README file

Forbidden Fruits:
The Political Economy of Science, Religion, and Growth

By Roland Bénabou, Davide Ticchi, Andrea Vindigni

Review of Economic Studies, 2021

Overview

The folder "Datafortheweb" contains the data and program files required to replicate the tables and figures reported in the main text and in the Online Appendix F of the paper. We here below report a brief description of each folder and its contents. Then, we provide the Data Availability and Provenance Statements containing a description of the origin, location and accessibility of the data used in the article. It follows a description of the contents of the data and program files. Online Data Appendix G complements the information provided in this README file with a detailed description for each variable used in the empirical analysis and the data source. Program files also contain additional comments.

Folder content:

- **"Data"**: the folder is divided into 3 subfolders.
    - **(a) "Source"**: it includes all the datasets recalled by the do files and required to replicate the analysis. The subfolder in turn contains two subfolders named:
      - **"ICC"**, which includes the datasets used to generate the variables of the international cross-country analysis;
      - **"USA"**, which includes the datasets employed to generate the variables of the US cross-state analysis.
    - **(b) "Temp"**: empty folder that will include all the datasets generated by and recalled by the do files to construct the two final datasets used to replicate the analysis.
    - **(c) "Final"**: empty folder that will include the two datasets (one for the international cross-country analysis and one for the US cross-state analysis) generated by the do files and required to replicate the tables and the figures.
- **"Programs"**: includes all the Stata do files required to replicate the results.
- **"Results"**: empty folder that works as a repository for the Tables and Figures.

Instructions:

Programs have been written using Stata 14.

Preliminary set-up before running the programs:

Install the following accessory user-written commands (in Stata, type command *ssc install* followed by):
    estout
    coefplot

Replication:

(i) Open the file "master.do" and replace the path to the folder "…/Datafortheweb/" at line 18 with your own directory path.

(ii) The output Tables and Figures created by the do files are stored in the folder **"Results"**.

The approximate time needed to reproduce the analyses on a standard desktop machine is less than 5 minutes.

Data Availability and Provenance Statements

All data used in this paper are publicly available. As specified below in detail, all datasets necessary to replicate our analysis are contained in the folder "/Data/Source", either in the subfolder "ICC" or "USA". However, some of the data are reported in the form of an "extract" of the full dataset, i.e. the datafile provided contains only the data of the variables of the full dataset used in the paper. Specifically, this is the case for the European Values Study and the World Values Survey datasets that are subject to a redistribution restriction. However, EVS and WVS have both allowed the authors of the paper to upload an extract of their full datasets. As reported in detail below, regarding all other data employed in the analysis, the full dataset is provided either because this is allowed by the conditions of use or because the authors have been authorized by the owners of the data to redistribute the dataset within the replication package. The full Pew Forum on Religion & Public Life 2007 datafiles are provided in Stata format rather than in the SPSS format that is the one available in the PEW website.

Details on each Data Source for the international cross-country analysis

EVS (2020). European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008). GESIS Data Archive, Cologne. ZA4804 Data file Version 3.1.0, https://doi.org/10.4232/1.13486. (accessed July 16, 2021).
Datafile: "ZA4804_v3-1-0.dta" (not provided)
        "ZA4804_v3-1-0_extract.dta" (the extracted datafile is provided)
Folder: "/Data/Source/ICC"
Data is subject to a redistribution restriction but can be freely downloaded from the EVS website. Specifically, the full dataset can be accessed starting from the Data Download section of the webpage
https://europeanvaluesstudy.eu/
that allows to go at the webpage with title "ZA4804: European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008)" at the URL
https://dbk.gesis.org/dbksearch/sdesc2.asp?db=e&no=4804.
Then, click on the link "ZA4804 Downloads and Data Access" to go in the download page of the required file.
The URL for direct download (at the date of the last access) is:
https://search.gesis.org/research_data/ZA4804
The data download requires to fill out the registration form and agree to the conditions of use. Save the file "ZA4804_v3-1-0.dta" in the directory "/Data/Source/ICC". Then, follow the instructions contained at the beginning of the file "/Programs/2_evs.do" to generate the extracted datafile "ZA4804_v3-1-0_extract.dta" with only the data necessary for the analysis; the extracted file has to be saved in the folder "/Data/Source/ICC".

The datafile "ZA4804_v3-1-0_extract.dta" is provided as part of this archive in the directory "/Data/Source/ICC". Therefore, the replication of the analysis does not require any download of the data from the EVS website.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014. World Values Survey: All Rounds – Country-Pooled Datafile Version: http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp. Madrid: JD Systems Institute. (accessed January 27, 2020).
Datafile: "WVS_Longitudinal_1981_2016_stata_v20180912.dta" (not provided)
         "WVS_Longitudinal_1981_2016_stata_v20180912_extract.dta"   (the extracted datafile is provided)
Folder: "/Data/Source/ICC"
Data is subject to a redistribution restriction but can be freely downloaded from http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp.
Choose "WVS_Longitudinal_1981_2016_stata_v20180912.dta", fill out the registration form, including a brief description of the project, and agree to the conditions of use. Save the file in the directory "/Data/Source/ICC". Then, follow the instructions contained at the beginning of the file "/Programs/1_wvs.do" to generate the extracted datafile "WVS_Longitudinal_1981_2016_stata_v20180912_extract.dta" with only the data necessary for the analysis; the extracted file has to be saved in the folder "/Data/Source/ICC".
The datafile "WVS_Longitudinal_1981_2016_stata_v20180912_extract.dta" is provided as part of this archive in the directory "/Data/Source/ICC". Therefore, the replication of the analysis does not require any download of the data from the WVS website.

Barro, Robert and Jong-Wha Lee (2013) "A New Data Set of Educational Attainment in the World, 1950-2010." Journal of Development Economics, 104, 184–198. (accessed December 28, 2015).
Datafile: "BL2013_MF1599_v2.0.dta"
Folder: "/Data/Source/ICC"
The data are in the public domain and available at:
http://www.barrolee.com/
The file "BL2013_MF1599_v2.0.dta" is provided as part of this archive in the folder "/Data/Source/ICC".

Norris, Pippa (2009). "Democracy Cross-national Data. Release 3.0 Spring 2009." (accessed March 8, 2013).
Datafile: "Democracy Crossnational Data Spring 2009 StataSE.dta"
Folder: "/Data/Source/ICC"
Data publicly available at the author's webpage:
http://www.pippanorris.com/
Data can be freely downloaded from the Pippa Norris's webpage but is subject to a redistribution restriction. However, the datafile "Democracy Crossnational Data Spring 2009 StataSE.dta" is provided as part of this archive in the folder "/Data/Source/ICC" as the authors have been authorized by Pippa Norris to redistribute it within the replication package.

Park, Walter G. (2008) "International patent protection: 1960–2005." Research Policy, 37(4), 761–766. https://doi.org/10.1016/j.respol.2008.01.006. (accessed January 28, 2020).
Datafile: "Patent index1960 - 2015.xlsx"
Folder: "/Data/Source/ICC"
Data publicly available at the author's webpage:

http://fs2.american.edu/wgp/www/?_ga=2.150063158.1045324815.1586191710-954683830.1586191710

Data can be freely downloaded from the Walter Park's webpage but is subject to a redistribution restriction. However, the datafile "Patent index1960 - 2015.xlsx" is provided as part of this archive in the folder "/Data/Source/ICC" as the authors have been authorized by Walter Park to redistribute it within the replication package.

World Development Indicators. Washington, D.C., The World Bank. (accessed January 27, 2020).
Datafile: "WDIData.csv"
Folder: "/Data/Source/ICC"
The data are in the public domain and available at:
http://data.worldbank.org/data-catalog/world-development-indicators
Data can be downloaded from the webpage with the URL above reported. Choose all variables and countries for the period 1960 to 2019 and download the csv file. The file "WDIData.csv" has to be saved in the folder "/Data/Source/ICC". A copy of the data is provided as part of this archive.

World Intellectual Policy Organization (WIPO) statistics database. "Intellectual property right: Patent". Last updated: October 2019. (accessed January 27, 2020).
Datafile: "patent_1980_2018.csv"
Folder: "/Data/Source/ICC"
The data are in the public domain and available at:
https://www3.wipo.int/ipstats/index.htm?tab=patent
Data can be freely downloaded from the webpage with the URL reported above. Choose "Patent" and the following entries: "Indicator: 1 - Total patent applications (direct and PCT national phase entries)", "Report type: Total count by filing office", "Year range: 1980 – 2018", "Office: all countries", "Type: Non-resident, Resident". Then, download the csv file "patent_1980_2018.csv" and save it in the folder "/Data/Source/ICC". A copy of the data is provided as part of this archive.

Wikipedia (2020) "Communist state", April 1, 2020. https://en.wikipedia.org/wiki/Communist_state. (accessed April 7, 2020).
Datafile: "14_communism_labels.do"
Folder: "/Programs"
The data are in the public domain and available at:
https://en.wikipedia.org/wiki/Communist_state
Data for the variable "communist" is imported in the final dataset using the do program file "14_communism_labels.do" contained in the folder "/Programs". Further details about the extraction of information from the above cited Wikipedia's webpage is reported in Online Appendix G.

Details on each Data Source for the US cross-state analysis

Pew Forum on Religion and Public Life (2007). "U.S. Religion Landscape Survey". (accessed November 3, 2013).
Datafiles: "Religious Landscape Survey Data - Continental US.dta"
         "Religious Landscape Survey Data - Alaska and Hawaii.dta"
Folder: "/Data/Source/USA"
Data is subject to a redistribution restriction but can be freely downloaded from the Pew Research Center website at:
https://www.pewforum.org/dataset/u-s-religious-landscape-survey/
Specifically, the data download requires to fill out the registration form and agree to the conditions of use. The file downloaded is named:
"dataset_Religious_Landscape_Survey_Data.zip"

and it contains (together with the documentation files in Word format) the following two datafiles in SPSS format:
"Religious Landscape Survey Data - Continental US.sav" and
"Religious Landscape Survey Data - Alaska and Hawaii.sav".
These two files need to be transformed in Stata format and saved in the directory "/Data/Source/ICC".
While the data is subject to a redistribution restriction, the two files (transformed by the authors) in Stata format
"Religious Landscape Survey Data - Continental US.dta"
"Religious Landscape Survey Data - Alaska and Hawaii.dta"
are provided as part of this archive in the directory "/Data/Source/ICC" as the authors have been authorized by the Pew Research Center to redistribute the data within the replication package.

U.S. Bureau of Economic Analysis, "Gross Domestic Product by State". Last updated: November 7, 2019. https://www.bea.gov/data/gdp/gdp-state. (accessed February 13, 2020).
Datafiles: "Per capita real GDP by state 2006 2007.xls"
          "Gross domestic product (GDP) 2006 2007.xls"
Folder: "/Data/Source/USA"
The data are in the public domain and available at:
https://www.bea.gov/data/gdp/gdp-state
Data can be freely downloaded from the webpage with the URL reported above. Choose to download "Per capita real GDP by State (Chained 2012 dollars) (SAGDP10N)" and "Gross domestic product (GDP) by State (SAGDP2N) (All industry total)" for the years 2006 and 2007. Save the files respectively with names "Per capita real GDP by state 2006 2007.xls" and "Gross domestic product (GDP) 2006 2007.xls" in the folder "/Data/Source/USA". A copy of the data is provided as part of this archive.

U.S. Bureau of Economic Analysis, "Foreign Direct Investment in the United States (FDIUS): 2006 Data Tables". https://www.bea.gov/international/foreign-direct-investment-united-states-fdius-2006-data-tables. (accessed February 13, 2020).
Datafile: "Tab III.D11.xls"
Folder: "/Data/Source/USA"
The data are in the public domain and available at:
https://www.bea.gov/international/foreign-direct-investment-united-states-fdius-2006-data-tables
Data can be freely downloaded from the webpage with the URL reported above. Choose to download "complete set of the Revised 2006 Data Tables (ZIP)" in Section "Revised 2006 Data Tables". Then, unzip the folder and save the file "Tab III.D11.xls" containing "Table III.D11. Gross Property, Plant, and Equipment of Affiliates, State by Use" in the folder "/Data/Source/USA". A copy of the data is provided as part of this archive.

U.S. Census Bureau, "Table 1. Intercensal Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2000 to July 1, 2010 (ST-EST00INT-01)". Last Revised: November 30, 2016. https://www.census.gov/data/tables/time-series/demo/popest/intercensal-2000-2010-state.html. (accessed February 13, 2020).
Datafile: "st-est00int-01.xls"
Folder: "/Data/Source/USA"
The data are in the public domain and available at:
https://www.census.gov/data/tables/time-series/demo/popest/intercensal-2000-2010-state.html
Data can be freely downloaded from the webpage with the URL reported above "State Intercensal Tables: 2000-2010". Choose to download "Intercensal Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April

1, 2000 to July 1, 2010" and save the file "st-est00int-01.xls" in the folder "/Data/Source/USA". A copy of the data is provided as part of this archive.

U.S. Census Bureau (2009), "Educational Attainment in the United States: 2007". Issued January 2009, P20-560. (accessed February 12, 2020).
Datafiles: "p20-560.pdf"
          "ba_US.xls"
Folder: "/Data/Source/USA"
The publication is in the public domain and available at:
https://www.census.gov/library/publications/2009/demo/p20-560.html
Download the PDF file "p20-560.pdf" from the link "Download Educational Attainment in the United States: 2007 [PDF - <1.0 MB]" in the webpage with the URL reported above. The data of the share (in percentage term) of the State population with bachelor's degree or more used for the analysis are contained in "Table 2. Educational Attainment for the Population Aged 25 and Over by Region, State, and Nativity Status: 2007" at page 8 of the file "p20-560.pdf". These data have been reported by the authors in the Excel file "ba_US.xls" that is saved in the folder "/Data/Source/USA". A copy of the data is provided as part of this archive.

U.S. Patent and Trademark Office, Electronic Information Products Division, Patent Technology Monitoring Team (PTMT) (2007). "Calendar Year 2007 Patent Counts by Patent Type and by State and Country of Origin (01-Jan-2007 to 31-Dec-2007)". http://www.uspto.gov/web/offices/ac/ido/oeip/taf/st_co_07.htm.   (accessed November 4, 2013).
Datafile: "patents_US.csv"
Folder: "/Data/Source/USA"
The data are in the public domain and available at:
http://www.uspto.gov/web/offices/ac/ido/oeip/taf/st_co_07.htm
The data of the total number of patents submitted by residents of the State in 2007 are contained in the report available in the webpage with the URL reported above. The data have been reported by the authors in the csv format file "patents_US.csv" that is saved in the folder "/Data/Source/USA". A copy of the data is provided as part of this archive.

Dataset list: folder "Data"

We here report a brief description about the folder "Data" that contains three subfolders: "Final", "Source", "Temp". All source datafiles are contained in the subfolder "Source" that, in turn, includes two subfolders: "ICC" and "USA". The subfolder "Final" is empty and will contain the following two final datasets (generated by combining the source files) used for the analysis. The subfolder "Temp" is empty and will include all the datasets generated by and recalled by the do files to construct the two final datasets.

Description of the datasets in the subfolder "Source/ICC"

The subfolder "ICC" contains the source datasets used to generate the final dataset for the international cross-country analysis.

"codes.dta": The dataset comes from our elaboration. It contains a list with the name of each country, variable named "country_wv", that is taken from the World Values Survey (WVS) and the European Values Study (EVS) datasets, and a three-letter code, variable named "code2", that corresponds to "Country Code" of the World Development Indicators (WDI) dataset. The variable "code2" uniquely

identifies each country and is used to combine the various datasets in a unique manner as country names may contain minor differences in the various datasets to combine.

"WVS_Longitudinal_1981_2016_stata_v20180912_extract.dta": Extracted datafile with just the variables used of the World Values Survey longitudinal dataset containing the 6 waves for the period 1980-2010 and that is used to obtain the five religiosity variables, the shares of main religions and the dummy variables for the predominant religion of each country.

"ZA4804_v3-1-0_extract.dta": Extracted datafile with just the variables used of the European Values Study longitudinal dataset containing the 4 waves for the period 1981-2008 and that is used to obtain the five religiosity variables, the shares of main religions and the dummy variables for the predominant religion of each country.

"patent_1980_2018.csv": Dataset that contains data on patents taken from the World Intellectual Policy Organization (WIPO).

"match_wipo.dta": Stata dataset that comes from our elaboration and that contains the country names and "code2". It is used to merge the WIPO data on patents with the other datasets.

"WDIData.csv": Dataset from the World Development Indicators (WDI) containing the data on Population, Gross Domestic Product per capita, and Foreign direct investment.

"BL2013_MF1599_v2.0.dta": Dataset from Barro and Lee (2013) that is used to obtain the average years of tertiary education.

"Patent index1960 - 2015.xlsx": Dataset by Park containing the index on patent protection.

"match_park.dta": Stata dataset that comes from our elaboration and that contains the country names and "code2", and that is used to merge the Park's dataset containing the index on patent protection with the other datasets.

"Democracy Crossnational Data Spring 2009 StataSE.dta": Democracy Cross-national Data, Release 3.0 Spring 2009, (Norris, 2009), that is used to obtain the variable Religious freedom.

Description of the datasets in the subfolder "Source/USA"

The subfolder "USA" contains the source datasets used to generate the final dataset for the U.S. cross-state analysis.

"Religious Landscape Survey Data - Continental US.dta" and "Religious Landscape Survey Data - Alaska and Hawaii.dta": These two datasets are taken from the Pew Research Center's Religion & Public Life Project, Religion Landscape Survey, 2007, and are used to obtain the three religiosity variables for each U.S. State. Specifically, the datasets available from the PEW website are in SPSS format and these two datasets have been transformed in Stata format by the authors.

"patents_US.csv": Dataset taken from the U.S. Patent and Trademark Office containing the data on the total number of patents submitted by residents of the State in 2007.

"Gross domestic product (GDP) 2006 2007.xls": Dataset containing the Gross domestic product (GDP) by State (SAGDP2N) in 2006 that is used to compute the Foreign direct investment as a share of GSP in 2006.

"st-est00int-01.xls": Dataset taken from the United States Census Bureau that is used to obtain the Population of each State.

"p20-560.pdf": File on the Educational Attainment for the Population Aged 25 and Over by Region, State, and Nativity Status: 2007. Source: United States Census Bureau, Educational Attainment in the United States: 2007; Issued January 2009, P20-560. We take the data on the share (in percentage term) of the State population with bachelor's degree or more from Table 2, at page 8, as explained below.

"ba_US.xls": Dataset containing the share (in percentage term) of the State population with bachelor's degree or more. The dataset comes from our elaboration and the data are taken from Table 2, at page 8, of the report cited above (the Educational Attainment for the Population Aged 25 and Over by Region, State, and Nativity Status: 2007) and contained in the file "p20-560.pdf".

"Tab III.D11.xls": Dataset taken from the Bureau of Economic Analysis, State or DC, from which we take the Foreign Direct Investment in the United States (FDIUS) in 2006 (Table III.D11. Gross Property, Plant, and Equipment of Affiliates, State by Use, in millions of dollars).

"Per capita real GDP by state 2006 2007.xls": Dataset taken from the Bureau of Economic Analysis, State or DC, from which we take the (SAGDP10N) Per capita real GDP by State (Chained 2012 dollars) in 2007.

Description of the datasets in the subfolder "Data/Final"

The subfolder "Final" will contain the following two final datasets (generated by combining the source files) used for the analysis.

"icc_final_dataset.dta": Final dataset used for the international cross-country analysis.

"usa_final_dataset.dta": Final dataset used for the U.S. cross-state analysis.

Description of the datasets in the subfolder "Data/Temp"

The subfolder "Temp" is empty and will include all the datasets generated by and recalled by the do files to construct the two final datasets used to replicate the analysis. The datasets contained in this folder are of no use and can be deleted after the final datasets have been generated.

Description of programs/code files: folder "Programs"

We here report a brief description of the do files contained in the folder "Programs". The do files also contain some comments inside that help explaining the various steps.

The file "master.do" is the master file. It calls the other do files (whose brief description of is reported below) to generate the datasets and replicate the results of the analysis.

STEP 1: BUILDING THE DATASET FOR THE INTERNATIONAL CROSS-COUNTRY ANALYSIS

Do files numbered from 1 to 14 are used to generate the final dataset "icc_final_dataset.dta" for the international cross-country analysis.

1_wvs.do: the code generates the five variables of religiosity (Religious person, Belief in God, Church attendance, Importance of religion, God very important), the four variables for the shares of religions (Protestant, Catholic, Muslim, Orthodox) and the correspondent predominant religions for all countries using the datafile named "WVS_Longitudinal_1981_2016_stata_v20180912_extract.dta" that has been extracted from the World Values Survey (WVS) dataset "WVS_Longitudinal_1981_2016_stata_v20180912.dta". At the beginning of the file there are reported the instructions on how the extracted datafile has been (and can be) obtained. The data are stored in the temporary dataset "wvs_dbc.dta".

2_evs.do: the code generates the five variables of religiosity (Religious person, Belief in God, Church attendance, Importance of religion, God very important), the four variables for the shares of religions (Protestant, Catholic, Muslim, Orthodox) and the correspondent predominant religions for all countries using the datafile named "ZA4804_v3-1-0_extract.dta" that has been extracted from the European Values Study (EVS) dataset "ZA4804_v3-1-0.dta". At the beginning of the file there are reported the instructions on how the extracted datafile has been (and can be) obtained. The data are stored in the temporary dataset "evs_dbc.dta".

3_wvs_evs.do: the code merges the dataset "wvs_dbc.dta" obtained from the WVS dataset at point 1 (i.e. from "1_wvs.do") with the dataset "evs_dbc.dta" obtained from the EVS dataset at point 2 (i.e. from "2_evs.do"). The merged dataset is named "wvs_evs.dta".

4_wipo.do: the code prepares the WIPO dataset on patents "patent_1980_2018.csv" for the merger with our dataset. The resulting data is stored in the file "wipo.dta".

5_merge_wipo.do: the code updates the dataset "wvs_evs.dta" (obtained at step 3) with the patents data contained in "wipo.dta" (obtained at step 4). The resulting dataset is named "w_evs_wipo.dta".

6_wdi.do: the code prepares the dataset "WDIData.csv" from the WDI containing the data on Population, Gross Domestic Product per capita, and Foreign direct investment, for merging with our dataset. The code stores the data in the file "WB_dataset.dta".

7_merge_wdi.do: the code updates the dataset "w_evs_wipo.dta" (obtained at step 5) with the WDI data in the file "WB_dataset.dta" (obtained at step 6) and stores the data in the file "w_evs_wipo_wdi.dta".

8_barrolee.do: the code prepares the dataset "BL2013_MF1599_v2.0.dta" from Barro and Lee that contains the data on tertiary education (years) for merging with our dataset. The code stores the data in the file "bl.dta".

9_merge_barrolee.do: the code updates the dataset "w_evs_wipo_wdi.dta" (obtained at step 7) with the Barro and Lee data on tertiary education contained in the file "bl.dta" (obtained at step 8) and stores the data in the file "w_evs_wipo_wdi_bl.dta".

10_park.do: the code prepares the dataset "Patent index1960 - 2015.xlsx" from Park containing the index of patent protection for merging with our dataset. The code stores the data in the file "Panel_Park.dta".

11_merge_park.do: the code updates the dataset "w_evs_wipo_wdi_bl.dta" (obtained at step 9) with the Park's index of patent protection contained in the file "Panel_Park.dta" (obtained at step 10) and stores the data in the file "w_evs_wipo_wdi_bl_ipr.dta".

12_norris.do: the code prepares the dataset "Democracy Crossnational Data Spring 2009 StataSE.dta" from the Democracy Cross-national Codebook (Norris, 2009) containing the data on Religious Freedom index for merge and saves the data in the file "norris.dta".

13_merge_norris.do: the code updates the dataset "w_evs_wipo_wdi_bl_ipr.dta" (obtained at step 11) with the Religious freedom index in the file "norris.dta" (obtained at step 12) and stores the data in the file "w_evs_wipo_wdi_bl_ipr_norris.dta".

14_communism_labels.do: the code updates the dataset "w_evs_wipo_wdi_bl_ipr_norris.dta" (obtained at step 12) with the dummy variable "communist" denoting whether the country is or was a communist state, add labels and macros, and stores the data in the file "icc_final_dataset.dta" that is our final dataset for the international cross-country analysis.

STEP 2: TABLES AND FIGURES FOR THE INTERNATIONAL CROSS-COUNTRY ANALYSIS

As reported in the master file, the following do files contain the code to generate the tables and figures (whose names are reported beside the name of the do file) for the international cross-country analysis starting from our final dataset "icc_final_dataset.dta".

15_table_2.do   Table 2

16_table_F1.do   Table F.1

17_table_F2.do   Table F.2

18_table_F3.do   Table F.3

19_table_F4.do   Table F.4

20_figures_1a_1b.do   Figures 1a and 1b

21_figures_F1a_F1b.do   Figures F.1a and F.1b

22_figures_F2a_F2b.do   Figures F.2a and F.2b

23_figures_F3a_F3b.do   Figures F.3a and F.3b

24_figures_F4a_F4b.do   Figures F.4a and F.4b

25_figures_F5a_F5b.do   Figures F.5a and F.5b

STEP 3: BUILDING THE DATASET FOR THE U.S. CROSS-STATE ANALYSIS

Do files numbered from 26 to 33 are used to generate the final dataset "usa_final_dataset.dta" for the U.S. cross-state analysis.

26_pew.do: the code merges the datasets "Religious Landscape Survey Data – Continental US.dta" and "Religious Landscape Survey Data – Alaska and Hawaii.dta" from the Pew Research and generates the three variables of religiosity (Importance

of religion, Belief in God, Church attendance) for the U.S. states. The data are then stored in the dataset "pew_usa.dta".

27_patents.do: the code prepares the dataset "patents_US.csv" containing the data on the total number of patents submitted by residents of the State in 2007 (from the U.S. PATENT AND TRADEMARK OFFICE) for merging with other datasets and stores the data in the file "patents_US.dta".

28_gsp.do: the code prepares the dataset "Gross domestic product (GDP) 2006 2007.xls" containing the data on the Gross domestic product (GDP) by State in 2006 for merging with other datasets and stores the data in the file "gsp_US.dta". This data will then be used to compute the Foreign direct investment as a share of GSP (in 2006).

29_pop.do: the code prepares the dataset "st-est00int-01.xls" containing the data on Population by State in 2007 for merging with other datasets and stores the data in the file "pop_US.dta".

30_edu.do: the code prepares the dataset "ba_US.xls" containing the data on Tertiary education in 2007 for merging with other datasets and stores the data in the file "ba_US.dta".

31_fdi.do: the code prepares the dataset "Tab III.D11.xls" containing the data of the Foreign Direct Investment by State in 2006 for merging with other datasets and stores the data in the file "fdi_US.dta".

32_gsp_c.do: the code prepares the dataset "Per capita real GDP by state 2006 2007.xls" containing the data on the Per capita real GDP by State in 2007 for merging with other datasets and stores the data in the file "gsp_c_US.dta".

33_merge.do: the code merges all the datasets generated before (in the steps 26-32), generates the variables required, adds the labels to the variables used, and stores the data in the file "usa_final_dataset.dta", that is our final dataset used for the U.S. cross-state analysis.

STEP 4: TABLES AND FIGURES FOR THE U.S. CROSS-STATE ANALYSIS

As reported in the master file, the following do files contain the code to generate the tables and figures (whose name is reported beside) for the U.S. cross-state analysis.

34_table_3.do   Table 3

35_figures_2a_2b.do   Figures 2a and 2b

36_figures_F6a_F6b.do Figures F.6a and F.6b

37_figures_F7a_F7b.do Figures F.7a and F.7b

Description of output files: folder "Results"

The folder "Results" will contain the tables and figures with all results of the analysis after the execution of the file "master.do". The name of the files in the folder after running the program are self-explaining about their content and are the following:
Figure_1a.pdf
Figure_1b.pdf

```
Figure_2a.pdf
Figure_2b.pdf
Figure_F.1a.pdf
Figure_F.1b.pdf
Figure_F.2a.pdf
Figure_F.2b.pdf
Figure_F.3a.pdf
Figure_F.3b.pdf
Figure_F.4a.pdf
Figure_F.4b.pdf
Figure_F.5a.pdf
Figure_F.5b.pdf
Figure_F.6a.pdf
Figure_F.6b.pdf
Figure_F.7a.pdf
Figure_F.7b.pdf
Table_2.rtf
Table_3.rtf
Table_F.1.rtf
Table_F.2.rtf
Table_F.3.rtf
Table_F.4.rtf
```

References

Barro, Robert and Jong-Wha Lee (2013) "A New Data Set of Educational Attainment in the World, 1950-2010." Journal of Development Economics, 104, 184-198. http://www.barrolee.com/. (accessed December 28, 2015).

EVS (2020). European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008). GESIS Data Archive, Cologne. ZA4804 Data file Version 3.1.0, https://doi.org/10.4232/1.13486. (accessed July 16, 2021).

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014. World Values Survey: All Rounds - Country-Pooled Datafile Version: http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp. Madrid: JD Systems Institute. (accessed January 27, 2020).

Norris, Pippa (2009). "Democracy Cross-national Data. Release 3.0 Spring 2009." http://www.pippanorris.com/. (accessed March 8, 2013).

Park, Walter G. (2008) "International patent protection: 1960-2005." Research Policy, 37(4), 761-766. https://doi.org/10.1016/j.respol.2008.01.006. (accessed January 28, 2020).

Pew Forum on Religion and Public Life (2007). "U.S. Religion Landscape Survey". https://www.pewforum.org/dataset/u-s-religious-landscape-survey/. (accessed November 3, 2013).

U.S. Bureau of Economic Analysis, "Foreign Direct Investment in the United States (FDIUS): 2006 Data Tables". https://www.bea.gov/international/foreign-direct-investment-united-states-fdius-2006-data-tables. (accessed February 13, 2020).

U.S. Bureau of Economic Analysis, "Gross Domestic Product by State". Last updated: November 7, 2019. https://www.bea.gov/data/gdp/gdp-state. (accessed February 13, 2020).

U.S. Census Bureau (2009), "Educational Attainment in the United States: 2007". Issued January 2009, P20-560. https://www.census.gov/library/publications/2009/demo/p20-560.html. (accessed February 12, 2020).

U.S. Census Bureau, "Table 1. Intercensal Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2000 to July 1, 2010 (ST-EST00INT-01)". Last Revised: November 30, 2016. https://www.census.gov/data/tables/time-series/demo/popest/intercensal-2000-2010-state.html. (accessed February 13, 2020).

U.S. Patent and Trademark Office, Electronic Information Products Division, Patent Technology Monitoring Team (PTMT) (2007). "Calendar Year 2007 Patent Counts by Patent Type and by State and Country of Origin (01-Jan-2007 to 31-Dec-2007)". http://www.uspto.gov/web/offices/ac/ido/oeip/taf/st_co_07.htm. (accessed November 4, 2013).

Wikipedia (2020) "Communist state", April 1, 2020. https://en.wikipedia.org/wiki/Communist_state. (accessed April 7, 2020).

World Development Indicators. Washington, D.C., The World Bank. http://data.worldbank.org/data-catalog/world-development-indicators. (accessed January 27, 2020).

World Intellectual Policy Organization (WIPO) statistics database. "Intellectual property right: Patent". https://www3.wipo.int/ipstats/index.htm?tab=patent. Last updated: October 2019. (accessed January 27, 2020).

Acknowledgements