Historical Persistence

Applied Economics Research Course

Bas Machielsen

Utrecht University

2023-11-13

Introduction

Institutions

- Economic or other outcomes are not only influenced directly by relative prices, costs and benefits.

- They are also influenced by more latent, long-term factors, often called institutions.

  - Example: Acemoglu, Robinson and Johnson (2001):

    > (..) estimate the effect of institutions on economic performance. Europeans adopted very different colonization policies in different colonies, with different associated institutions. In places where Europeans faced high mortality rates, they could not settle and were more likely to set up extractive institutions. (..) Exploiting differences in European mortality rates as an instrument for current institutions, we estimate large effects of institutions on income per capita.

Institutions

- Here's a world map with GDP/capita in 1995 for a selected number of countries:

```r
library(sf); library(rnaturalearthdata); library(ivdoctr)
world_map ← rnaturalearthdata::countries50 ▷ st_as_sf()
wm ← world_map ▷ left_join(ivdoctr::colonial,
                           by = c("iso_a3" = "shortnam"))
wm ▷ ggplot(aes(fill=logpgp95)) + geom_sf()
```
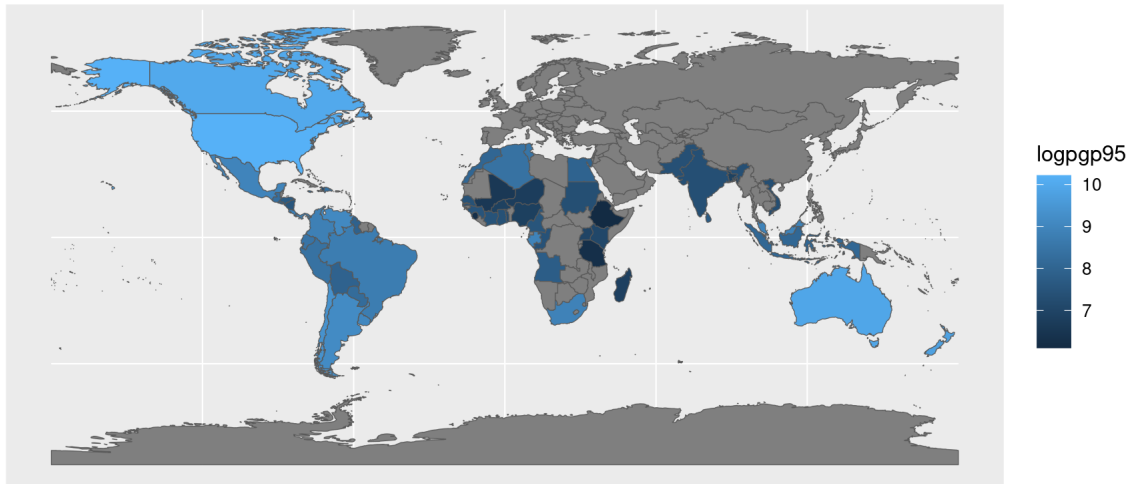
Institutions

- In this paper, Acemoglu, Johnson and Robinson (2001) take institutions to be *expropriation risk*: Measures risk of government appropriation of foreign private investment on a scale from 0 (least risk) to 10 (most risk), averaged over all years from 1985-1995.

- If one wants to find the relationship between institutions and economic well-being, one runs into the problem of *endogenity*:

  - Institutions causally affects economic growth
  - But economic growth also affects institutions!

  - Formally:

    - $\text{GDP}/\text{Cap}_i = \alpha_0 + \alpha_1 \text{Institutions}_i + \epsilon_i^1$
    - $\text{Institutions}_i = \beta_0 + \beta_1 \text{GDP}/\text{Cap}_i + \epsilon_i^2$

- If this is the *true* data-generating process, an OLS regression will not yield the correct coefficient you are interested in (α_1).

Instrumental Variables

- People have devised various strategies to solve this problem. The most often-used strategy is to use an *instrumental variable*

    - An instrumental variable is a variable that exogenously causes a change in X while not directly affecting Y.

    - In the AJR (2001) case:
    $$\text{Institutions}_i = \beta_0 + \beta_1 \text{GDP/Cap}_i + \underbrace{\beta_2 \text{SettlerMortality}_i}_{\text{Instrumental Variable}} + \epsilon_i^2$$

    - Now, GDP per capita and Institutions are *endogenously* determined as a result of *exogenous* settler mortality

Instrumental Variables

- You can rearrange these two equations to find that:

    - $\text{Institutions}_i = \cdots + \underbrace{\left(\frac{\beta_1 \alpha_1 \beta_2}{1 - \alpha_1 \beta_1} + \beta_2 \right)}_{= \frac{\beta_2}{1 - \alpha_1 \beta_1}} \cdot \text{SettlerMortality}_i + \ldots$

    - $\text{GDP/Cap}_i = \cdots + \left(\frac{\alpha_1 \beta_2}{1 - \alpha_1 \beta_1} \right) \cdot \text{SettlerMortality}_i + \ldots$

- Now you can find (*identify*) α_1 by dividing $\text{Cov}(\text{GDP/Cap}, \text{SettlerMort})$ by $\text{Cov}(\text{SettlerMort}, \text{Institutions})$

- The key (and untestable) assumption here is that settler mortality does not directly influence GDP per capita in 1995.

- Loosely speaking, settler mortality represents an exogenous shock to institutions, allowing us to identify the influence of a change in institutions on GDP per capita

Historical Persistence

Back to historical persistence

- In general, we are dealing with outcomes that are determined endogenously: outcomes now and outcomes in the past are determined by other, latent factors

- To identify the influence of these latent factors, economists often use large-scale, influential events that shock these institutions. They compare outcomes in places that were initially similar, but some of them have been coincidentally exposed to certain events, whereas others have not.

- Usually, this research is based on spatial regression discontinuity designs, comparing places at one side of the border with places at the other side of the border (e.g. Dell, 2010, Lowes and Montero, 2021).

Historical Persistence

- The set-up in general is:

  - Outcome Y_i has "deep roots" caused by some D_i long ago
  - This D_i usually did not came to be randomly, but at the margin near some threshold, it did
  - Example (Dell 2011): Spanish conquest of Latin America was bounded by mountainous areas

- Comparing villages just within vs. just outside of the old imperial borders

- Other seminal papers:

  - Acemoglu et al. (2011), Occupied areas by Napoleon in Germany do better
  - Dell & Olken (2020): Positive development effects of extractive colonialism in Indonesia
  - Voigtlander & Voth (2012): Deep roots of persecution in Nazi Germany

Spatial Regression Discontinuity

- Consider the example of **Roman roads**. Let's suppose that Roman roads, by influencing trade networks, stimulate development:

  - $Y_i = \alpha_0 + \alpha_1 \text{RomanRoad}_i + \alpha_2 \text{Environment}_i + \epsilon_i^1$

- But Roman Roads have not been built randomly by the Romans, but might have built in areas with more suitable environments for development:

  - $\mathbb{P}[\text{RomanRoad}_i] = \beta_0 + \beta_1 \text{Environment}_i + \epsilon_i^2$

- Then, comparing areas with roads to areas without roads doesn't work due to differences in the environment:

  - $\mathbb{E}[Y|RR = 1] = \alpha_0 + \alpha_1 + \alpha_2 \mathbb{E}[\text{Environment}|\text{RR} = 1]$

  - $\mathbb{E}[Y|RR = 0] = \alpha_0 + \alpha_2 \mathbb{E}[\text{Environment}|\text{RR} = 0]$

Identification Strategy

- If we can find two places with the same expected value of getting a road, but only *one* of them gets a road..

- According to the equations on the previous slide, this means two places with the same environment (take the expected value of Y_i conditional on Environment to see that)

- Then, we can estimate α_1 by:

  - $\mathbb{E}[Y|\mathrm{RomanRoad} \ \& \ \mathrm{P}[\mathrm{RomanRoad}]]$ - $\mathbb{E}[Y|\mathrm{NoRomanRoad} \ \& \ \mathrm{P}[\mathrm{RomanRoad}]]$

- This is what we attempt to do in a spatial regression discontinuity design: compare two arguably identical places with the same probability of treatment, where only one of them actually gets the treatment

The Data

Data

- I have two a couple of data sets on offer to you:

  - A dataset of French and German municipalities overlaid with Roman roads (`france_germany.shp`)

  - This dataset contains the distance of each municipality to a Roman road, as well as an indicator, whether a road streches through a municipality or not

  - A dataset of Dutch municipalities overlaid with the border of the former Roman empire (`netherlands_rome.geojson`)

  - Again including a variable for distance to the border, and an indicator of whether a present-day municipality was inside or outside the former Roman empire

- You are also free to use or compile a custom dataset using different settings and countries.
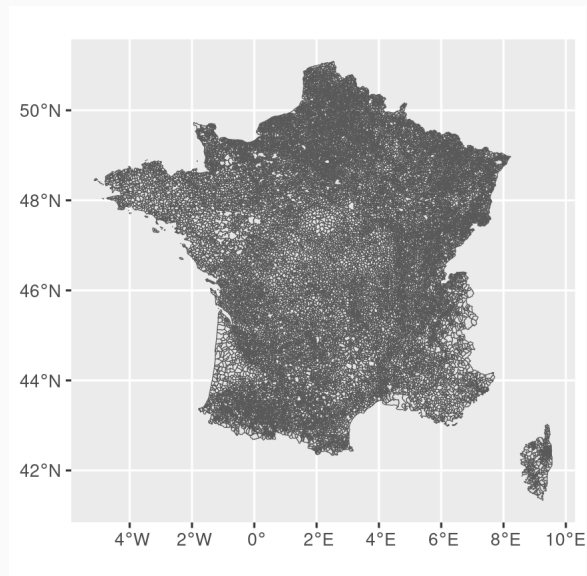
Demonstration

Demonstration

```
library(cawd); library(giscoR)
roads ← cawd::darmc.roman.roads.major.sp ▷ st_as_sf()
lau ← giscoR::gisco_get_lau(year = '2019', country = c("France"))

xmin ← -10; xmax ← 20; ymin ← 40; ymax ← 80
box ← st_bbox(c(xmin=xmin, xmax=xmax, ymax=ymax, ymin=ymin), crs = st_crs(4326))
square ← st_sf(st_as_sfc(box))
fr ← st_intersection(lau, square)

fr ▷ ggplot() + geom_sf()
```

What does this data.frame look like?

```
fr ▷ head(5)
```

```
## Simple feature collection with 5 features and 10 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 4.248763 ymin: 49.54231 xmax: 5.31495 ymax: 50.12165
## Geodetic CRS:  WGS 84
##         id GISCO_ID CNTR_CODE LAU_ID            LAU_NAME POP_2019 POP_DENS_2019  AR
## 1 FR_08026 FR_08026        FR  08026  Aubigny-les-Pothées      318      21.48498 14.
## 2 FR_08027 FR_08027        FR  08027 Auboncourt-Vauzelles      102      19.06222  5.
## 3 FR_08028 FR_08028        FR  08028             Aubrives      874      82.53602 10.
## 4 FR_08029 FR_08029        FR  08029             Auflance       85      13.92512  6.
## 5 FR_08030 FR_08030        FR  08030                 Auge       61      13.42680  4.
##                 _ogr_geometry_
## 1 POLYGON ((4.404236 49.75043 ...
## 2 POLYGON ((4.465646 49.56644 ...
## 3 POLYGON ((4.770839 50.09282 ...
## 4 POLYGON ((5.289269 49.62845 ...
## 5 POLYGON ((4.260257 49.86763 ...
```
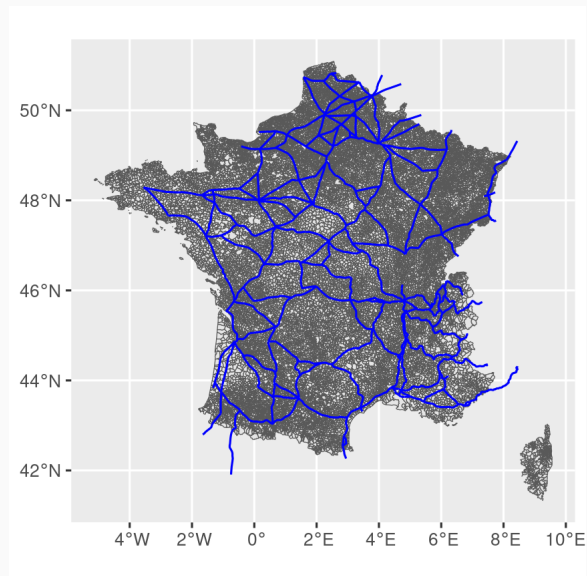
Merge this with the roads

```
numbers ← st_intersects(roads, fr) ▷
  as.data.frame() ▷ select(row.id) ▷ pull() ▷ unique()

roads_in_fr ← roads ▷ filter(is.element(row_number(), numbers))


ggplot() +
  geom_sf(data = fr) +
  geom_sf(data = roads_in_fr, color = 'blue')
```

Compute the distance

```
minimum_distances ← fr ▷
  st_centroid() ▷
  st_distance(roads_in_fr) ▷
  apply(1, min)


minimum_distances[1:10]


## [1] 12389.074  5728.599 34012.564  7624.885 24728.569 11728.691  2126.628  7900.327


fr ← fr ▷
  mutate(min_dist_to_road = minimum_distances)
```
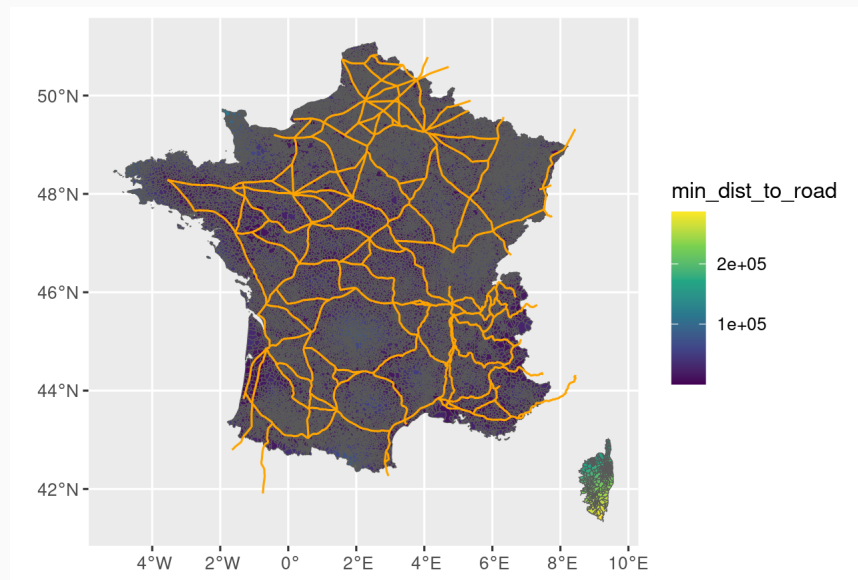
Plot the result

```
ggplot() +
  geom_sf(data = fr, aes(fill = min_dist_to_road), size = 0.0001) +
  scale_fill_viridis_c() +
  geom_sf(data = roads_in_fr, color = 'orange')
```

A small analysis

```r
library(fixest); library(modelsummary)
model1 ← fixest::feols(POP_DENS_2019 ~ min_dist_to_road, data = fr)

modelsummary(model1, gof_map = c("r.squared", "nobs"), stars=T)
```

	(1)
(Intercept)	195.676***
	(4.821)
min_dist_to_road	-0.002***
	(0.000)
R2	0.004
Num.Obs.	35227
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Your task

- Find municipality-level or geographical outcomes that suit an *interesting research question*

    - Present-day or more historical/macro-level outcomes

- Make sure you control for environments!

    - Overlapping regions (provinces, supersets of municipalities)
    - But also possible: terrain, other initial conditions (coastal municipalities)
    - Make it plausible that you compare apples with apples
    - Some roads have also been built with a destination in mind: how to control for that?

- Explore mechanisms through which the effect can work

Course Schedule

- We have 8 meetings

- Meeting 1 (16 Nov): Plenary meeting / Introduction

- Meeting 2 (23 Nov): Plenary meeting / Preliminary Research Question
- Meeting 3 (30 Nov): Plenary meeting / Research Design - Data- Methodology
  - **Send Research Proposal**
- Meeting 4 (7 Dec): **Presentations** and discussion of Research Plan / Feedback (possibly individual)
- Meeting 5 (14 Dec): Plenary meeting / Execution of research plan
- Meeting 6 (21 Dec):
  - 1st half: Plenary meeting: **Presentations** of progress
  - 2nd half: Individual meetings: discussion & feedback
- Meeting 7 (11 Jan): Individual meeting / Execution of research plan
  - Send Draft paper
- Meeting 8 (18 Jan): **Presentations** of draft paper & feedback

Expected output

- At the end of week 3 you will upload your research plan in Osiris, including problem statement, (preliminary) literature review and research approach. This plan and your performance in the research group so far will be the input for your supervisors' go/no go decision (crucial with respect to your compliance with the effort requirements).

- In the meeting in week 4, the focus is on the research plan, which is a crucial element in the overall research process. It should contain the motivation for the central problem statement and a plan how to do the actual research.

- Clearly, all members of the supervision group should receive each other's' draft research plan in time, that is, a few days before the meeting itself; to make the meeting as useful as possible, everyone should have carefully read the different research plans, should have given each of these a thought and should have prepared some questions or suggestions.

Expected output

- In each presentation (meeting 4, meeting 6, meeting 8) session, every student gives a short and focused presentation of his/her research project – corresponding to the stage the project is in – and elaborates both on the contents and the used research method and methodology, on choices that have been made and on solved or unsolved problems they have encountered in their work.

- Each presentation is followed by a general discussion where all participants are supposed to contribute by asking questions for further explanation as well as critical questions about the actual research and by providing constructive suggestions and comments. *This feedback needs to be written down in a document and handed in at the start of each session*

- **Final Paper Due**: 25 January 2024

Contact

- You can always contact me at a.h.machielsen@uu.nl

- But I also create a **Microsoft Teams group**

  - Note to self: do this now!

- And I can send you the data in due time (Microsoft Teams files)

Data Sources

- World Values Survey: https://www.worldvaluessurvey.org - Geographically coded information about norms & values
- Demographics & Health Survey: https://dhsprogram.com/data/ - Also geocoded information about D&H
- Eurostat: https://ec.europa.eu/eurostat/web/main/data/database - Database about European countries on aggregate (country) and more disaggregated levels
- Waar Staat Je Gemeente?: Dutch Municipality Data
- Statistiques Locales (FR): French Municipality Data
- Global Climate Database: http://www.worldclim.org/
- My own website data overview (includes municipality data): here
- Clio-Infra: https://github.com/basm92/Clio
- Standard scholarly measures of politics: https://github.com/xmarquez/democracyData
- Correlates of war, armed conflicts: https://rdrr.io/cran/peacesciencer/
- Political protests: https://acleddata.com/
- Labor conflicts: https://datasets.iisg.amsterdam/dataverse/labourconflicts
- World Bank, IMF, Countries' National Statistical Agencies, Countries' Parliaments

References & Literature

Acemoglu, D., Cantoni, D., Johnson, S., & Robinson, J. A. (2011). The consequences of radical reform: The French Revolution. American economic review, 101(7), 3286-3307.

Beach, B., & Hanlon, W. W. (2022). Culture and the historical fertility transition. The Review of Economic Studies.

Dell, M. (2010). The persistent effects of Peru's mining mita. Econometrica, 78(6), 1863-1903.

Dell, M., & Olken, B. A. (2020). The development effects of the extractive colonial economy: The dutch cultivation system in java. The Review of Economic Studies, 87(1), 164-203.

Lowes, S., & Montero, E. (2021). Concessions, violence, and indirect rule: evidence from the Congo Free State. The Quarterly Journal of Economics, 136(4), 2047-2091.

Lowes, S. (2022). Kinship Structure and the Family: Evidence from the Matrilineal Belt (No. w30509). National Bureau of Economic Research.

References & Literature

Jones, B. F., & Olken, B. A. (2005). Do leaders matter? National leadership and growth since World War II. The Quarterly Journal of Economics, 120(3), 835-864.

Voth, H. J. (2021). Persistence–myth and mystery. In The handbook of historical economics (pp. 243-267). Academic Press.

Voigtländer, N., & Voth, H. J. (2012). Persecution perpetuated: the medieval origins of anti-Semitic violence in Nazi Germany. The Quarterly Journal of Economics, 127(3), 1339-1392.