SPATIAL DATA

Valerio Leone Sciabolazza

valerio.leonesciabolazza@uniroma1.it

In this course, you will be exposed to various pieces of empirical economics papers that use spatial datasets. In addition, you will have hands-on experiences of using R to replicate actual pieces of the empirical research conducted in these papers. At the end of the course, you will have the ability to process spatial datasets needed for your own research.

Syllabus

Lecture 1: Introduction to Causal Inference and Spatial Data

In this lecture, we will discuss how satellite image data has recently been used by economists and learn the basics of spatial datasets, including polygons, raster data, and map projections.

Reference papers:

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." Journal of Economic Perspectives, 30(4): 171–198.

Kudamatsu, Masayuki. 2018. "GIS for Credible Identification Strategies in Economics Research." CESifo Economic Studies, 64(2), 327-338.

Lecture 2: Spatial Join

We will replicate how this paper spatially merges ethnic group level data with the grid of weather data, a great source of exogenous variation.

Reference paper:

Alsan, Marcella. 2015. "The Effect of the Tse Tse Fly on African Development." American Economic Review, 105(1): 382–410.

Lecture 3: Buffer

We will replicate how Conley & Udry (2010) match each agricultural plot with its neighboring plots within a radius of one kilometer in order to control for common shocks in peer effect estimation.

Reference papers:

Conley, T. G. 1999. "GMM estimation with cross sectional dependence." Journal of Econometrics 92 (1): 1–45.

Conley, Timothy G., and Christopher R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana." American Economic Review, 100(1): 35–69.

Manski, C. F. 1993. "Identification of endogenous social effects: The reflection problem." The Review of Economic Studies, *60*(3), 531-542.

Lecture 4: Distance

This lecture demonstrates how Nunn (2008) constructs instrumental variables using distance from each country's centroid to the nearest coastal point. We will also replicate how slave export data at the ethnic group level are aggregated to the country level. Finally, we will discuss the challenges and limitations of using distance-based instruments, highlighting potential pitfalls in their application.

Reference paper:

Nunn, Nathan. 2008. "The Long-Term Effects of Africa's Slave Trades." Quarterly Journal of Economics, 123(1): 139–176.

Lecture 5: Zonal Statistics

We will replicate how this paper calculates the standard deviation of land suitability for agriculture at the 0.5-degree cell level within each "virtual country" (i.e., a 2.5-degree cell), where the boundary is exogenously defined.

Reference paper:

Michalopoulos, Stelios. 2012. "The Origins of Ethnolinguistic Diversity." The American Economic Review, 102(4): 1508–1539.

Lecture 6: Elevation

This lecture replicates how this paper constructs instrumental variables for dam construction using the slope of rivers.

Reference paper:

Duflo, Esther, and Rohini Pande. 2007. "Dams." Quarterly Journal of Economics, 122(2): 601–646.

Lecture 7: Spatial Regression Discontinuity Design

We will replicate how Dell (2010) obtains each observation's distance to the treated area boundary (i.e., the cutoff) and the nearest boundary segment in order to implement a regression discontinuity design (RDD). In addition, we will replicate how the paper accounts for elevation in calculating road length in Peru, a hilly country. The lecture will also cover recent methodological advances in RDD, focusing on strategies to improve identification around cutoffs and accurately estimate changes in outcomes.

Reference papers:

Dell, Melissa. 2010. "The Persistent Effects of Peru's Mining Mita." Econometrica, 78(6): 1863–1903.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2019. "Regression discontinuity designs using covariates," Review of Economics and Statistics, 101, 442–451.

Lecture 8: Map Algebra

Map algebra allows you to edit raster data at the cell-by-cell level. We will replicate how this technique is used to figure out the origin of states.

Reference paper:

Mayshar, J., Moav, O., & Pascali, L. (2022). "The Origin of the State: Land Productivity or Appropriability? " Journal of Political Economy, 130(4), 1091-1144.

Lecture 9: Interference and Spillover Effects

A recurring challenge with spatial data is that observations are rarely independent across space, and treatments may generate spillover effects on neighboring units. In this lecture, we will learn how to identify potential spatial autocorrelation using tests such as Moran's I and Geary's C, and how to detect neighborhood-level common trends. We will examine methods for modeling spatial autocorrelation both in observational settings (e.g., Conley's standard errors) and in quasi-experimental designs (e.g., spatial first differences, ring-based approaches) leveraging the fact that treatment effects often decay with distance

Reference Papers:

Butts, K. 2023. "Difference-in-differences with geocoded microdata." Journal of Urban Economics 133, 103493.

Conley, T. G. 1999. "GMM estimation with cross sectional dependence." Journal of Econometrics 92 (1): 1–45.

Druckenmiller, H., Hsiang, S. 2018. "Accounting for unobservable heterogeneity in cross section using spatial first differences. " NBER working paper series, No. w25177.

Manski, C. F. 1993. "Identification of endogenous social effects: The reflection problem." The Review of Economic Studies, *60*(3), 531-542.

Required materials:

In order to participate to the class, you'll need the following:

- Your own laptop.

- R and RStudio installed on your laptop (please make sure to have the latest available version).

- The relevant materials downloaded to a folder in your laptop. Provided during the first class.

- Install the following R packages: conleyreg, cubelyr, dplyr, elevatr, exaxtextract, fixest, haven, ggplot2, ggrepel, ggthemes, lwgeom, modelsummary, parallel, pbapply, raster, readr, readxl, reshape2, rgdal, sf, stars, starsExtra, terra, tidyr, viridis, WDI.

Exam

The final exam will consist of an applied exercise. Students will be assigned excerpts from a recent empirical paper describing the setting, research question, and data. Based on this information, students will be required to design an identification strategy that leverages the methods learned throughout the course. The exam will also require a critical discussion of the proposed approach, including its strengths and limitations. The exam will be held in person and will have a duration of two hours.