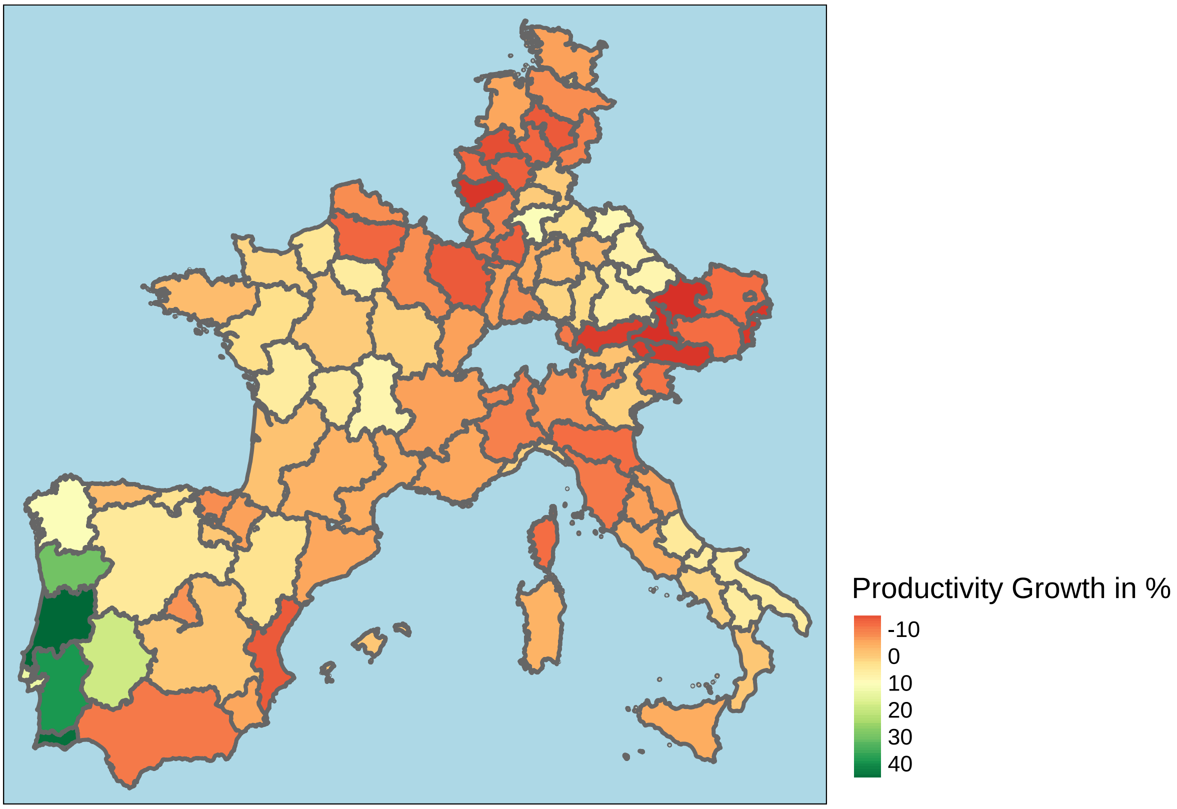
Contents

The code that was used in compiling the assignment is available on GitHub at https://github.com/gustavpirich/spatial_econ/blob/main/02_assignment/02_assignmnet.Rmd.

Exercise A

Calculate the growth rate of productivity from 1980 to 2013 and create a map that shows the productivity growth for each region.

The map shows the productivity growth rates in NUTS-2 regions for the selected countries. We can see that many regions especially in West Germany, Austria, and France exhibited negative productivity growth over the selected period. Notably, Portugal's productivity has been growing the fastest. We suspect that the negative growth rates can be explained by the fact that high-income countries had a high baseline productivity to begin with, while Portugal started from a rather low baseline productivity. Thus we can interpret this as productivity convergence across Europe.



Generate three different spatial weights matrixes using (i) a distance threshold, (ii) smooth distance-decay, and iii) a contiguity-based measure.

(i) Distance Threshold

First, we create a binary distance threshold spatial weights matrix based. Any region is being assigned a '1' with respect to another region, if it is less than 3 km away. We have chosen this threshold so that every region has a neighbor. We use the nb2mat function from the 'spdep' package. We row-normalize the matrix.

```
coords <- st_coordinates(st_centroid(EU27))

# checking the maximum distance as to include all observations which have a
# matrix
nb1 <- knn2nb(knearneigh(coords, k = 1))

dist1 <- nbdists(nb1, coords)

distw <- dneargneigh(coords, 0, 3)

# creating matrix based on distance threshold up to 3 kilometers
dist_w_matrix <- nb2mat(distw, style = "W", zero.policy = TRUE)
```

Table 1: Summary of Distance Threshold Graph

Property	Value
Number of vertices	103
Number of edges	514
Average path length	0.6750662
Graph density	0.09784885
Average degree	9.980583
Max Eigenvector Centrality	1.00
Min Eigenvector Centrality	0.008694479
Average Eigenvector Centrality	0.1497447
Most Central Unit (Vertex ID)	49

(ii) Smooth-Distance Decay

Next, we create a spatial weights matrix based on a smooth distance-decay. We use the following simple distance decay function $w_{i,j} = 1/d^{\lambda}_{i,j}$, where d denotes the distance between observation i and j, and λ is the distance decay parameter. By ease of convention we set $\lambda = 0$. We calculate the weights for each neighboring region based on the k=20 nearest neighbors. We do *not* row-normalize the matrix.

```
k1 <- knearneigh(coords, k = 20)
k2 <- knn2nb(k1)

dists <- nbdists(k2, coords)

ids <- lapply(dists, function(d) {
    1/d
})

decay_weights_matrix_list <- nb2listw(k2, glist = ids, style = "B", zero.policy = TRUE)

decay_weights_matrix <- listw2mat(decay_weights_matrix_list)
```

(iii) Contiguity-based measure

Finally, we calculate a contiguity based measure, which we row normalize as well.

```
# Create a contiguity-based spatial weights matrix
queen_weights <- poly2nb(EU27, queen = TRUE)

contig_w_matrix <- nb2mat(queen_weights, style = "W", zero.policy = TRUE)
```

Compare the matrices; use your knowledge of graph theory and linear algebra

We can gain deeper insights into these spatial weights matrices as well as the networks they represent by comparing key measures of the graphs that are derived from them.

We compare the matrices based on a set of characteristics. We compare the row-normalized matrices for the queen contiguity and distance threshold matrix. Note that normalization procedure does not preserve the structure of the network.

Number of edges

The Smooth Distance-Decay Graph has the most edges (1266). The Distance Threshold Graph has fewer edges (514) than the Smooth Distance-Decay Graph, implying stricter criteria for edge creation based on a fixed distance threshold. The Contiguity-Based Graph has the fewest edges (222), since only directly contiguous or neighboring entities are connected, leading to a more sparse graph structure.

Average path length

The Contiguity-Based Graph has the highest average path length (1.2888), reflecting the sparse connectivity where nodes are less directly connected. The Distance Threshold Graph has a medium average path length

Table 2: Summary of Smooth Distance-Decay Matrix

Property	Value
Number of vertices	103
Number of edges	1266
Average path length	0.49328
Graph density	0.2410051
Average degree	24.58252
Max Eigenvector Centrality	1
Min Eigenvector Centrality	0.001002582
Average Eigenvector Centrality	0.2836761
Most Central Unit (Vertex ID)	23

Table 3: Summary of Contiguity-Based Graph

Property	Value
Number of vertices	103
Number of edges	222
Average path length	1.288804
Graph density	0.04226156
Average degree	4.31068
Max Eigenvector Centrality	1.00
Min Eigenvector Centrality	3.590996e-18
Average Eigenvector Centrality	0.08975192
Most Central Unit (Vertex ID)	48

(0.6751). The Smooth Distance-Decay Graph has the lowest average path length (0.4933), indicative of a denser network where nodes are more directly accessible to one another.

Graph density

Consistent with the number of edges, the Smooth Distance-Decay Graph is the densest (0.2410), followed by the Distance Threshold Graph (0.0978), with the Contiguity-Based Graph being the least dense (0.0423).

Average degree

Again, the Smooth Distance-Decay Graph shows the highest average degree (24.5825), the Distance Threshold Graph shows a medium degree (9.9806), and the Contiguity-Based Graph has the lowest (4.3107).

Minimum Eigenvector Centrality

The Contiguity-Based Graph shows the most significant variation in centrality (minimum near zero), reflecting a few very poorly connected nodes, or nodes that only connect to other low-influence nodes. The Smooth Distance-Decay Graph and the Distance Threshold Graph have higher minimum values, indicating a more uniform distribution of node influence.

Average Eigenvector Centrality

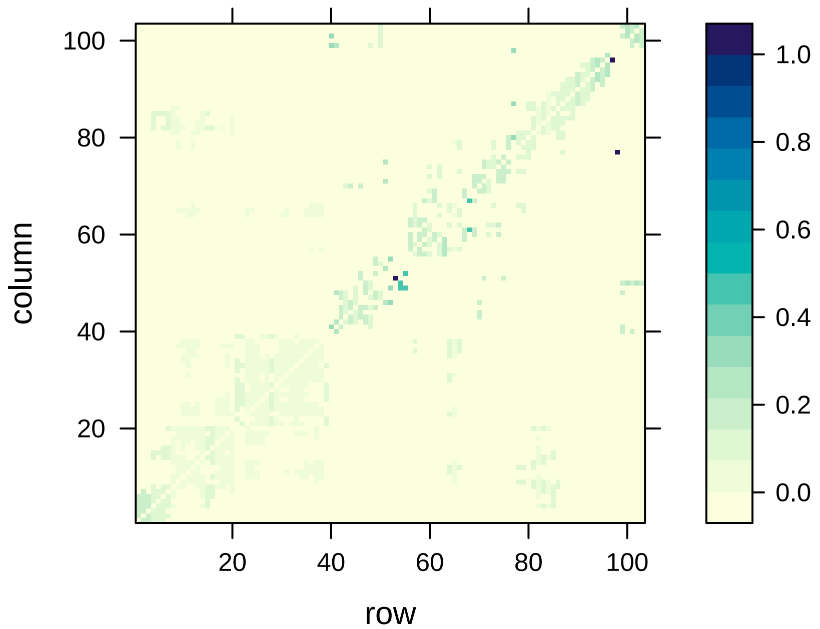
Higher on average in the Smooth Distance-Decay Graph (0.2837), suggesting that, on average, nodes are better positioned or more influential within the network. It's lowest in the Contiguity-Based Graph (0.0898), consistent with its sparse and uneven connectivity.

Looking at the most central unit through eigenvector centrality shows that different nodes are identified as most central in each graph, reflecting the impact of the underlying connection logic on the perceived importance or centrality of nodes.

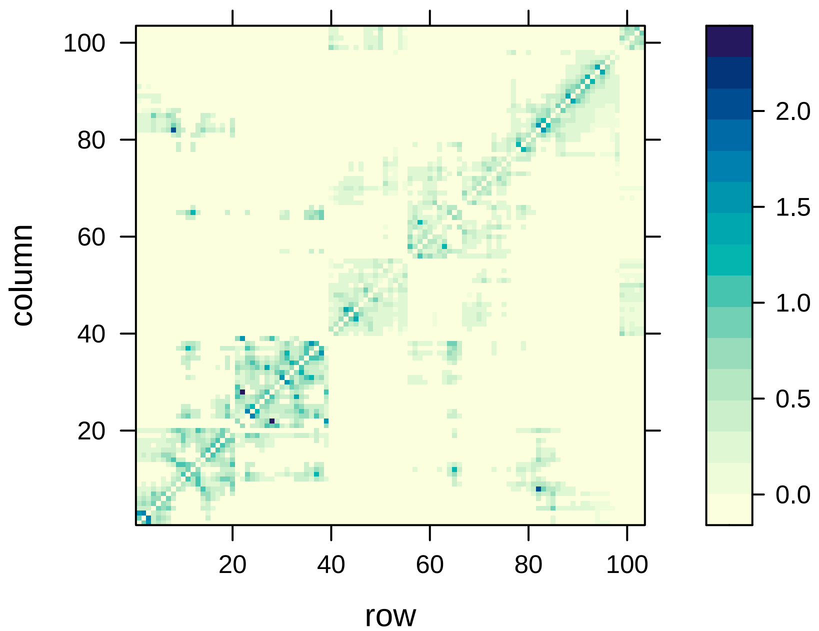
Plot the matrix

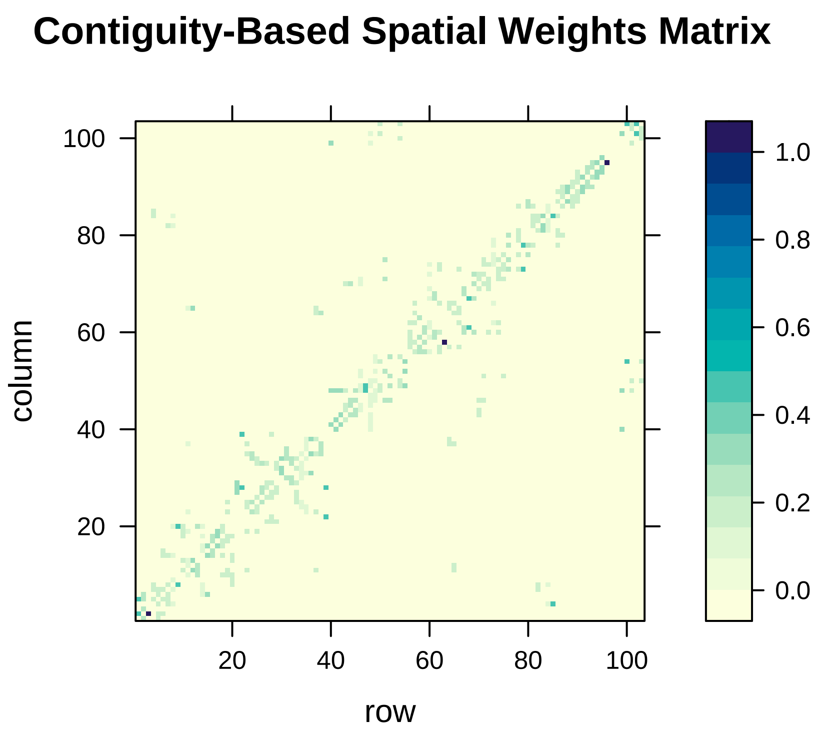
We now plot the three spatial weight matrices. We see that the distance decay weight matrix is symmetric. The distance decay matrix is

Distance Threshold Spatial Weights Matrix



Smooth Distance-Decay Spatial Weights Matrix





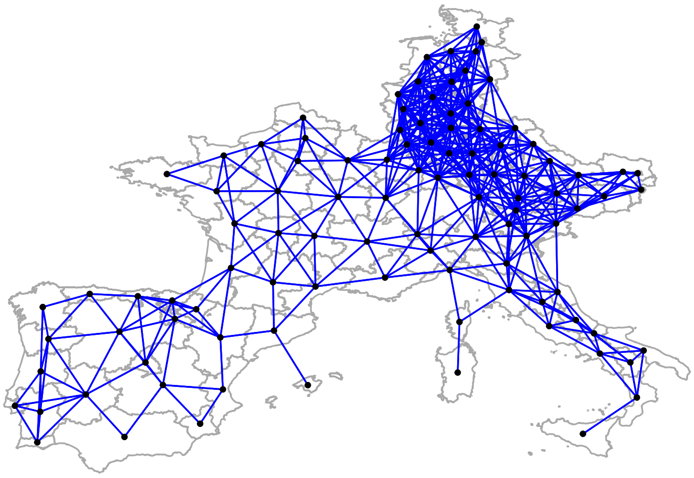
Try to visualize the network they represent

Let us first visualize the distance based spatial matrix. The first plot shows the map of Europe and the blue lines indicate the connections. The map shows the connectivity in Europe based on the distance threshold.

The second map visualizes the network based on the distance decay matrix. However, the edges do not display the intensity of connections, but just the connectivity to neighboring regions.

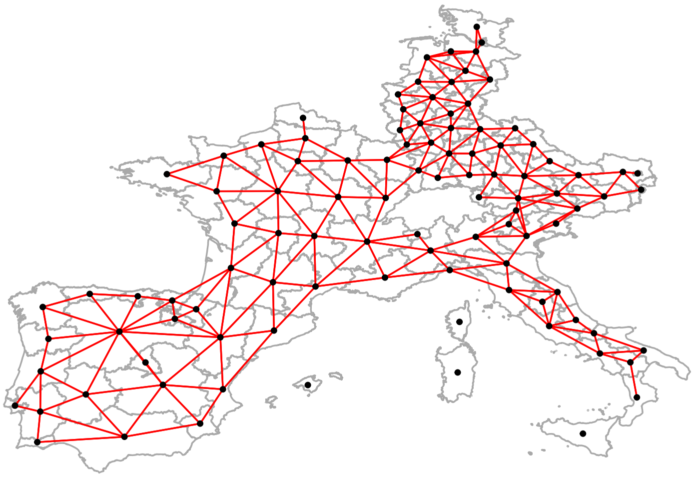
The last map displays the queen contiguity based measure. The islands in the middle sea are not being counted as neighbors. This should caution the use of this network, as it seems implausible that Siciliy for example is not connected to the mainland Italian provinces.

Distance Threshold



Distance Decay

Queen Contiguity



Test	Moran_I	Expectation	Variance	p_value
Distance Threshold	0.5938	-0.0098	0.0026	0
Smooth Distance-Decay	0.2925	-0.0098	0.0010	0
Contiguity-Based	0.5380	-0.0102	0.0045	0

Compute a suitable measure of spatial autocorrelation for productivity growth using these matrices. Point out differences, if there are any.

We calculate Global Moran's I as a measure of spatial autocorrelation for all three spatial weight matrices. All three matrices display the strong positive spatial autocorrelation between 0.62 - 0.54, which are all highly statistically significant with p-values < 0.01. Thus there is strong evidence for the presence of sizeable levels of spatial autocorrelation. This result is robust to the choice of the spatial weights matrix.

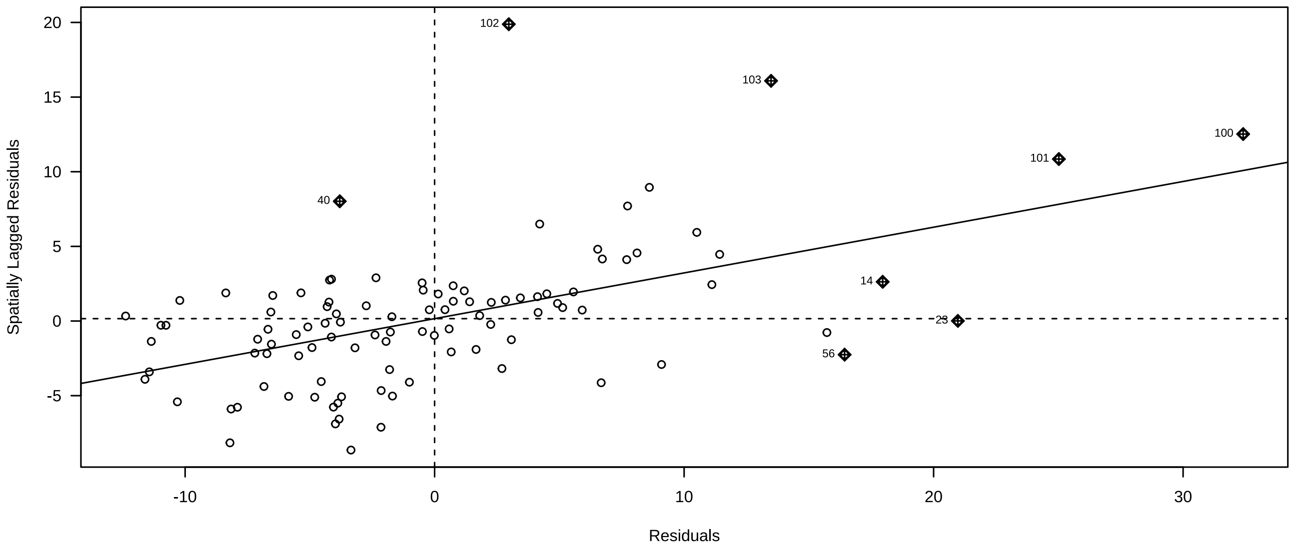
Estimate a linear regression model using OLS.

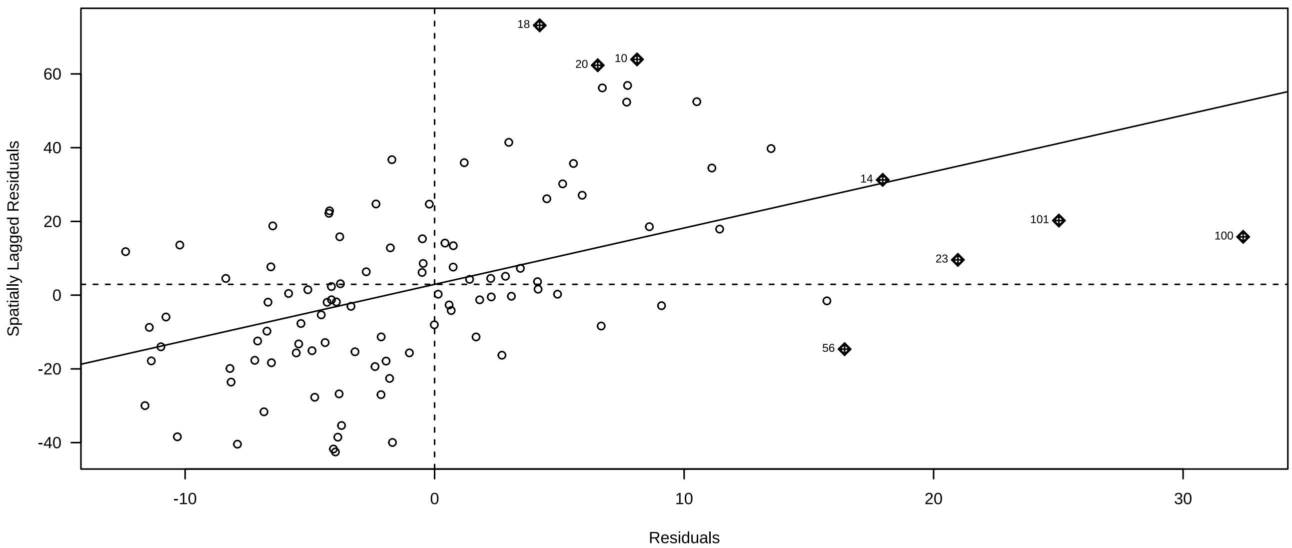
We estimate the specified model and obtain the following output.

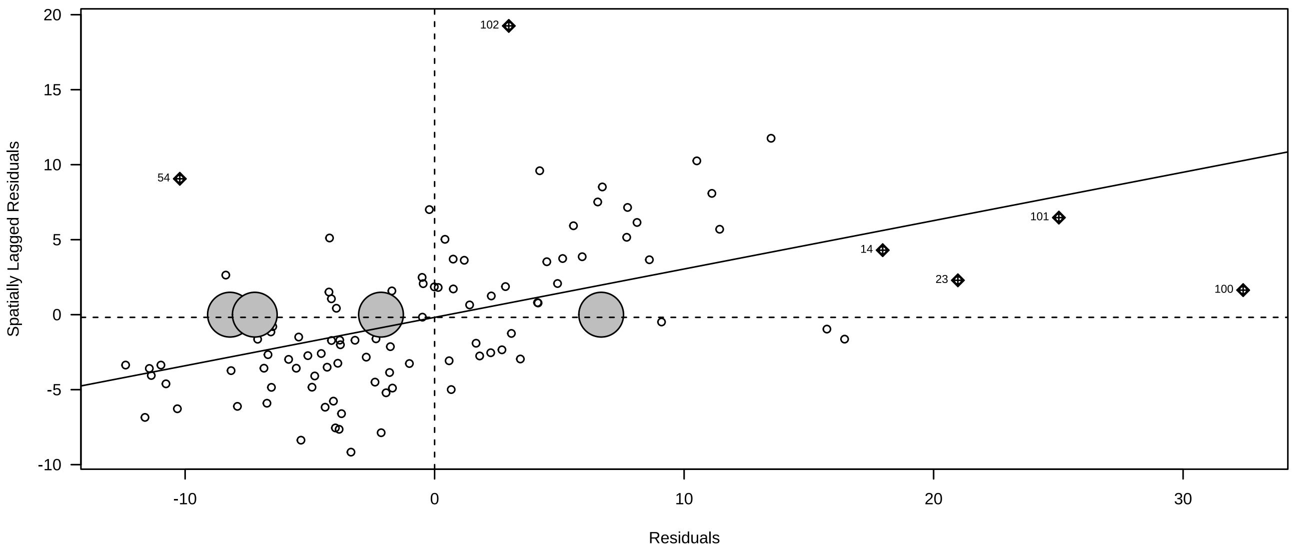
Table 4:

	Dependent variable:
	prod_growth
pr80b	-0.253^{***}
	(0.025)
lninv1b	0.032^{***}
	(0.008)
lndens.empb	0.007
	(0.009)
Constant	0.314^{***}
	(0.061)
Observations	103
R^2	0.528
Adjusted R^2	0.514
Residual Std. Error	0.080 (df = 99)
F Statistic	36.983^{***} (df = 3; 99)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

We observe strong evidence for spatial autocorrelation. This holds for all spatial weights matrices used. The different weighting schemes highlight differnt countries. Thus the neglect fo this spatial dimension might give rise to bias in the OLS estimated coefficients.

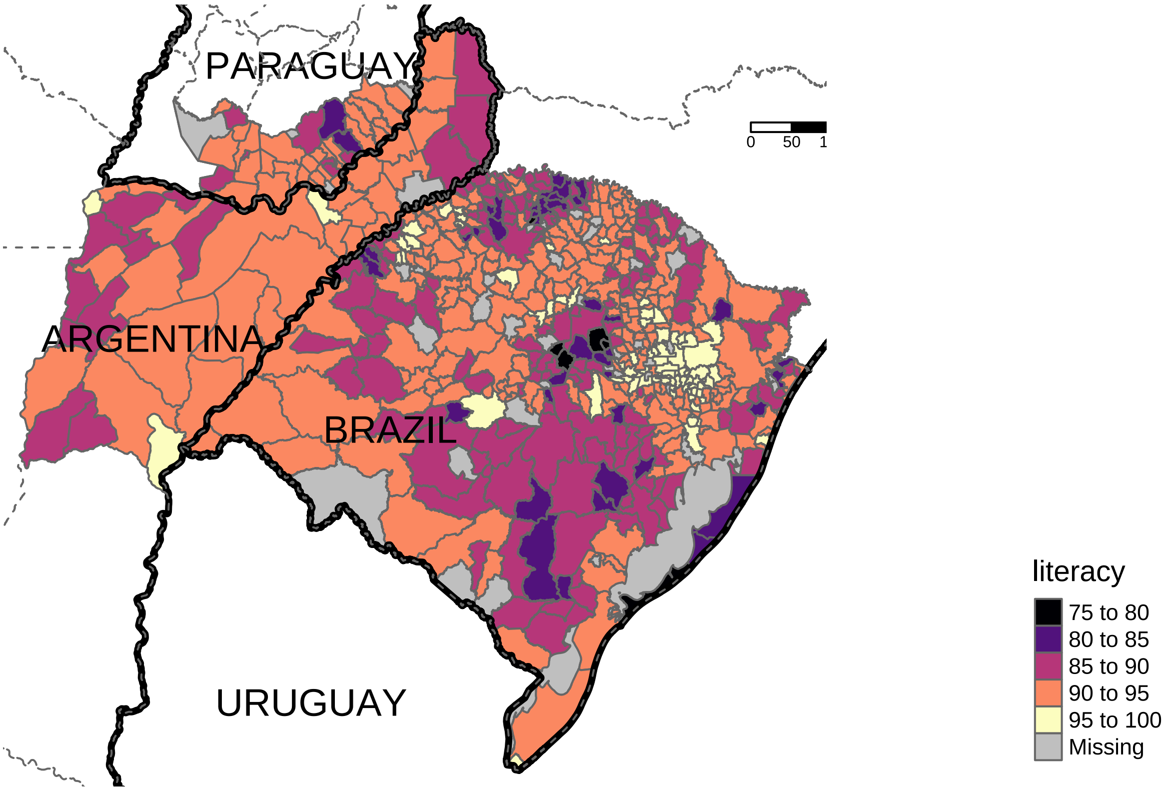

Spatially Lagged Residuals
Residuals
102
103
100
101
40
14
23
56


Spatially Lagged Residuals
Residuals
18
20
10
14
101
100
23
56


Spatially Lagged Residuals
Residuals
102
54
101
14
23
100

Exercise B

Creating maps



Below, we see each municipality and its proximity to its nearest Jesuit mission via color intensity.

Reproducing Table 2

```
####### Data Loading ####

setwd("/Users/gustavpirich/Library/Mobile
Documents/com~apple~CloudDocs/Wirtschaftsuniversitaet/MASTER/summer_term_2024/spatial_economics/data/Ta

argentina_brazil_paraguay <- read_dta("./Data/Tables/Argentina Brazil Paraguay Spatial.dta")
argentina_literacy <- read_dta("./Data/Tables/Argentina Literacy Spatial.dta")
brazil_literacy <- read_dta("./Data/Tables/Brazil Literacy Spatial.dta")
paraguay_literacy <- read_dta("./Data/Tables/Paraguay Literacy Spatial.dta")
```

In replicating the table, we used the author's Stata code and approximated our r code as best as possible to the variables included in the Stata code and its considerably different syntax.

Formulas for OLS

Below, we specified the 8 individual regressions used by the author into objects for multiple uses.

What stuck out was the fact that the author did indeed alter the variables for individual countries, beyond the inclusion of mesorregion controls for Brazil. For Paraguay, e.g., the variables for longitude and altitude were not included for the model but rather only used to compute the Conley standard errors.

We then specified fixed effects models to solely to extract within R^2 values.

For the Conley SE's we duplicate latitude and longitude entries as apparently, they cannot be simultaneously taken as explanatory variables and as autocorrelation variables with the same name (unlike in Stata).

Table 5: Effect on Illiteracy

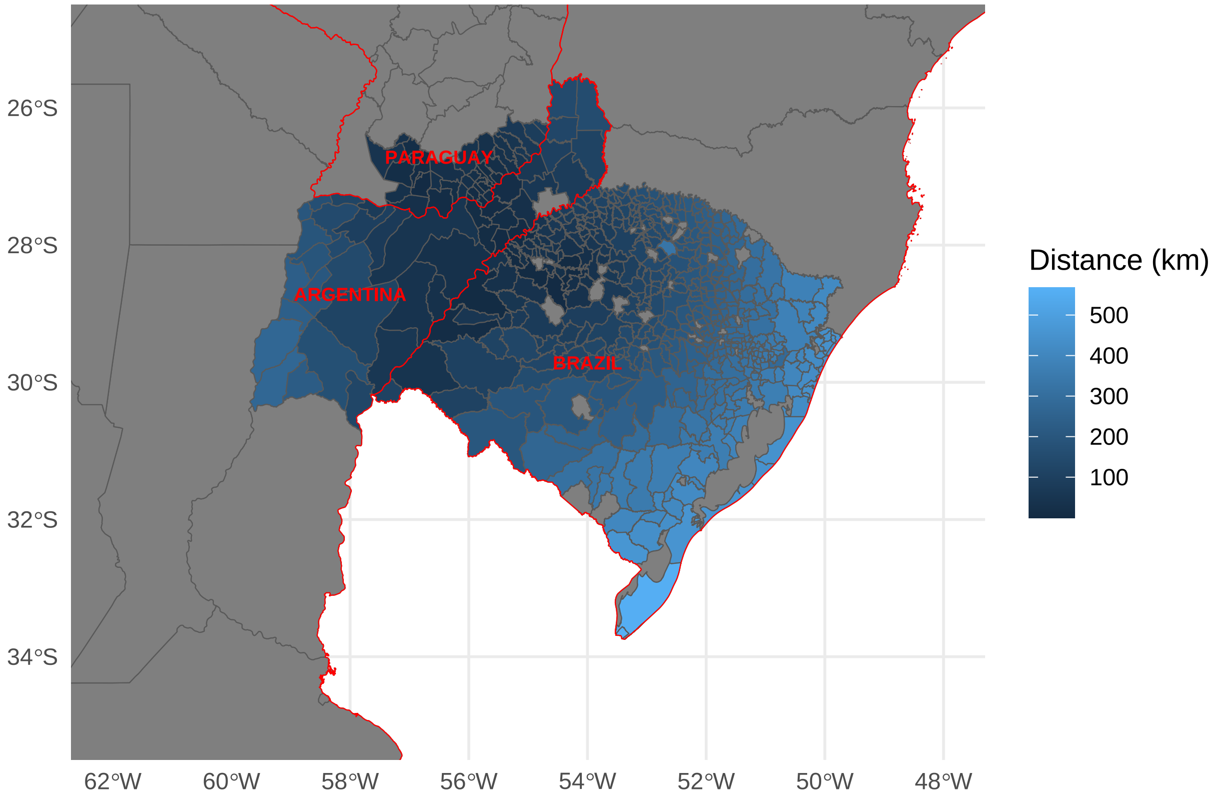
	Dependent variable:							
	illiteracy							
	Argentina, Brazil, and Paraguay		Brazil		Argentina		Paraguay	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mission distance	0.0105**	0.0112*	0.0200***	0.0313***	0.0157	0.0669**	0.0043	0.0138
	(0.0039)	(0.0046)	(0.0056)	(0.0077)	(0.0081)	(0.0232)	(0.0163)	(0.0264)
Conley SE	0.004	0.005	0.006	0.009	0.007	0.019	0.012	0.023
Geo controls	No	Yes	No	Yes	No	Yes	No	Yes
Within R^2	0.037	0.068	0.013	0.057	0.109	0.647	0.002	0.25
Observations	548	548	467	467	42	42	39	39
R^2	0.0419	0.0730	0.0562	0.0951	0.1651	0.6689	0.0039	0.2513

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

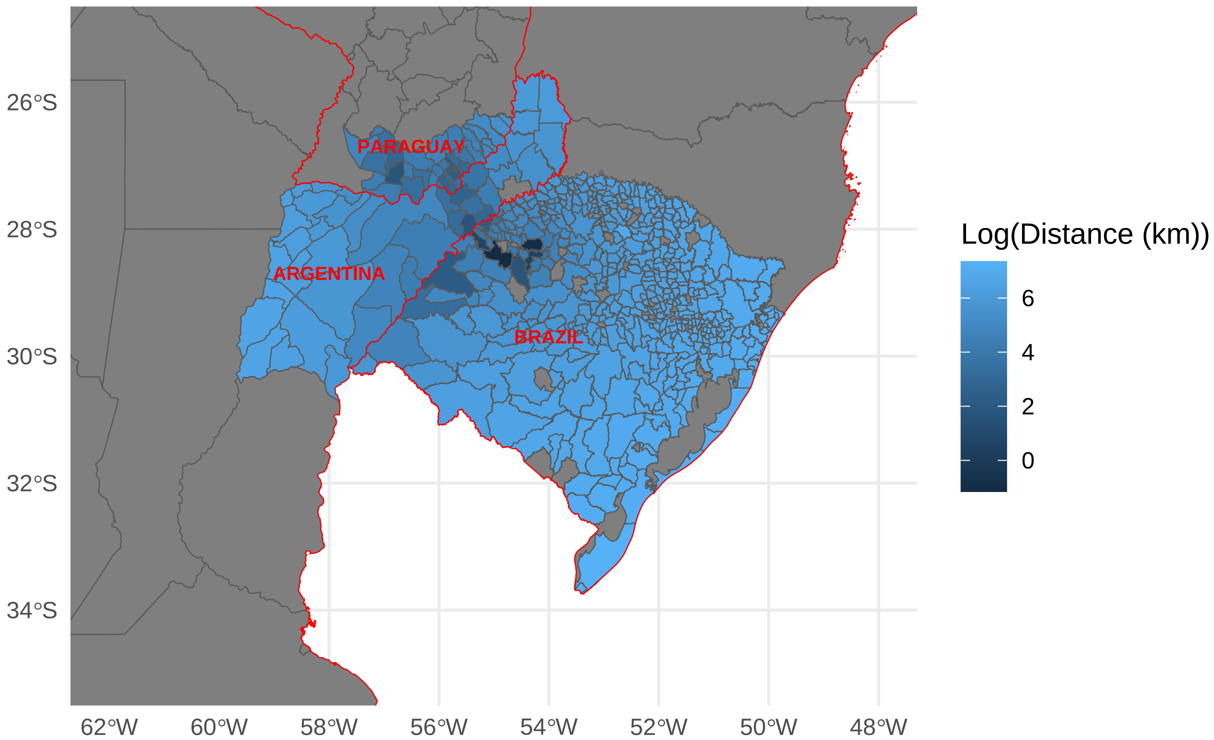
Operationalizations of Distances

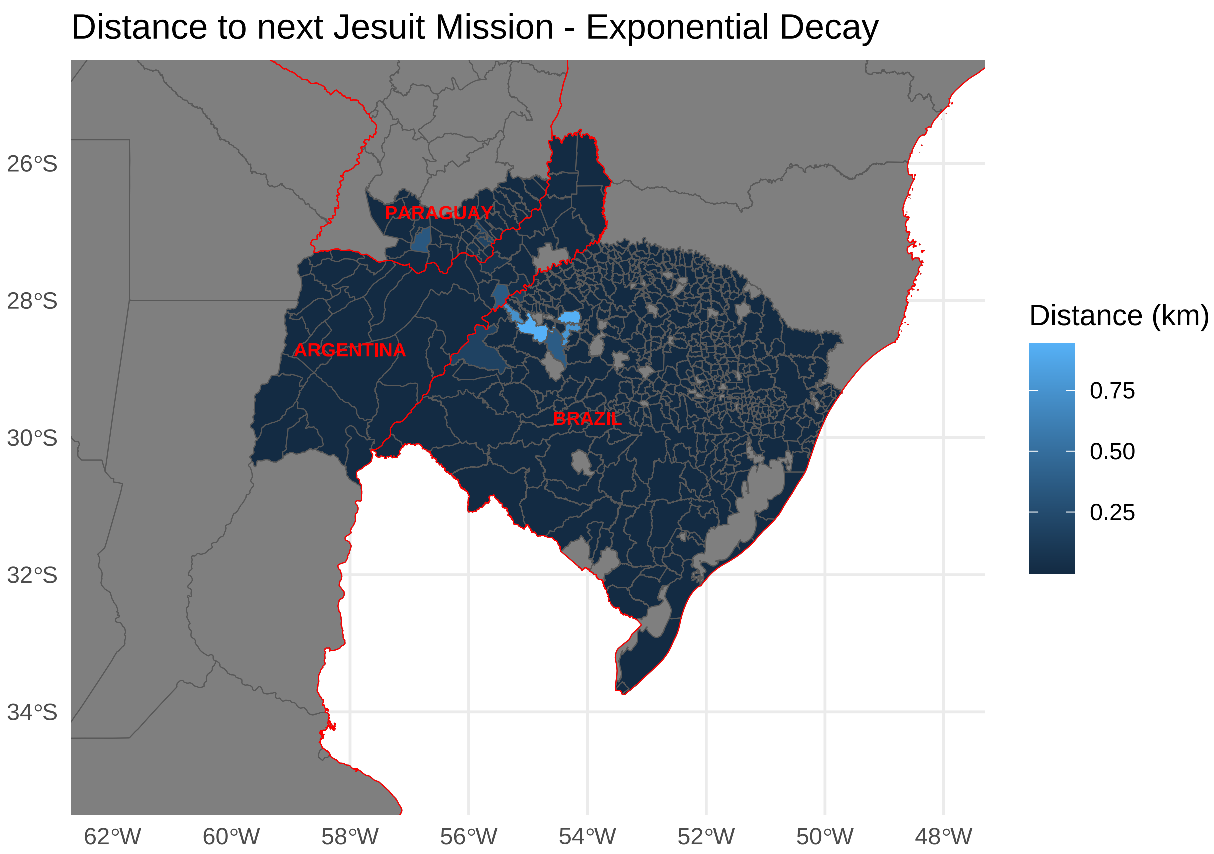
Caceido uses the nominal ditance in km from each municipality's centroid to arrive at his distance measures. Alternative transformations to that could be log -transformations (i.e. $log(d) + 1$) or methods for computing decay such as $\frac{1}{x}$ or exponential decay. For the latter we used a beta of -0.5.

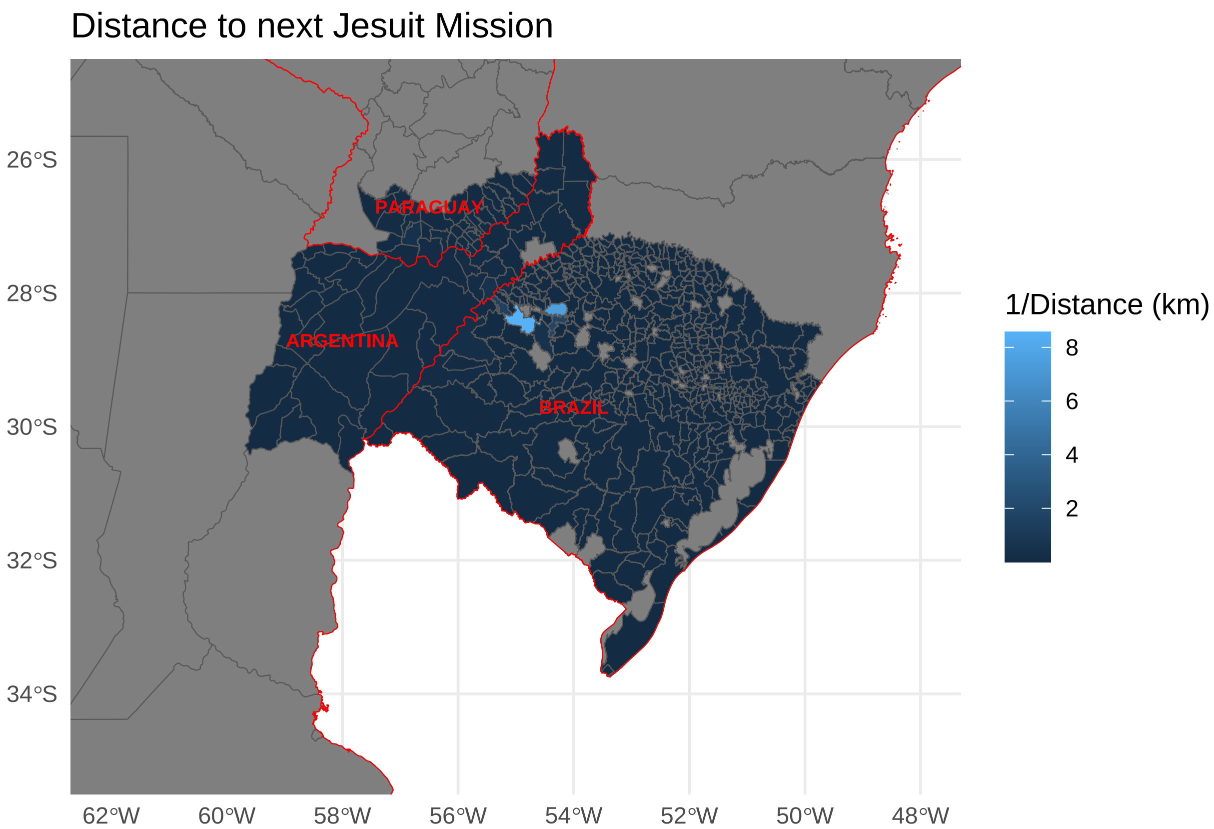
Distance to next Jesuit Mission



Distance to next Jesuit Mission







We can see that distance operationalization can indeed change the notion of distance considerably and hence our results. Especially in the case of strong exponential decay, the propagation of effects over space may be understated and hence lead to a potential underestimation of the effects. That can be caused, e.g., by creating a perimeter of influence of Jesuit human capital effects that is too small, and affected units would not be included in the appropriate subset. In doing so, one would falsely assign units that were indeed "treated" into "untreated", which can reasonaby expected to worsen the efficiency of estimates.

Exercise C

Recall 'The perils of peer effects' (Angrist, 2014). Write a short text (not more than 800 words) on the 'The perils of ignoring peer effects'.

- Touch on the topics of drawing valid inference and the trade-off between internal and external validity (think of an experimental setting vs, e.g. an actual classroom), and the goals of (applied and methodological) scientific research.
- Briefly explain how network dependence (spatial, social, etc.) may impact *validity* and *relevance* of a certain instrument. Consider weather instruments, the quarter of birth instrument by Angrist and Krueger (2001), or some instrument that yoiu are familiar with as an example.

In a sweeping review, Angrist (2014) provides a critique of the economics literature estimating peer effects. He derives and demonstrates potential pitfalls by linking the behaviour of IV estimation with group-level dummies with OLS. In a linear-in-means model with exogenous effects, he shows that the ‚social multiplier' is equivalent to the *ratio* between the IV 2SLS and OLS estimand. $Y_i = WX_i\beta + \varepsilon i \; \phi_1 = \frac{IV}{OLS}$

For an endogenous effects model the *difference* between the IV estimate and the OLS estimated is equivalent to the social multiplier. $Y_i = WY_i\delta + \varepsilon i \; \phi_2 = IV - OLS$

Selection bias, omitted variables, and measurement error can inflate or deflate the IV estimate of the coefficient and thus Angrist cautions attributing the ratio (or difference) to peer effects.

Applied researchers might have considered network dynamics as sources of potential contamination in a randomised experiment or observational study with quasi experimental variation. Thus researchers conceptualised peer effects as a violation of the stable unit treatment variable assumption (Rubin, 1978). Only recently economists have begun to explicitly model these network dependencies as ‚spillover effects'. Ignoring peer effects threatens the internal validity of both experimental and non-experimental findings. Consider for example the effect of a program at an individual level (some sort of mentoring program) on grades. By employing class or school fixed effects, the estimated effect size of the treatment at the individual level might be affected, because of positive peer effects (the mental health improve the mental health of others). Miguel and Kremer (2004) study the effects of a deworming RCT in Kenya. They highlight that by not including peer effects and reduced network transmission, the reduction in school absenteeism of the intervention are doubly undercounted. The reduced disease transmission, which spilled over to control schools leads to an undercount of the positive benefits of the intervention. Thus, ignoring peer effects is equivalent to ignoring an externality.

As peer effects and network dependency are everywhere and always present in the realm of social science research it is necessary to be explicit about their existence. While empirical estimation is bedevilled by problems, better data can provide a way forward to empirically establish the impact of peer effects.

The External Validity Tradeoff

Randomisation of peers in experimental settings might allow researchers to draw internally valid inference, the problem is that these research strategies might lack external validity. Network structure emerge endogenously, making the randomisation of peers in studies infeasible. In the real-world friends are not being randomly assigned, but people sort into groups and networks (often based on homophile). In many settings, it is practically infeasible to vary the structure of peers exogenously. For example, when trying to estimate the long-run effect of peers it is hard to imagine a practically feasible experiment where the network structure was determined exogenously and evolved so over time.

Policy Relevance and Peer Effects

How then can we learn something about peer effect and the importance of social networks? An innovative methodological contribution that sheds light on peer effects is Chetty et al. (2022). The researchers demonstrate that economic connectedness and social mobility demonstrates importance of social networks for economic outcomes. Chetty et al. (2022) do not rely on random assignment, nor do they estimate the impact of a certain policy. Thus the relationship between economic connectedness and economic mobility is fraught by a set of issues, but the incredible rich data allows to demonstrate that even broad based correlations can give us societally (and policy) relevant insights into the dynamics of social networks. This data-driven approach and correlational evidence coupled with extremely granular data can preclude concerns over threats to identification like reverse causality.

Network Dependence and Instruments

To consider how network dependence might affect the *relevance* and *validity* of an instrument let us consider the paper by Ramsay (2011). The author studies the relationship between natural resource wealth and a countries level of political freedom. More specifically, he is interested in the causal effect of a countries annual oil income per capita on the level of democracy as measured by the polity IV score. Let us consider the simplified equation: $Democracy_{i,t} = \mu + OilIncomepercapita_{i,t} + \delta X_{i,t} + \varepsilon_{i,t}$ where X is a vector of controls and for country i in year t.

However, this relationship likely suffers from reverse causality, as a countries political institutions will determine its oil income, while concurrently oil income determines democracy (the effect we are interested in). Ramsay (2011) proposes the out-of-region natural disaster as an IV for annual oil income. He splits the world into five regions and uses natural disasters, which impact the global oil price in other regions as an IV for annual oil income. Thus, Ramsay (2011) identifies the variation in annual oil income of a country in lets say Africa that is caused by a natural disaster in Colombia. $Oilincomepercapita_{i,t} = \mu + \theta OutofRegionNaturalDisaster_{i,t} + \delta X_{i,t} + \varepsilon_{i,t}$

Now how might spatial dependence in networks impact the validity (i.e. the exclusion restriction) of the instrument. $Cov(\text{OutofRegionNaturalDisaster}_{i,t}, \varepsilon_{i,t}/\delta X_{i,t}) = 0$ Network dependence through spatial interdependence can lead to a violation of the instruments validity of the exclusion restriction. Changes and levels of countries political institutions are correlated and clustered in space. Take for example the recent wave of military coups in Africa propagated through Africa and is likely to afflict political stability in other regions as well. The coarse aggregation of the world into five regions is creating spatial dependence. Additionally, the shocks induced by natural disasters is affecting other channels. Consider maybe a natural disaster in South America which might induce changes in the US foreign aid network, by redirecting resources away from one country to another. A reduction in foreign aid in Africa might then concurrently impact political freedom in Africa. Another obvious candidate for violations of the assumptions is that natural disasters will affect shock trade networks which will have a direct impact on countries political institutions. Thus the effects of shocks induced by natural disasters are clustered in space through countless channels. The authors assumption that, by dividing the world into five regions and that the shocks in those regions do not spill over in systematic ways seems implausible. The *relevance* of the instrument, however, is not directly impaired by network dependence as the first stage F-test still demonstrates that the instrument is highly correlated with oil income per capita.

Bibliography

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics* 30 (October): 98–108. https://doi.org/10.1016/j.labeco.2014.05.008.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2020. "Peer Effects in Networks: A Survey." *Annual Review of Economics* 12 (1): 603–29. https://doi.org/10.1146/annurev-economics-020320-033926.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, et al. 2022. "Social Capital II: Determinants of Economic Connectedness." *Nature* 608 (7921): 122–34. https://doi.org/10.1038/s41586-022-04997-3.

Kopczewska, Katarzyna, and Paul Elhorst. 2024. "New Developments in Spatial Econometric Modelling." *Spatial Economic Analysis* 19 (1): 1–7. https://doi.org/10.1080/17421772.2023.2281173.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217. https://doi.org/10.1111/j.1468-0262.2004.00481.x.

Ramsay, Kristopher W. 2011. "Revisiting the Resource Curse: Natural Disasters, the Price of Oil, and Democracy." *International Organization* 65 (3): 507–29. https://doi.org/10.1017/S002081831100018X.

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *The Annals of Statistics* 6 (1). https://doi.org/10.1214/aos/1176344064.