

Overview

The codes and datasets in this replication package construct the analyses of “Kin Networks and Institutional Development” from 7 data sets using Stata 15.1. The replication file is structured into 6 sub-folders along those datasets. Each sub-folder contains a do-file and the accompanying data set. To replicate a table, open a do-file in the respective subfolder. The following is the list of folders and the associated regression outputs:

01_Country_Table1: Country dataset to replicate **Table 1** and appendix tables B1, B2, B3, & B4.

02_Ethnicity_Table2: Ethnicity dataset to replicate **Table 2** and appendix table B6.

03_Commune_Table3_Table4: Two historical city-level data sets.

- (i) Panel dataset to replicate **Table 3** and appendix tables C1, C2, C3, C4, C5, C7, & C8.
- (ii) City cross-sectional data set to replicate **Table 4** and appendix tables C9 & C10.

04_EuropeanRegions_Table5: Regional dataset to replicate **Table 5** and appendix tables B5 & D1.

05_Immigrants_Table6: contains two data sets.

- (i) 01_Table6: second-generation immigrants’ dataset to replicate **Table 6** and app. tables D3.
- (ii) 02_TableD4: second-generation immigrants’ dataset (based on parents originating ethnicity) to replicate app. table D4.

06_Appendix_Italy: Italian provincial dataset to replicate appendix table D2.

07_Appendix_HistoricalCountry: Historical country dataset to replicate app. table C6.

Data Availability and Provenance Statements

There are several sources of data in this paper. How they are constructed is described in the manuscript and the appendix. Data sources are detailed below.

Statement about Rights

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Summary of Availability

All data are publicly available.

Details on each Data Source

The replication package contains the following data sets (in the stata format: `.dta` and in `.csv`) in the respective sub-folders. Data sources for all indicators are also listed.

01_Country_Table1\data: **CountryData.dta**.

It consists of various publicly available country-level indicators. Sources:

Adherence to major religions: Taken from Barro and McCleary (2003) for the year 2000. Retrieved on March 14, 2016 from https://scholar.harvard.edu/barro/data_sets. Adherents in a country to Catholicism, Protestantism, Orthodox Christianity, other Christian denominations, Islam, Hinduism and Buddhism, as fractions of the country's population.

Absolute latitude: Taken from Ashraf and Galor (2013). The absolute latitude of a country's approximate geodesic centroid, as reported by the CIA's World Factbook.

Caloric suitability: Using data from Galor and Özak (2016), the Caloric Suitability Index captures the average potential agricultural output (measured in calories) based on crops that were available for cultivation after 1500 CE. Caloric Suitability therefore captures the variation in potential crop yield across the globe, as accounted for by calories per hectare per year. The Caloric Suitability Index is constructed based on data from the Global AgroEcological Zones (GAEZ) project of the Food and Agriculture Organization (FAO). The GAEZ project supplies global estimates of crop yield for 48 crops in grids with cell sizes of 5-degree cells. We use the medium level rain-fed potential output.

Caloric suitability for oat: Using data from Galor and Özak (2016), the Caloric Suitability Index for oats captures the medium level rain-fed potential agricultural outputs (measured in calories) of oat.

Caloric suitability for rye: Using data from Galor and Özak (2016), the Caloric Suitability Index for rye captures the medium level rain-fed potential agricultural outputs (measured in calories) of rye.

Genetic heterogeneity (ancestor adjusted): Based on Ashraf and Galor (2013). The expected heterozygosity (genetic diversity) of a country's population, predicted by migratory distances from East Africa (i.e., Addis Ababa, Ethiopia). It is a prediction based on the worldwide sample of 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel. The measure is ancestor adjusted using the World Migration Matrix, 1500-2000 CE, from Putterman and Weil (2010).

Church exposure: constructed by the author (see manuscript for details)

Cousin-marriage preference: Based on the Ethnographic Atlas compiled by D-PLACE (Kirby et al., 2016) and aggregated to the country level by Bahrami-Rad, D. et al. (2021).

Cousin-term differentiation: Based on the Ethnographic Atlas compiled by D-PLACE (Kirby et al., 2016) and aggregated to the country level by Bahrami-Rad, D. et al. (2021).

Irrigation potential: Taken from Bentzen et al. (2017). Irrigation potential measures the fraction of land that would have experienced at least a doubling of yields if irrigation were to be introduced into an area where agriculture was previously rainfed. The measure is in relation to all land suitable for agriculture. The measure is based on data from the global Agro-Ecological Zones (GAEZ) 2002 database of the Food and Agriculture Organization.

Log GDP per capita: Real GDP per capita in 2000 CE, in international dollars (adjusted for Purchasing Power Parity), as reported by the Penn World Table, version 6.2. Natural logs are taken.

Distance to navigable waterways: Taken from Gallup, Sachs and Mellinger (1999). The distance, in thousands of km, from a GIS grid cell to the nearest ice-free coastline or sea-navigable river, averaged across the grid cells of a country. It is part of Harvard University's CID Research Datasets on General Measures of Geography.

Parasite stress: The measure is Fincher and Thornhill (2012)'s combined parasite-stress indicator (both non-zoonotic and zoonotic parasites). It is based on the GIDEON database (Global Infectious Disease & Epidemiology Network; www.gideononline.com).

Percent cousin marriage: Bittles as published on www.consang.net (see details in the manuscript)

Ruggedness: Taken from Nunn and Puga (2012). At one (grid-cell level) point, the index is given by the square root of the sum of the squared differences in elevation between the central point and the eight adjacent points. The country-level indicator is the average across all the grid cells within a country.

Timing of Neolithic Transformation (Ancestor adjusted): The number of years elapsed, up to the year 2000 CE, since the majority of the population residing within a country's modern national borders began practicing sedentary agriculture as the primary mode of subsistence. This measure is based on Putterman (2008). It is compiled using a wide variety of both region- and country-specific archaeological studies as well as more general encyclopedic works on the transition from hunting and gathering to agriculture during the Neolithic

Revolution. We use the ancestry adjusted indicator to take account of migration post-1500 CE. The ancestry weights are obtained from the World Migration Matrix of Putterman and Weil (2010).

Tropical area: Taken from Nunn and Puga (2012). It is based on Kottke et al. (2006), who classify each cell on a 30 arc-minute grid covering the entire land area of the Earth into one of 31 climates in the widely used Köppen-Geiger climate classification (these categories are formed using temperature and precipitation data from the Climatic Research Unit of the University of East Anglia and the Global Precipitation Climatology Centre of the German Weather Service). Based on these data, Nunn and Puga (2012) calculated the percentage of the land surface area of each country that has any of the four Köppen-Geiger tropical climates.

02_Ethnicity_Table2\data: *EthnographicAtlas.dta*.

It consists of the Ethnographic Atlas as prepared by D-Place (Kirby et al. 2016) plus data on Deep Christianization prepared by Korotayev (2003).

03_Commune_Table3_Table4\data: *IncestLegislation.dta & urbanflexible.dta*.

This dataset is based on data by Bosker et al. (2013), Bairoch (1988), and data on Church exposure constructed by myself (detailed in the manuscript)

04_EuropeanRegions_Table5\data: *EuropeanRegions2.dta*.

This data set is based on the European Social Survey. It is merged with historical data on Church exposure, data on Cousin marriage, and other historical data (detailed in the manuscript).

Absolute latitude: Absolute latitude of the centroid of a region.

Caloric suitability: Taken from Galor and Özak (2016), the Caloric Suitability Index captures the average potential agricultural output (measured in calories) based on crops that were available for cultivation after 1500 CE. Caloric Suitability therefore captures the variation in potential crop yield across the globe, as accounted for by calories per hectare per year. The Caloric Suitability Index is constructed based on data from the Global AgroEcological Zones (GAEZ) project of the Food and Agriculture Organization (FAO). The GAEZ project supplies global estimates of crop yield for 48 crops in grids with cell sizes of 5-degree cells, which allowed us to construct regional indicators. We use the medium level rain-fed potential output.

Caloric suitability for oat: Using data from Galor and Özak (2016), the Caloric Suitability Index captures the medium level rain-fed potential agricultural outputs (measured in calories) of oat.

Caloric suitability for rye: Using data from Galor and Özak (2016), the Caloric Suitability Index captures the medium level rain-fed potential agricultural outputs (measured in calories) of rye.

Carolingian Empire: Based on Shepherd's map (1911), this variable indicates the areal fraction of a region that fell within the boundaries of the Carolingian Empire in the year 814 CE.

Distance to the coast: Distance of the centroid of a region from the coast, constructed based on a coastline physical vector map in 1:10m resolution. Source: Natural Earth (<http://www.naturalearthdata.com/>).

Elevation: Mean elevation is constructed based on the global map (30 by 30 arcsecond cells) obtained from Global 30 Arc-Second Elevation dataset. Source: GTOPO30 dataset (<https://lta.cr.usgs.gov/GTOPO30>).

Lake or rivers (presence of): A dummy variable indicating whether there is a river or lake within a region. Rivers primarily derive from World Data Bank 2. Data of Europe primarily derives from Catchment Characterization and Modelling (CCM) Database 2.1 by the European Commission, Joint Research Centre, Institute for Environment and Sustainability. Data for North America derives the North American Environmental Atlas, a collaboration of government agencies in Canada, Mexico and the United States and the trilateral Commission for Environmental Cooperation (CEC). Source: Natural Earth (<http://www.naturalearthdata.com/>).

Monastic presence. The monastic presence indicator is based on five different monastic orders: Cluniac, Cistercians, Premonstratensians, Franciscans, and Dominicans. For each order we drew 50km radii around all the monastic houses of the corresponding order. Based on the pixels that fall within the 50km radii I calculated the percentage of each region that was exposed to at least one order at some point in time. The underlying geo-located data on Cluniac (existing between before 998 to 1109 or later), Dominican (existing

between 1216 to 1500), Franciscan (existing around 1300), and Premonstratensians Houses (existing between 1120 to 1500) are based on the Atlas zur Kirchengeschichte (Hubert et al. 1980) and are taken from the Digital Atlas of Roman and Medieval Civilization (DARM). Cistercian Houses (existing between 1095 and 1675) are based on Donkin (1978) and are taken from Andersen et al. (2017).

Roman roads: Using data from McCormick et al., this measure captures the length of Roman roads within a region (as identified in the Barrington Atlas) per area of the region.

Ruggedness: The regional measure is constructed based on the global map (30 by 30 arc-second cells) obtained from the grid-cell-level data on ruggedness based on Nunn and Puga (2012). For details, see the country-level indicator.

Socialist history: Indicator variable capturing whether a European region has a socialist history (see map in Figure S2 in the main text). Regions that belonged to Yugoslavia are coded as having a socialist history, even though they were not part of the Warsaw pact. (This follows Churchill's original (1948) statement on the "Iron curtain".)

Temperature: The means of the entire annual cycles of temperature is constructed for the period between 1901 and 2014 CE based on monthly global maps (0.5 by 0.5 degree cells) obtained from the CRU-TS 3.1 Climate Database. Source: Harris et al. (2013).

Population density in 500 CE: Taken from Goldewijk et al. (2010), this measures population density of a region in the year 500 CE. These estimates are based on the country estimates by McEvedy and Jones (1978), broken down to the pixel level according to geographic factors that relate to the probability of settlement (e.g., proximity to waterways, temperature).

Percent cousin marriage: The data was compiled by population geneticists who were granted access to archives of the Catholic Church and contains data for Spain (average of years 1911 to 1943, Pinto-Cisternas et al., 1979), Italy (average of years 1910 to 1964, Cavalli-Sforza et al., 2004), and France (average of years 1926 to 1958, Sutter and Goux, 1964). I augmented this data by Turkish cousin-marriages rates based on the second wave of the Demographic and Health Survey (year 1998).

Precipitation: The means of the entire annual cycles of precipitation constructed for the time period between 1901 and 2014 CE. Based on monthly global maps (0.5 by 0.5 degree cells) obtained from the CRU-TS 3.1 Climate Database. Source: Harris et al. (2013).

05_Immigrants_Table6: contains two data sets.

- (i) 01_Table6\data: ***ESS_immigrants.dta***. This data set is based on the European Social Survey (wave 2-8). It is merged with historical data on Church exposure and country level indicators of the originating countries of the parents of second-generation immigrants (detailed in the manuscript). The sources of the country level indicators are detailed above under "01_Country_Table1"
- (ii) 02_TableD4\data: ***SecondGen_AncetralEthnicity.dta***. This data set is based on the European Social Survey (wave 2-8). It is merged with historical data on Church exposure and country level indicators of the originating countries of the parents of second-generation immigrants, and ethnicity level characteristics of parents' ancestral ethnicity based on the Ethnographic Atlas (Kirby et al, 2016).

06_Appendix_Italy\data: ***Italydata.dta***.

This dataset is constructed based on cousin marriage data (Cavalli-Sforza et al., 2004) and data of Guiso et al. (2004)

07_Appendix_HistoricalCountry\data: ***HistoricalCountry.dta***.

Historical country dataset (detailed in the manuscript). Data sources are McEvedy and Jones (1978), while the data on urban population is calculated based on Bairoch et al. (1988).

Computational requirements

To conduct the replication Stata version 15.1 is required. There are a few user-generated commands. One is an ado file by Jean-Pierre Dube which is needed to calculate Conley standard errors based on genetic distance and which I slightly adopted. The ado file “x_ols2” is saved in the folder “01_Country_Table1”. Other user-written commands such as the “asrec” command by Colella et al. (2019) need to be installed.

Software Requirements

The replicate package consists of a total of 8 do-files. In total, they will take about one hour to run.

Description of programs/code

For each dataset there is one do-file attached.

Instructions to Replicators

The replication file is structured into 6 sub-folders. Each sub-folder contains a do-file and an accompanying dataset. To replicate a table open and run a do-file in the respective subfolder. The output for each do-file will be stored in a sub-sub-folder called “tables”.

List of tables and programs

The following is the list of folders and the associated regression outputs. Note that Tables are not always saved in one but multiple files (reflecting that tables contain multiple panels). Each output is clearly named such that it is clear which table and which panel of a table is contained in a particular output file.

01_Country_Table1: Country dataset to replicate **Table 1** and appendix tables B1, B2, B3, & B4.

02_Ethnicity_Table2: Ethnicity dataset to replicate **Table 2** and appendix table B6.

03_Commune_Table3_Table4: Two historical city-level data sets.

- (i) Panel dataset to replicate **Table 3** and appendix tables C1, C2, C3, C4, C5, C7, & C8.
- (ii) City cross-sectional data set to replicate **Table 4** and appendix tables C9 & C10.

04_EuropeanRegions_Table5: Regional dataset to replicate **Table 5** and appendix tables B5 & D1.

05_Immigrants_Table6: contains two data sets.

- (i) 01_Table6: second-generation immigrants' dataset to replicate **Table 6** and app. tables D3.
- (ii) 02_TableD4: second-generation immigrants' dataset (based on parents originating ethnicity) to replicate app. table D4.

06_Appendix_Italy: Italian provincial dataset to replicate appendix table D2.

07_Appendix_HistoricalCountry: Historical country dataset to replicate app. table C6.

References

- Andersen, T. B., Bentzen, J., Dalgaard, C. J., Sharp, P. "Pre-reformation roots of the protestant ethic," *Economic Journal*, 127 (2017), 1756–1793.
- Ashraf, Quamrul, and Oded Galor, "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103 (2013), 1-46.
- Bairoch, Paul, Jean Batou, and Pierre Chèvre, *La Population des villes Europeennes de 800à 1850: Banque de Données et Analyse Sommaire des Résultats*. (Geneva: Centre d'histoire économique Internationale de l'Université de Genève, Librairie Droz, Geneva., 1988).
- Barro, Robert J., and Rachel M. McCleary, "Religion and Economic Growth across Countries," *American Sociological Review*, 68 (2003), 760-781.
- Bentzen, Jeanet Sinding , Nicolai Kaarsen, and Asger Moll Wingender (2017). "Irrigation and Autocracy," *Journal of the European Economic Association*,
- Bosker, Maarten, Eltjo Buringh, and Jan Luiten van Zanden, "From Baghdad to London: Unraveling Urban Development in Europe, the Middle East, and North Africa 800-1800," *Review of Economics and Statistics*, 95 (2013), 1418-1437.
- Colella, F., Lalive, R., Sakalli, S. O., Thoenig, M. (2019). 'Inference with arbitrary clusters', IZA discussion paper No 12584
- Donkin, R., *The Cistercians: Studies in the Geography of Medieval England and Wales* (Pontifical Institute of Medieval Studies, Toronto, 1978).
- Droysen, Gustav, *Allgemeiner historischer Handatlas* (Bielefeld: Velhagen & Klasing, 1886)
- ESS Round 9: European Social Survey Round 9 Data (2018). Data file edition 3.1. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS9-2018.
- ESS Round 8: European Social Survey Round 8 Data (2016). Data file edition 2.2. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS8-2016.
- ESS Round 7: European Social Survey Round 7 Data (2014). Data file edition 2.2. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS7-2014.
- ESS Round 6: European Social Survey Round 6 Data (2012). Data file edition 2.4. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS6-2012.
- ESS Round 5: European Social Survey Round 5 Data (2010). Data file edition 3.4. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS5-2010.
- ESS Round 4: European Social Survey Round 4 Data (2008). Data file edition 4.5. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS4-2008.
- ESS Round 3: European Social Survey Round 3 Data (2006). Data file edition 3.7. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS3-2006.
- ESS Round 2: European Social Survey Round 2 Data (2004). Data file edition 3.6. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS2-2004.

- Fincher, Corey L., and Randy Thornhill, "Parasite-stress promotes in-group assortative sociality: the cases of strong family ties and heightened religiosity," *Behavioral and Brain Sciences*, 35 (2012), 61-119.
- Gallup, John L., and Jeffrey D. Sachs, "The economic burden of malaria," *American Journal of Tropical Medicine and Hygiene*, 64 (2001), 85-96.
- Galor, Oded, and Ömer Özak, "The Agricultural Origins of Time Preferences," *American Economic Review*, 106 (2016), 3064-3103.
- Goldewijk, K. K. , Beusen, A., van Drecht, G., "The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12 ,000 years." *Glob. Ecol. Biogeogr.* 20 (2010), 73–86.
- Guiso, Luigi, Paola Spanienza, and Luigi Zingales, "The Role of Social Capital in Financial Development," *American Economic Review*, 94 (2004), 526-556.
- Harris, I., Jones, P. D., Osborn, T. J. , Lister, D. H., "Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset." *Int. J. Climatol.* (2013), doi:10.1002/joc.3711.
- Hubert, J., Latourette, K. S., Martin, J., *Atlas zur Kirchengeschichte, die Christlichen Kirchen in Geschichte und Gegenwart* (Herder, Freiburg, 1980).
- Kirby, K.R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H-J, Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R., Lee, D., Low, B. S., McCarter, J., Divale, W. and Gavin, M. C. (2016). D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity. *PLoS ONE*, vol. 11(7): e0158391.
- Korotayev, Andrey V., "Unilineal Descent Organization and Deep Christianization: A Cross-Cultural Comparison," *Cross-Cultural Research*, 37 (2003), 133-157.
- Kottek, M. J. Grieser, C. Beck, B. Rudolf, F. Rubel, World map of the Köppen-Geiger climate classification updated. *Meteorol. Zeitschrift*. 15, 259–263 (2006).
- McCormick, M., et al. The Digital Atlas of Roman and Medieval Civilizations. <https://darmc.harvard.edu/>
- McEvedy, C., Jones, R., *Atlas of world population history* (Penguin, New York, 1978).
- Nunn, Nathan, and Diego Puga, "Ruggedness: the Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 94 (2012), 20-36.
- Putterman, Louis, "Agriculture, Diffusion, and Development: Ripple Effects of the Neolithic Revolution," *Economica*, 75 (2008), 729-748.
- Putterman, Louis, and David N Weil, "Post-1500 Population Flows and the Long Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, 125 (2010), 1627-1682.
- Shepherd, W., *Historical atlas* (Henry Holt and Company, New York, 1911).
- Spolaore, E. and Wacziarg, R. "Ancestry and development: new evidence" *Journal of Applied Econometrics*, 33(5), (2018).
- Wahl, Fabian, "Participate Political Institutions in Pre-Modern Europe. Introducing a New Database," *Historical Methods. A Journal of Quantitative and Interdisciplinary History*, 49 (2016), 67-79.