AEA project ID: openicpsr-120243

Replication package for *"The origins of common identity: Evidence from Alsace-Lorraine",*
by Sirus H. Dehdari (sirus.dehdari@statsvet.uu.se) and Kai Gehring (mail@kai-gehring.net).

The following package contains French municipal-level data and individual-level survey data as well as R-scripts required to replicate the results in the paper and the appendix. Note that the replication R-scripts run on our final Stata data file, which is based on a large variety of underlying raw data. Due to various restrictions, we are unable to provide several of these raw data sets. However, we include the Stata do file (Scripts/AlsaceLorraine.do) that produces the final Stata data file. Below, we list which data sets we have provided and which could not be provided, together with detailed descriptions on how to access the data (see Data Availability Statement). **Note that the included Stata do file cannot be executed since many of the raw data sets are not included.**

Several of the manipulations in GIS over the last years were made directly in ArcGIS, which involved using the georeferencing tools, creating maps about the borders, administrative division and depiction of the outcomes. We also used the spatial statistics to compute distances and spatial averages. The projection that was used is "ETRS_1989_UTM_Zone_32N". We provide sufficient information in the appendix about all sources and are happy to guide researchers who aim to work with those data, or provide modified shapefiles of the borders that we georeferenced upon request. A comprehensive list of all variables as well as detailed description of their sources are available in the Appendix Section E, Figures E5 to E9. The data sets used are

- AlsaceLorraine.dta: comprise municipal-level data, mostly from *Institut national de la statistique et des études économiques* (INSEE). The distance variables have been computed using ArcGIS.
- border_pair_distance_final_merged.dta: comprise municipal-level data for all municipalities included in pairs of bordering départements.
- cahiers.dta: data based on city-level average loyalty assessments from the Cahier de doléances.
- natidentity2003_dep.dta: survey data from the ISSP National Identity 2003, aggregated on département level, together with département-level referenda results for 1992 and 2005.
- oip99_01.dta and oip2003.dta: individual-level survey data from the *Observatoire Interrégional du Politique* (OIP).

Note that Figure 6 was not created by the R-scripts, but in Excel. This figure is not an analysis, but rather a descriptive depiction of historical raw data that we collected. We have included the underlying data in the replication package (other_data\regionalseats.txt).

Individual-level survey data from *Observatoire Interrégional du Politique* are used in the paper. The data has been de-identified and do not compromise the privacy of the respondents.

Exact steps to follow for replication:

Data and code are available as AEA Project ID: openicpsr-120243

The results were obtained on a HP ZBook 15 G3, OS Microsoft Windows 10 Enterprise, using R version 3.5.1 (2018-07-02) [platform: x86_64-w64-mingw32/x64 (64-bit)]. The tex-files were

compiled in Texmaker. It was replicated on two alternative computers to ensure the results are identical and the files can be compiled.

1. Run the file "Rscripts/clean_data.R" to obtain all R data files

2. Run the file "Rscripts/analysis.R" to obtain all the figures and tables.

3. Compile the Appendix file "Paper/Dehdari_Gehring_Appendix.tex" by running Latex, then Bibtex, then Latex twice.

4. Compile "Paper/Dehdari_Gehring_The_origins_of_common_identity.tex" by running Latex, then Bibtex, then Latex twice.

Data Availability Statement

Here we list the sources to all data sets used in the paper. We list the data sets that are provided in the replication package and for those that, due to various restrictions, are not, we provide detailed descriptions on how to access the data. All sources are also listed in Section F in the Online Appendix.

- Newspaper subscription data from *Le Republicain Lorraine* for 2014. The data was accessed through e-mail contact with the newspaper.
- Data on location of churches from Vlachos (2019). The data was graciously shared by the author.
- Data based on the Cahiers de Doléances. The data has been accessed from Hyslop (1968). The raw data set is provided in Stata_data/cahiers1789.xlxs.
- 1969 Constitutional referendum results from L'Est Republicain. The data was accessed through e-mail contact with the newspaper L'Est Republicain. The raw data set is provided in Stata_data/ref69_d_final.xlxs.
- French 1866 census data and French 1916-1946 population data from Talandier et al. (2016). The data sets were graciously shared by the authors.
- Historical railway data from Mimeur et al. (2018). The data was graciously shared by the authors.
- Historical road network data from Perret et al. (2015). The data was graciously shared by the authors.
- Data from Twitter on share of tweets about the German and French national football team. We include a list of the Twitter tweet IDs in other_data/twitter_ids.txt.
- Data on reported crime from L'Express. The data was accessed through e-mail contact with the newspaper L'Express. The raw data set is provided in Stata_data/Securite Alsace-Lorraine 25 fev 2015.xlxs.
- Several data sets used in this paper have been accessed through https://commande.progedo.fr/en/utilisateur/connexion. This includes municipal-level data from The National Institute of Statistics and Economic Studies (INSEE) (2006-2013), Observatoire Interrégional du Politique (OIP) Survey data (1999-2003), Regional election results (2015), Center for socio-political data (CDSP) referenda results (1994 and 2005), ISSP (2003). The data can be accessed by creating an account at the website and making a request. The available data sets are searchable and multiple data sets can be added to each request. Once the request has

been processed, the user is provided with a link from where the data can be downloaded. The names of the raw data sets are:

- CDSP: cdsp_refe1992_comm, cdsp_refe2005_comm, cdsp_presi2007t1_comm, cdsp_presi2007t2_comm, cdsp_legi2002t1_comm-m3500, cdsp_muni1983t1_commp9000-ag, cdsp_muni1989t2_commp9000-ag, cdsp_muni1995t1_commp9000-ag, cdsp_muni1995t2_commp9000-ag, cdsp_muni2001t1_commp9000-ag, cdsp_muni2001t2_commp9000-ag, cdsp_muni2008t1_commp3500, cdsp_muni2008t2_commp3500, cdsp_muni2001t1_comm-m3500, cdsp_muni2001t1_comm3500-9000, cdsp_muni2001t2_comm-m3500, cdsp_muni2001t2_comm3500-9000
- INSEE: fam01sc, fam1ss, mig01sc, mig2ss, 2014 VNO ABTS ER PAR COMMUNES, equip-serv-action-sociale-com-13, equip-serv-commerce-com, equip-serv-ens-1er-degre-com, equip-serv-ens-2eme-degre-com-2013, equip-serv-ens-sup-form-serv-com-13, equip-serv-medical-para-com, equip-serv-particuliers-com-13, equip-serv-sante-com-13, equip-sport-loisir-socio-com-13, equip-tour-transp-com-13, base-cc-diplomes-formation-2010, HIST_POP_COM_RP11, base-cc-demo-entreprises-11, base-cc-demo-entreprises-12, base-cc-demo-entreprises-13, base-cc-evol-struct-pop-2011, base-cc-irpp-nouv-serie-09, RFDM2008COM, reg15_northeast, Reg_15_Resultats_Communes_T1_c
- ISSP: National Identity II - ISSP 2003
- OIP: fr_cdsp_ddi_oip_1999_F1, fr_cdsp_ddi_oip_2000_F1, fr_cdsp_ddi_oip_2001_F1, fr_cdsp_ddi_oip_2003_F1, fr.cdsp.ddi.PEFV2_F1

The following GIS data shapefiles were used in GIS, using georeferencing tools, creating maps about the borders, administrative division and depiction of the outcomes. The processed GIS data files can be found in Stata_data and are named AL_railway.dta, AL_rivers_length.dta, AL_road_length.dta, border_departements_distance.xlxs, Dep_Closeness_to_Border.xlxs, FRA_adm5.xlxs, FRA_mun_cent_border_points.xlxs, FRA_mun_cent_raw.xlxs, Is_maginot_Lorraine_Alsace.dta. A detailed description of the methods used is provided in Section E in the Online Appendix.

- GIS shapefiles for French administrative borders. The data can be accessed by visiting https://gadm.org/download_country_v2.html, choosing "France" in the dropdown menu, and picking "Shapefile" (version 2.8).
- GIS data on historical language borders. The data has been geoprocessed from Harp (1998).
- GIS data on the Maginot Line. The data has been geoprocessed from https://upload.wikimedia.org/wikipedia/commons/8/8b/1940-Fall_Gelb.jpg.
- NASA SRTM elevation data. The data and documentation can be found at https://www2.jpl.nasa.gov/srtm/. The data set can also be access through ArcGIS, by adding data from ArcGIS online. The data set is called "Elevation 90m".
- Geographic farming and grazing data from Klein Goldewijk et al. (2011) is used in this paper. The data can be accessed by downloading the baseline.zip file from https://dataportaal.pbl.nl/downloads/HYDE/HYDE3.2/.
- Data on ruggedness from Nunn and Puga (2012) is used in this paper. The data can be accessed
- from https://diegopuga.org/data/rugged/.
- Data on soil suitability for crops from the Food and Agricultural Organization of the United Nations is used in this paper. The data can be accessed from http://www.fao.org/nr/gaez/en/.

References

Harp, S. L. (1998): Learning to be Loyal: Primary Schooling as Nation Building in Alsace and Lorraine, 1850-1940., De Kalb: Northern Illinois University Press.

Mimeur, C., F. Queyroi, A. Banos, and T. Thévenin (2018): "Revisiting the structuring effect of transportation infrastructure: an empirical approach with the French Railway Network from 1860 to 1910," Historical Methods: A Journal of Quantitative and Interdisciplinary History, 51, 65–81.

Nunn, N. and D. Puga (2012): "Ruggedness: The Blessing of Bad Geography in Africa," Review of Economics and Statistics, 94, 20–36.

Perret, J., M. Gribaudi, and M. Barthelemy (2015): "Roads and cities of 18th century France," Scientific data, 2, 1–7.

Talandier, M., V. Jousseaume, and B.-H. Nicot (2016): "Two centuries of economic territorial dynamics: the case of France," Regional studies, Regional science, 3, 67–87.

Vlachos, S. (2019): "On war and political radicalization," Vienna Economics Papers Nr. 1904.