

Pillars of Prosperity

The Political Economics of Development Clusters

Codebook

Accompanying Dataset¹
(Pillars 2.dta)

Tim Besley

STICERD and Department of Economics, The London School of Economics

Torsten Persson

Institute for International Economic Studies, Stockholm University

¹Please email Chris Dann at c.m.dann@lse.ac.uk for any information, queries or unnoticed errors concerning the dataset and accompanying slides on the website.

Contents

1	Introduction	3
2	Measurements of State Capacity	4
2.1	Fiscal Capacity	4
2.1.1	Measures of the Tax System	4
2.1.2	Introduction of VAT and Income Taxes and “Information Capacity”	5
2.1.3	Measures of the Informal Economy	5
2.2	Legal Capacity	6
2.2.1	Government Antidiversion Policies	6
2.2.2	Doing Business Indexes	7
2.3	Collective Capacity	8
2.3.1	Educational Attainment and Human Capital	8
2.3.2	Indicators of Health	8
2.3.3	Investments in Collective Capacity	9
3	Measures of Violence	9
3.1	Civil War	9
3.2	Repression	10
3.3	Other Measures	10
4	Measures of Model Parameters	11
4.1	Common Interests	11
4.2	Cohesive Institutions	11
4.2.1	Executive Constraints	11
4.2.2	Alternative Measures of Cohesiveness	13
4.3	Political Stability	13
4.4	Legal Origins	13
5	Other Measures	14
5.1	Fragility Indexes	14
5.1.1	The Fund for Peace	14
5.1.2	Polity IV Fragility Index	14
5.2	GDP	15
5.3	Natural Disasters	16
5.4	Inequality	16
5.5	Private Credit	17
5.6	Corruption	17
5.7	Investment	17
5.8	Membership of Security Council	18
5.9	Aid	18
5.10	Attitudes and Values	19
A	World Bank Doing Business Data Collection City Observations	25
B	Discrepancies in Low Inequality Dummy	26

1 Introduction

This document describes the variables in the *Pillars 2.dta* dataset accompanying Tim Besley and Torsten Persson, 2011, *Pillars of Prosperity: The Political Economics of Development Clusters*, Princeton University Press. This dataset serves as an update to most of the original variables in *Pillars.dta*, using, where applicable:

- i) more recent time periods (up to 2016)
- ii) newer and better data sources
- iii) multiple data sources for certain variables to permit robustness checks
- iv) measures of ‘collective capacity’
- v) a panel data structure

For all variables, we have outlined whether they stem from either i) brand new datasources, ii) the original datasources used in *Pillars.dta* but are updated to a more recent year, or iii) are kept, and hence remain unchanged, from *Pillars.dta*. This dataset can be used in conjunction with the new .do files provided on the book’s website to replicate the graphs and tables in the book with this updated data. Thus, the respective .do files associated with *Pillars.dta* and *Pillars 2.dta* allow for comparisons of graphs and tables across a decade. However, whilst elements of this dataset employ a panel structure versus *Pillars.dta*, almost all regressions are still cross-sectional by design. Note that in order to generate the maps (Figures 1.1, 1.2, 8.1, 8.2) in the book you need to download *WB_countries_shp_clean.dta*. All variables from the original *Pillars.dta* dataset are kept in the dataset labelled with an “_old” suffix. For all updates and downloading lecture materials please refer to the book website at www.PillarsofProsperity.info.

The following variables are used to identify observations:

country variable contains the name of the country.

ncode is a numeric code that identifies each country.

ccode is a three-letter code for each country.

id is a numeric code that is used in drawing the maps. This is used in conjunction with *world-c.dta* that contains shape information for countries of the world.

year shows the year of observation. Note that for cross-sectional variables, we simply fill out that variable for all years and this variable becomes irrelevant.

c_year shows the year each country became independent.

The next section explains the variables used for measuring state capacity. The third section describes the measures we use for political violence. The fourth section contains the variables used to proxy for model parameters. The last section contains all other variables present in the dataset.

2 Measurements of State Capacity

2.1 Fiscal Capacity

We use two sets of measures for fiscal capacity: first a set of variables capturing the tax system. Second, we use a measure of the informal economy.

2.1.1 Measures of the Tax System

Brand new measures of the tax system are taken from the [International Centre for Taxation and Development \(ICTD\) Government Revenue Dataset](#), which were not used in the original *Pillars.dta*. This is an unbalanced panel variable between 1980-2016. There are 140 observations (countries) with non missing values for 2016 as a cross-section (the most recent year of available data). For *incometax_share* in 2016 there are 87 observations, and for *tradetax_share* in 2016 there are 112 observations². In the dataset, we also include a set of “mipolated” versions of the following tax variables from the ICTD, with prefix *mipo_* for each respective variable. This simply carries forward the last available non-missing entry of each variable to maximise country coverage for the Pillars of Prosperity index cluster analysis in Chapter 8.

taxrevenuegdp Share of Taxes in GDP

The ratio of total tax revenue to GDP from 1980 to 2016. This variable corresponds to the ICTD’s *total non-resource tax revenue, excluding social contributions*. Social contributions are excluded as recommended by the ICTD, purely for data consistency purposes across countries.

incometax_share Share of Income Taxes in Total Taxes

The share of revenue from incomes taxes in total taxes from 1980 to 2016. The closest variable in the ICTD dataset corresponding to income tax was *total income, capital gains and profit taxes on individuals*, which is measured as a share of GDP. Hence, we divide this variable by *taxrevenuegdp*.

tradetax_share Share of Trade Taxes in Total Taxes

The share of revenue from trade taxes in total taxes 1980 to 2016. This variable corresponds to the ICTD’s *total taxes on international trade, including both import and export taxes*, which is also measured as a share of GDP. Hence, we divide this variable by *taxrevenuegdp*.

nontradetax_share Share of Non-Trade Taxes in Total Taxes

This variable is constructed from the above variable as: $1 - \text{tradetax_share}$.

taxshare_bias Difference between income and trade tax shares

This is constructed by subtracting share of trade taxes from share of income tax in total tax revenue ($= \text{incometax_share} - \text{tradetax_share}$).

²Macro panel data on taxes for 31 countries going back to the early 19th century is available from Andersson and Brambor (2020). However, given the smaller sample range relative to ICTD, we do not use this dataset.

2.1.2 Introduction of VAT and Income Taxes and “Information Capacity”

These variables are collected by the authors (and were kept from the original *Pillars.dta*). The information capacity dummies stem from Lund University’s STANCE [Information Capacity Dataset](#).

vatdummy_ext VAT dummy

This variable takes a value of zero until a country introduces VAT after which it takes the value of one.

incometax_conf Dummy=1 if country has income tax

This variables take a value of zero until a country introduces income taxation after which it takes the value of one.

popreg Population register dummy

This variable takes a value of zero until a country introduces a population register after which it takes the value of one.

statag_intro Statistical agency dummy

This variable takes a value of zero until a country introduces a statistical agency after which it takes the value of one. The dummy variable is maintained even if a statistical agency is eventually abolished by a country due to its operations being subsumed by other government departments.

2.1.3 Measures of the Informal Economy

Gathered from the [Quality of Government](#), we use the data generated by Elgin and Oztunali’s (2012) ‘Shadow Economies around the World: Model Based Estimates’. The data is brand new relative to *Pillars.dta*. This variable indicates the size of the “shadow economy” (i.e. informal economy) as a % of GDP, and is a panel variable from 1950 to 2009. The correlation coefficient between this variable and the old one from *Pillars.dta* is 0.8528 over years with corresponding non-missing observations.

minform Size of Informal Economy % GDP

This is the original variable from Elgin and Oztunali (2012).

formal Size of Formal Economy % GDP

Size of the formal economy is defined as $1 - \text{minform}$.

2.2 Legal Capacity

2.2.1 Government Antidiversion Policies

mgadp97 Property Rights Protection Index (ICRG)

We use the ICRG dataset ([International Country Risk Guide](#), © The PRS Group, Inc., East Syracuse, NY, 1980-present.) to construct a measure of government antidiversion policies. It is calculated as an average of indexes of “law and order”, “bureaucratic quality”, “corruption”, “risk of expropriation” and “government repudiation of contracts” from ICRG dataset in 1997. This measure is a cross-sectional variable and is defined for 129 countries in 1997. This variable is kept from *Pillars.dta* to gauge convergent validity between this variable and the new variables we use of i) V-Dem’s similar measure of property rights, *v2xcl_prpty*, ii) the Fraser Institute’s ‘Legal Structure and Security of Property Rights’ variable, *fi_legprop_pd*, and iii) the World Justice Project’s rule of law index (below)³.

v2xcl_prpty Property Rights (V-Dem)

This new variable, which did not feature in *Pillars.dta*, is constructed by [V-Dem](#) by asking project coders “Do citizens enjoy the right to private property?” V-Dem accordingly classify property rights as “. . . the right to acquire, possess, inherit, and sell private property, including land. Limits on property rights may come from the state which may legally limit rights or fail to enforce them; customary laws and practices; or religious or social norms. This [variable] concerns the right to private property, not actual ownership of property”. It is normalised between 0-1.

WJP_score Rule of Law Index (World Justice Project)

This new variable, which did not feature in *Pillars.dta*, is a separate index on legal capacity from the [World Justice Project \(WJP\)](#). It is an overall score across eight subindices: 1) Constraints on Government Power, 2) Absence of Corruption, 3) Open Government, 4) Fundamental Rights, 5) Order and Security, 6) Regulatory Enforcement, 7) Civil Justice and 8) Criminal Justice (see WJP for further details). This data exists for 2014-2019, years where an overall index was provided by the WJP. This variable ranges from 0-1. It is non-missing for 110 countries for 2016.

fi_legprop_pd Legal structure and security of property rights (Fraser Institute)

This new variable, which did not feature in *Pillars.dta*, is a separate index on legal capacity by the [Fraser Institute](#), as gathered from the [Quality of Government](#). This variable ranges from 0-10, with 0 meaning “no judicial independence”, “no trusted legal framework exists”, “no protection of intellectual property”, “military interference in rule of law”, and “no integrity of the legal system”, and 10 meaning “high judicial independence”, “trusted legal framework exists”, “protection of intellectual property”, “no military interference in rule of law”, and “integrity of the legal system”. It is normalised between 0-1 by dividing each observation by 10. This is a panel variable, covering the period 1970-2015 (although between 1970-2000 this variable was only measured in five year

³We also keep this variable because of its wide usage in the literature (e.g. see Acemoglu et al. (2001)).

intervals)⁴.

2.2.2 Doing Business Indexes

We use the following variables from the [World Bank's Doing Business Project](#) in 2016. There are 188 countries with non missing observations in this year across all variables. This data stems from the same source as the original *Pillars.dta*, and so is essentially an update of the following variables to include ten additional years⁵.

mcontracts16 Index of contract enforcement (rank)

This variable contains the country's normalised rank on the measure of enforcing contracts in 2016. This was constructed using the Doing Business variable *Score-Enforcing contracts (DB16 methodology)*, which is a continuous variable from 0-100, with 0 being the worst regulatory performance and 100 the best. Country rankings from these scores were subsequently constructed. To normalise the ranking, we minus one from each country's rank, and then divide by the total number of countries with non-missing data minus one (i.e. $188-1=187$).

mproperty16 Ease of registering property (rank)

This variable contains the country's normalised rank on ease of registering property in 2016. This was constructed using the Doing Business variable *Score-Registering property (DB16 methodology)*, which is a continuous variable from 0-100, with 0 being the worst regulatory performance and 100 the best. Country rankings from these scores were subsequently constructed. To normalise the ranking, we minus two from each country's rank, and then divide by the total number of countries with non-missing data minus two (i.e. $188-2=186$). We subtract by two versus one here given Libya, Marshall Islands and the Federated States of Micronesia were all equally ranked by the World Bank for this variable as 0.

mbusiness16 Ease of doing business (rank)

This variable contains the country's normalised rank on the overall Doing Business indicator in 2016. This was constructed using the Doing Business variable *Ease of doing business score (DB16 methodology)*, which is a continuous variable from 0-100, with 0 being the worst regulatory performance and 100 the best. Country rankings from these scores were subsequently constructed. To normalise the ranking, we minus one from each country's rank, and then divide by the total number of countries with non-missing data minus one (i.e. $188-1=187$).

mcredit16 Ease of getting credit (rank)

This variable contains the country's normalised rank in the ease of access to credit in 2016. This was constructed using the Doing Business variable *Score-Getting credit (DB15-19 methodology)*, which is a continuous variable from 0-100, with 0 being the worst regulatory performance and 100 the best. Country rankings from these scores were subsequently constructed. To normalise the

⁴Papers such as Nunn (2007) have similarly used this variable to measure "legal quality".

⁵See Appendix A for further details on data collection.

ranking, we minus 3.5 from each country's rank, and then divide by the total number of countries with non-missing data minus 3.5 (i.e. $188-3.5=184.5$). We subtract by 3.5 versus one here given Eritrea, Iraq, Jordan, Libya, Somalia and Yemen were all equally ranked by the World Bank for this variable as 0.

2.3 Collective Capacity

All the following variables that measure collective capacity are brand new, as collective capacity was not measured in the original *Pillars.dta*.

2.3.1 Educational Attainment and Human Capital

The following variable for educational attainment stems from Lee and Lee (2016), gathered from the [Barro-Lee Educational Attainment Dataset](#) website. Both variables count observations every five years, over the period 1870-2010.

leelee_educattainment Average years of total schooling for population aged 15-64

2.3.2 Indicators of Health

The following indicators of health are initially gathered from the [Quality of Government](#) database, but their respective original sources have been listed for each respective variable.

wdi_lifexp Life expectancy at birth, total (years)

This variable measures life expectancy at birth, total (years), and is sourced from the World Bank's World Development Indicators (WDI), with "total" referring to males and females.

wdi_mortinf Infant mortality rate (per 1,000 live births)

This variable measures infant mortality per 1,000 live birth, and is again sourced from the World Bank's World Development Indicators (WDI). None of the .do files for the updated charts/tables include this variable, but we have left this on for the discretion of the user to explore alongside other variables in the dataset.

who_dwtot Share of population using at least basic drinking water services

This variable measures the percentage of the total population (both males and females) with access to basic drinking water services, and is sourced from the World Health Organisation (WHO). None of the .do files for the updated charts/tables include this variable, but we have left this on for the discretion of the user to explore alongside other variables in the dataset.

coll_cap Collective capacity index

Based on Besley & Persson (2014), this index is the average of educational attainment and life expectancy. We first use min-max normalisation on *leelee_educattainment* and *wdi_lifexp*, and then

average these two variables. This index is defined for 2010 as a cross-section (this is also the last year of data available for *leelee_educattainment*) for 107 countries with non-missing data.

2.3.3 Investments in Collective Capacity

The following two variables stem from the WDI and measure government expenditure/investments in health and education – the two main forms of collective capacity, as outlined by the previous two subsections. None of the .do files for the updated charts/tables include this variable, but we have left this on for the discretion of the user to explore alongside other variables in the dataset.

wdi_dgovhexp Domestic general government health expenditure (% GDP)

wdi_expedu Government expenditure on education, total (% GDP)

3 Measures of Violence

3.1 Civil War

The following variables are all constructed from the [UCDP/PRIO Armed Conflict Dataset version 19.1](#), produced by the Uppsala Conflict Data Program (UCDP) at the Department of Peace and Conflict Research, Uppsala University. We recreate the ‘war incidence’ dummy variable (see Besley & Persson, 2009) by looking at the entire episode of “internal conflicts” (type 3 or type 4 from the UCDP), meaning at some point the conflict reached 1,000+ battle-related deaths in a given year (i.e. ‘intensity level’ of 2 from the UCDP) during its full duration. Stemming from the same source as the data used to measure civil war in *Pillars.dta*, this is essentially an update of the original data to a more recent year. The correlation coefficient between this new dummy and the old dummy over years with non-missing observations across both variables (1946-2006) is 0.9840.

wardum Civil war dummy

This variable takes the value of one if there was at least one internal armed conflict without intervention from other states (i.e. non-internationalised) in a given country-year.

avg_civilwar10 Proportion of years in civil war 1950-2010

This variable shows the proportion of years with civil war (where the *wardum* is equal to one) over 1950-2010 for each country (excluding missing values).

avg_civilwar75 Proportion of years in civil war 1975-2016

This is constructed in the same way as the above variable but for the 1975-2016 period.

countryavg_civilwar Proportion of years in civil war 1950-2016

This is the same variable as *avg_civilwar10* except for the fact that it uses a longer time span.

yearavg_civilwar Prevalence of civil war

This variable summarises the fraction of countries in our sample that has civil war in a given year. This is a time series variable and is defined from 1950 until 2016.

3.2 Repression

The following variables are all constructed from the number of purges variable (*purges*) taken from the [Banks and Wilson \(2019\) Cross-National Time-Series Data Archive](#). This data stems from the same source as in the original *Pillars.dta*, and so the following variables are just an update to include ten additional years.

repressiondum Repression dummy

This variable takes the value of one if there was at least one purge in a given country-year.

avg_repression10 Proportion of years in repression between 1950-2010

This variable is constructed by dividing the sum of years with positive number of purges by total years with non missing data on purges in the 1950 - 2010 period. There are 198 countries with non missing observations for this variable.

avg_repression75 Average proportion of years in repression 1975-2016

This is constructed in the same way as *avg_repression10* but for the 1975-2016 period. There are 198 countries with non missing observations for this variable.

countryavg_repression Proportion Years in Repression 1950-2016

This is the same variable as *avg_repression10* except for the fact that it uses a longer time span. Again there are 198 observations.

yearavg_repression Prevalence of repression

This variable summarises the fraction of countries with positive purges in any given year using the *purges* variable (above). This is a time series variable and is defined from 1950 until 2016.

3.3 Other Measures

repdum_purPRIO Ordered Variable for Violence

This variable takes 0 if there is peace. It takes the value of 1 if there is repression (as defined by positive number of purges from the *purges* variable above) in the absence of civil war (defined from the *wardum* variable above). Finally it takes the value of 2 when there is civil war. Note that the ordering matters for the analysis in chapter 4 and the specific numbers given to each state is irrelevant. Notwithstanding using updated CNTS data and a new, self-constructed civil war

dummy, the correlation coefficient between this variable and the one in *Pillars.dta* is 0.9728.

violence Repression or Civil War

This variable takes the value of 1 if the ordered variable (*repdum_purPRIO*) takes 1 or 2. Notwithstanding using updated CNTS data and a new, self-constructed civil war dummy, the correlation coefficient between this variable and the one in *Pillars.dta* is 0.9543.

4 Measures of Model Parameters

4.1 Common Interests

avg_war00 Proportion of years in external conflict up to 2000

This variable captures the parameter ϕ in the model. It measures the proportion of years in external war from 1816 to 2000. This variable is kept from the original *Pillars.dta* dataset, given the Correlates of War (COW) (original data source) has not been updated (see original Pillars codebook for complete definition of variable). Although Reiter et al. (2014, 2016) have updated COW, this is only for interstate wars and not extrastate wars, hence we have kept the original variable as it is comprised of both types of wars.

fe_etfra Ethnic Fractionalisation (Fearon, 2003)

This is ethnic fractionalisation from Fearon (2003) (taken from [Quality of Government](#) database). This variable is defined for 157 countries, and is kept from *Pillars.dta*.

fe_ethom Ethnic Homogeneity (Fearon, 2003)

This is defined as $1 - fe_etfra$. This variable is defined for 157 countries, and is also kept from *Pillars.dta*. This variable also captures parameter ϕ (common interests) in the model.

4.2 Cohesive Institutions

4.2.1 Executive Constraints

We construct several variables from the variable *xconst* in the Polity IV dataset. We also use similar indicators from V-Dem, *v2x_jucon* and *v2xlg_legcon*, to permit robustness checks. Whilst the Polity IV data was used in the original *Pillars.dta*, and is thus updated to include ten additional years, the V-Dem data is brand new.

v2x_exconst Executive constraints (V-Dem)

This variable measures executive constraints in a country by taking the average of V-Dem's two main variables concerning executive constraints: *v2x_jucon* (executive constraints by the judiciary)

and *v2xlg_legcon* (executive constraints by the legislature and government agencies). If one variable is missing but the other is not, then *v2x_exconst* simply takes on the value of the non-missing type of executive constraints. Both of the aforementioned variables are already normalised by V-Dem so that each country's scores lie between 0 and 1 (0 being no executive constraints and 1 being the highest level of executive constraints). This is an unbalanced panel variable from 1789 (or later if uncoded for historical years) to 2016, and is analogous to Polity IV's *xconst* variable (discussed below).

mxconst10 Average executive constraints up to 2010 (Polity)

This measures the average value of the variable *xconst* from 1800 (or independence date if later) up to 2010. The average is taken over non missing values of *xconst* (values outside [1,7] are treated as missing). This variable is normalised so that each country's scores lie between 0 and 1 (subtract 1 and divide by 6 as the possible range for the average score is from 1 to 7). There are 168 non missing observations for this variable. This variable captures the parameter θ in the model.

mv2x_exconst10 Average executive constraints up to 2010 (V-Dem)

This measures the average value of the variable *v2x_exconst* from 1789 (or independence if later) up to 2010. The average is taken over non missing values of *v2x_exconst*. As mentioned for *v2x_exconst*, this variable is normalised to exist between 0-1 by design. There are 178 non missing observations for this variable. This variable captures the parameter θ in the model.

goodex_polity High executive constraints (Polity)

This variable is equal to one if the measure of executive constraints (*xconst*) is getting its maximum value (7) for each country and year. This is an unbalanced panel variable over the period 1800-2016 for 157 countries.

goodex_vdem High executive constraints (V-Dem)

This variable is equal to one if the measure of executive constraints (*v2x_exconst*) is getting a sufficiently high value (≥ 0.8) for each country and year (0.8 is just over one standard deviation from the mean, pooled over all available data). This is an unbalanced panel variable over the period 1789-2016 for 173 countries.

mgoodex_polity High Executive Constraints Polity (189 countries)

This variable shows the fraction of countries with the highest executive constraint score (*xconst*=7) for all the countries that Polity IV provides data on *xconst* continuously from 1900-2010 (there are 189 countries with *at most* one year of missing data; only 9 countries covered the full 1900-2010 series). Note that that is a time series variable that spans 1900-2010.

mgoodex_vdem High Executive Constraints V-Dem (181 countries)

This variable shows the fraction of countries with a sufficiently high executive constraint score ($v2x_{exconst} > 0.8$) for all countries that V-Dem provides data on continuously from 1900-2010 (there are 181 countries; only 15 countries covered the full 1900-2010 series). Note that that is a time series variable that spans 1900-2010.

4.2.2 Alternative Measures of Cohesiveness

parl_dem Parliamentary Democracy

This variable measures whether a country has parliamentary democracy. The period covered is 1950-2016. It takes the value of one for country-year observations where the *polity2* variable from the Polity IV dataset is positive (democracy) and the country is a parliamentary democracy, as determined by Bjørnskov and Rode’s (2020) database (which updates Cheibub et al.’s (2010) prominent dataset). The variable takes zero otherwise (given both *polity2* and Bjørnskov and Rode’s (2020) variables are non-missing). Despite being brand new data that was not used in *Pillars.dta*, the correlation coefficient between this variable and the old one is 0.7180.

4.3 Political Stability

mgamma10_polity Average non-open executive recruitment up to 2010 (Polity)

This measures average values of the sum of *xropen* and *xrcomp* variables in the Polity IV dataset from 1800 to 2010. Note that the average is taken when both *xropen* and *xrcomp* are not missing (we treat *xropen* and *xrcomp* as missing if they are less than one). The sum of *xropen* and *xrcomp* takes values between 2 and 7 in any given year, so in order to normalise the average we subtract 2 and divide by 5. To get a measure of political stability this average is inverted (multiplied by minus one and add with one). This variable is defined for 168 countries and corresponds to the parameter $1 - \gamma$ in the model.

mgamma10_vdem Average non-open executive recruitment up to 2010 (V-Dem)

This variable is analogous to the above variable (for robustness check purposes), except it takes the average of V-Dem’s variable *v2xel_frefair* from 1789 to 2010. It is already normalised to exist between 0-1. Again, to get a measure of political stability this average is inverted (multiplied by minus one and add with one). This variable is defined for 178 countries and corresponds to the parameter $1 - \gamma$ in the model.

4.4 Legal Origins

The legal origin dummy variables are from La Porta et al. (1998). These variables are kept from *Pillars.dta*.

l_eng English Legal Origin

This variable is equal to one if the country has a British legal origins and is equal to zero otherwise.

l_soc Socialist Legal Origin

This variable is equal to one if the country has a Socialist legal origins and is equal to zero otherwise.

l_sca Scandinavian Legal Origin

This variable is equal to one if the country has a Scandinavian legal origins and is equal to zero otherwise.

l_ger German Legal Origin

This variable is equal to one if the country has a German legal origins and is equal to zero otherwise.

l_fre French Legal Origin

This variable is equal to one if the country has a French legal origins and is equal to zero otherwise.

5 Other Measures

5.1 Fragility Indexes

5.1.1 The Fund for Peace

ffp_fsi Fragile States Index (2016)

This is the summated score across twelve different social, economic, and political indicators that form the overall Fragile States Index from the [Fund for Peace](#), each scaled between 0-10 (originally gathered from [Quality of Government](#)). The variable is normalised between 0-1 by dividing the Fragile States Index by 120 (the max. attainable score). Finally, for comparability with the Brookings Index for State Weakness (2008) from *Pillars.dta*, we subtract this variable by 1 so that higher values mean lower fragility. This is new data that did not feature in *Pillars.dta* (the Brookings Index was never updated to include ten additional years, and so a new variable had to be constructed from a different data source). The correlation coefficient between this variable and the Brookings Index for 2008 is 0.8125.

5.1.2 Polity IV Fragility Index

The variable *fragilityPolityIV* is from [Marshall and Elzinga-Marshall \(2017\)](#) and corresponds to their State Fragility Index (SFI). It is a panel variable from 1995-2016. This variable contains the quantile of the SFI for each country, and is represented as missing if also missing from the SFI variable. The other variables listed below are dummy variables for each quantile of fragility. This data did appear in *Pillars.dta*, and so is simply updated to include ten additional years.

fragilityPolityIV Fragility Index quantile

Given the original State Fragility Index is represented on a scale between 0-25, the following quantiles are based on the following coding structure:

- *extreme* = 4 if State Fragility Index $\in [21, 25]$
- *high* = 3 if State Fragility Index $\in [16, 20]$
- *moderate* = 2 if State Fragility Index $\in [11, 15]$
- *low* = 1 if State Fragility Index $\in [6, 10]$
- *nofragility* = 0 if State Fragility Index $\in [0, 5]$

extremefragility16 Dummy=1 if country defined as extremely fragile in 2016

highfragility16 Dummy=1 if country defined as highly fragile in 2016

moderatefragility16 Dummy=1 if country defined as moderately fragile in 2016

lowfragility16 Dummy=1 if country has low fragility in 2016

nofragility16 Dummy=1 if country has no fragility in 2016

5.2 GDP

We have several variables that are constructed from [Penn World Tables'](#) *rgdpna* variable (Real GDP at constant 2011 national prices (in mil. 2011US\$)), of which we calculate the per capita amount by dividing *rgdpna* by the population variable, *pop*. As a robustness check, we also develop the exact same indicators from V-Dem's *e_migdppc*, which corresponds to (and is sourced from) the [Maddison Project Database's](#) *rgdpnapc* (Real GDP per capita in 2011US\$, 2011 benchmark (suitable for cross-country growth comparisons)). This acts as a robustness check to the Penn World Tables data, although this variable tends to go much further back in history. Whilst the Penn World Tables data was used in *Pillars.dta*, and is hence updated to include ten additional years, the Maddison data is brand new. We primarily use the PWTs data in the new .do files for the updated charts/tables.

loggdp_pwt $\text{Ln}(\text{rgdpnapc})$

This is the natural logarithm of *rgdpnapc*. It is an unbalanced panel variable over 1950-2016, with 175 non missing observations for the year 2016.

loggdp_maddison $\text{Ln}(e_migdppc)$

This is the natural logarithm of *e_migdppc*. It is an unbalanced panel variable over 1789-2016, with 159 non missing observations for the year 2016.

mloggdp10_pwt Log Income per capita in 2010 (Penn World Table's)

This takes the value of *loggdp_pwt* in the year 2010. It is non missing for 175 countries.

mloggdp10_maddison Log Income per capita in 2010 (Maddison)

This takes the value of *loggdgdp_maddison* in the year 2010. It is non missing for 162 countries.

mloggdgdp00_pwt Log Income per capita in 2000 (Penn World Table's)

This takes the value of *loggdgdp_pwt* in the year 2000. It is non missing for 175 countries.

mloggdgdp00_maddison Log Income per capita in 2000 (Maddison)

This takes the value of *loggdgdp_maddison* in the year 2000. It is non missing for 160 countries.

mloggdgdp90_pwt Log Income per capita in 1990 (Penn World Table's)

This takes the value of *loggdgdp_pwt* in the year 1990. It is non missing for 175 countries.

mloggdgdp90_maddison Log Income per capita in 1990 (Maddison)

This takes the value of *loggdgdp_maddison* in the year 1990. It is non missing for 156 countries.

5.3 Natural Disasters

nat_dum =1 if a country experienced a natural disaster, as defined by EM-DAT.

This is a dummy variable generated from different disaster variables from the [EM-DAT](#) database. It is equal to one if a country had any of the following natural disasters in a given year: extreme temperature, earthquakes, floods and/or landslides. In EM-DAT, this corresponds to keeping events with disaster types 'Extreme temperature' and 'Flood' and disaster subtypes 'Landslide' and 'Tsunami'. EM-DAT appear to have changed their categorisation of natural disaster types since Besley and Persson (2009), and so the aforementioned four categories were chosen to be as comparable to the original variable in *Pillars.dta* (which was itself derived from EM-DAT). This is a panel variable from 1900 to 2016. For years with no natural disasters for a given country, we simply assume a value of 0. Notwithstanding this updated data, the correlation coefficient between this variable and the old variable in *Pillars.dta* is 9609.

5.4 Inequality

lowineq Low Inequality ($\text{Gini} < 0.3$)

This is a binary variable that is equal to 1 if a country has a low level of income inequality as measured by the average Gini coefficient being below 30. The Gini coefficient is gathered from the [Quality of Government](#) database, but is originally sourced from the World Bank's WDI (named *wdi_gini*). We take the average of the Gini coefficient for each country over all available years of data for that country. This average is used to construct the *lowineq* variable. There are 155 countries with at least one year of data for the Gini Index from the WDI. Note that the earliest date for which a country has Gini data is 1979 while the latest date is 2016. Note that also there are only 97 countries with five or more years of observations for Gini in this dataset. This variable is brand new as it includes new data that did not feature in the original *Pillars.dta*. The correlation

coefficient between this variable and the old variable is 0.5949 (see Appendix B for list of countries from both this dataset and *Pillars.dta* that are identified as having low inequality).

5.5 Private Credit

mpcy Private credit / GDP

This variable was taken from the World Bank Development Indicators ‘Domestic credit to private sector (% of GDP)’ variable. For comparability with *Pillars.dta*, we divided each observation by the yearly standard deviation. This is a panel variable from 1960 to 2016, with over 100 countries with consistent data from 1990 onwards. As a cross-section in 2016, there are 173 non missing observations.

5.6 Corruption

mcorr Corruption Perceptions Index (Transparency International)

This variable is the normalised value of the overall score of each country in Transparency International’s Corruption Perceptions Index (gathered from the [Quality of Government](#) database variable, *ti_cpi*). We divide *ti_cpi* by its respective standard deviation for each year of data to obtain *mcorr*, which is non missing for 172 countries for the year 2016, and is a panel variable from 2012 to 2016, with 164 balanced panels over this period. This data did feature in the original *Pillars.dta* but only for 2006 as a cross-section.

mv2x_corr Political Corruption Index

This variable stems from V-Dem whereby expert coders are asked ‘How pervasive is political corruption?’. This is the average of several corruption related subindices in V-Dem: public sector corruption index (*v2x_pubcorr*), executive corruption index (*v2x_execorr*), legislative corruption (*v2lgcrrpt*), and judicial corruption (*v2jucorrdc*) (see V-Dem for further details). The variable is inverted by multiply by negative 1 and then adding 1 so that higher scores mean less corruption, similar to the Corruption Perceptions Index. Second, albeit already normalised between 0-1, we divide observations by the annual standard deviation for comparability with our *mcorr* variable above. This variable covers an unbalanced panel of countries going as far back as 1789, with 171 non-missing observations for 2016 as a cross-section. The correlation coefficient of this variable with *mcorr* is 0.6557. None of the .do files for the updated charts/tables include this variable, but we have left this on for the discretion of the user to explore alongside other variables in the dataset.

5.7 Investment

Data was collected from the World Bank’s [TCdata360](#) regarding total investment as a percentage of GDP (although the original dataset stems from the IMF World Economic Outlook). This is brand new data that did not feature in *Pillars.dta*.

minvest Investment % GDP

This variable is formed from measuring the ratio of total investment in local currency and GDP in local currency. Investment or gross capital formation is measured by the total value of the gross fixed capital formation and changes in inventories and acquisitions less disposals of valuables for a unit or sector. This is a panel variable from 1980 to 2016, and is non missing for 171 countries for the year 2016. *minvest* is created by dividing by the yearly standard deviation for all non missing observations between 1980-2016. The correlation coefficient between this variable and the old one (for 2006 as a cross-section as per *Pillars.dta*) is 0.4805.

5.8 Membership of Security Council

New data for UN Security Council membership for 1951-2016 was collected from [Axel Dreher's page](#) at the Alfred-Weber-Institute for Economics, the Heidelberg University. However, data on security council membership between 1946-1950 was kept from the original *Pillars.dta* (which was collected by the authors). Over years that were non-missing between the old variable in *Pillars.dta* and Dreher's dataset, the correlation coefficient is 0.9911.

scmem Security Council Member

This variable is equal to 1 if a country is a member of the security council. Note that the five permanent members of the council are assigned a value of zero. This is a panel variable and is defined for 210 countries over the 1946-2016 period.

scmem_cold Security Council Member during Cold War

This variable is equal to *scmem* during the cold war (1946 - 1989). It takes the value of zero from 1990 to 2016.

5.9 Aid

Regarding total aid as a share of GNI, data was collected directly from the [World Bank's WDI regarding net official development assistance \(ODA\) as a share of GNI](#). For received total aid excluding debt, data was collected from the OECD Development Database on Aid regarding aid to countries from Developmental Assistance Countries (DAC). Both of these variables stem from the datasource used in *Pillars.dta*, and so the following variables are simply an update to include ten additional years of data.

aidgni Aid / GNI

This variable corresponds to the WDI's *Net ODA received (% of GNI)*. As defined by the World Bank, "Net official development assistance (ODA) consists of disbursements of loans made on concessional terms (net of repayments of principal) and grants by official agencies of the members of the Development Assistance Committee (DAC), by multilateral institutions, and by non-DAC countries to promote economic development and welfare in countries and territories in the DAC list of ODA recipients. It includes loans with a grant element of at least 25 percent (calculated at a rate of discount of 10 percent)".

receivedaidtotalexcludebt Received aid total excluding debt

This is the total amount of aid received by each country excluding debt repayments. We use the variable “Official Development Assistance, Excl Debt (Constant Prices, 2019 USD millions)” from the OECD Development Database on Aid from DAC members (subset 2a).

5.10 Attitudes and Values

All variables except the Cantril Ladder score stem from the [World Values Survey \(WVS\)](#). The Cantril Ladder score stems from the [World Happiness Report](#). All variables were carried forward from the year they were asked until the next year/wave/round of the WVS they were asked (e.g. if a variable was asked in 2009, and not asked again until 2016, the years 2010-2015 would also contain the 2009 cross-national average responses).

demo_better Democracy has problems but is better than any other form of government

Respondents were asked on a scale of 1 (strongly agree) to 4 (strongly disagree) whether they agreed with the above statement. We transformed the scale so that higher values mean people are in more agreement by multiplying by negative 1 and adding 5. This variable is available for at most 79 countries between 1996-2016.

democ_important Importance of democracy

Respondents were asked to rate on a scale of 1 (not at all important) to 10 (absolutely important) how important it is to live in a country that is democratically governed. This variable is available for at most 77 countries between 2007-2016.

demo_prtctlibrt Civil rights protect people’s liberty against oppression

Respondents were asked to place on a scale of 1 (not an essential characteristic of democracy) to 10 (essential characteristic of democracy) how vital it is for a democracy to have civil rights that protect people’s liberty against oppression. This variable is available for at most 76 countries between 2007-2016.

democ_womenrights Women have the same rights as men

Respondents were asked to place on a scale of 1 (not an essential characteristic of democracy) to 10 (essential characteristic of democracy) how vital it is for a democracy to have women’s rights that are equal to men’s. This variable is available for at most 76 countries between 2007-2016.

stf_demdvlp Satisfaction with way democracy is developing

Respondents were asked “On the whole are you very satisfied, rather satisfied, not very satisfied or not at all satisfied with the way democracy is developing in our country?”. We transformed the scale so that higher values mean people are more satisfied by multiplying by negative 1 and adding

5. This variable is available for at most 71 countries between 1996-2016.

conf_govt Confidence in the government

Respondents were asked on a scale of 1 (a great deal) to 4 (none at all) how confident they were with their country's government. We transformed the scale so that higher values mean people have more confidence by multiplying by negative 1 and adding 5. This variable is available for at most 102 countries between 1991-2016.

conf_justicesystem Confidence in the judicial system/courts

Respondents were asked on a scale of 1 (a great deal) to 4 (none at all) how confident they were with their country's judicial system/the courts. We transformed the scale so that higher values mean people have more confidence by multiplying by negative 1 and adding 5. This variable is available for at most 99 countries between 1983-2016.

jaccept_bribe Justifiable: accept a bribe

Respondents were asked on a scale of 1 (never justifiable) to 10 (always justifiable) as to whether it was justifiable to accept a bribe. This variable is available for at most 104 countries between 1983-2016.

jcheat_taxes Justifiable: cheat on taxes

Respondents were asked on a scale of 1 (never justifiable) to 10 (always justifiable) as to whether it was justifiable to cheat on their taxes. This variable is available for at most 101 countries between 1983-2016.

jsteal_prpty Justifiable: steal property

Respondents were asked on a scale of 1 (never justifiable) to 10 (always justifiable) as to whether it was justifiable to steal property. This variable is available for at most 58 countries for 2012.

jviolence_others Justifiable: violence towards others

Respondents were asked on a scale of 1 (never justifiable) to 10 (always justifiable) as to whether it was justifiable to enact violence upon others. This variable is available for at most 58 countries for 2012.

rspcttol_childqual Important child qualities: tolerance and respect for other people

Respondents were asked "Here is a list of qualities that children can be encouraged to learn at home. Which, if any, do you consider to be especially important? Please choose up to five." Other child qualities were independence, hard work, feeling of responsibility, obedience, amongst several others (see WVS for details). This variable is available for at most 104 countries between 1983-2016.

trust_ppl Trust people in general

Respondents were asked on a scale of 1 (most people can be trusted) to 2 (can't be too careful) "[g]enerally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" We transformed the scale so that higher values mean people are more trustworthy by multiplying by negative 1 and adding 3. This variable is available for at most 104 countries between 1983-2016.

cantril_ladder Cantril Ladder score

This variable measures average life satisfaction across countries. Respondents are asked the following: "Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?". This is an unbalanced panel variable from 2005-2016, with non-missing observations for 135 countries in 2016.

References

- [1] Acemoglu, Daron, Simon Johnson and James Robinson, 2001, “The Colonial Origins of Comparative Development: An Empirical Investigation”. *The American Economic Review*, 1369-1401.
- [2] Andersson, P., & Brambor, T. (2020, February 20). Financing the State: Government Tax Revenue from 1800 to 2012. Retrieved from Swedish National Data Service: <https://snd.gu.se/en/catalogue/study/snd1148>
- [3] Banks, Arthur S., & Wilson, Kenneth A. (2019). Cross-National Time-Series Data Archive. Databanks International. Jerusalem, Israel; see <https://www.cntsdata.com/>
- [4] Barro, R., & Lee, J. (2013). A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*, 184-198.
- [5] Besley, T., & Persson, T. (2014). The Causes and Consequences of Development Clusters: State Capacity, Peace, and Income. *Annual Review of Economics*, 927-949.
- [6] Bjørnskov, C., & Rode, M. (2020). Regime types and regime change: A new dataset on democracy, coups, and political institutions. *The Review of International Organizations*, 531–551.
- [7] Cheibub, J. A., Gandhi, J., & Vreeland, J. R. (2010). Democracy and dictatorship revisited. *Public Choice*, 67-101.
- [8] Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Anna Lührmann, Kyle L. Marquardt, Kelly McMann, Pamela Paxton, Daniel Pemstein, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Steven Wilson, Agnes Cornell, Lisa Gastaldi, Haakon Gjerløw, Nina Ilchenko, Joshua Krusell, Laura Maxwell, Valeriya Mechkova, Juraj Medzihorsky, Josefine Pernes, Johannes von Römer, Natalia Stepanova, Aksel Sundström, Eitan Tzelgov, Yi-ting Wang, Tore Wig, and Daniel Ziblatt. 2019. “V-Dem [Country-Year/Country-Date] Dataset v9”, Varieties of Democracy (V-Dem) Project. <https://doi.org/10.23696/vdemcy19>
- [9] Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer. 2007. “Private Credit in 129 Countries.” *Journal of Financial Economics* 84(2): 299–329.
- [10] Doing Business 2019: Training for Reform, 2019, World Bank, Washington DC: The World Bank Group. <https://www.doingbusiness.org/en/doingbusiness>
- [11] Dreher, Axel; Jan-Egbert Sturm and James Vreeland, 2009, Development Aid and International Politics: Does membership on the UN Security Council influence World Bank decisions? *Journal of Development Economics* 88: 1-18 2009.
- [12] Elgin, Ceyhan and Oztunali, Oguz, (2012), Shadow Economies around the World: Model Based Estimates, No 2012/05, Working Papers, Bogazici University, Department of Economics, <https://EconPapers.repec.org/RePEc:bou:wpaper:2012/05>.
- [13] EM-DAT: The OFDA/CRED International Disaster Database, Université Catholique de Louvain, Burssels, Belgium. URL: <https://www.emdat.be/>

- [14] Fearon, James, 2003, “Ethnic and Cultural Diversity by Country”, *Journal of Economic Growth*, vol. 8, 195-222.
- [15] Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), “The Next Generation of the Penn World Table” *American Economic Review*, 105(10), 3150-3182
- [16] Gwartney, J., Lawson, R., & Hall, J. (2016). *Economic Freedom Dataset 2016*, published in *Economic Freedom of the World: 2016*. Fraser Institute. Retrieved from <https://www.fraserinstitute.org/economic-freedom/dataset>
- [17] Haken, N., Messner, J., Hendry, K., Taft, P., Lawrence, K., Anderson, T., . . . Whitehead, A. (2018). *Failed states index 2006-2018*. The Fund for Peace (FFP). Retrieved from <https://fragilestatesindex.org/>
- [18] Harbom, Lotta, Erik Melander & Peter Wallensteen, (2008) *Dyadic Dimensions of Armed Conflict, 1946-2007*. *Journal of Peace Research* 45(5): 697-710.
- [19] ICTD/UNU-WIDER, ‘Government Revenue Dataset’, (version date) <https://www.wider.unu.edu/project/government-revenue-dataset>
- [20] International Country Risk Guide, © The PRS Group, Inc., East Syracuse, NY, 1980-present. URL: <http://www.prsgroup.com/>
- [21] La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny (1998), “Law and Finance,” *Journal of Political Economy*, vol. 106, 1113-1155.
- [22] Lee, Jong-Wha & Lee, Hanol, 2016. “Human capital in the long run,” *Journal of Development Economics*, Elsevier, vol. 122(C), pg 147-169.
- [23] Maddison Project Database, version 2018. Bolt, Jutta, Robert Inklaar, Herman de Jong and Jan Luiten van Zanden (2018), “Rebasing ‘Maddison’: new income comparisons and the shape of long-run economic development”, [Maddison Project Working paper 10](#)
- [24] Marshall, Monty G. and Gabrielle Elzinga-Marshall, 2017, *Global Report 2017: Conflict, Governance, and State Fragility*, Center for Systemic Peace and Center for Global Policy. URL: <http://www.systemicpeace.org/inscrdata.html>
- [25] Marshall, Monty G., Ted Robert Gurr and Keith Jagers. 2019. *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2018*. Version p4v2017 [Computer File]. College Park, MD: Center for Systemic Peace. URL: <http://www.systemicpeace.org/polityproject.html>
- [26] Nunn, N. (2007). Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade. *Quarterly Journal of Economics*, 569-600.
- [27] Pemstein, Daniel, Kyle L. Marquardt, Eitan Tzelgov, Yi-ting Wang, Juraj Medzihorsky, Joshua Krusell, Farhad Miri, and Johannes von Römer. 2019. “The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data”, V-Dem Working Paper No. 21. 4th edition. University of Gothenburg: Varieties of Democracy Institute.
- [28] Pettersson, Therese; Stina Höglbladh & Magnus Öberg, 2019. *Organized violence, 1989-2018 and peace agreements*, *Journal of Peace Research* 56(4).

- [29] Scartascini, C., Cruz, C., & Keefer, P. (2018, February). Database of political institutions codebook 2017 (dpi2017). Numbers for Development. Retrieved from <https://mydata.iadb.org/Reform-Modernization-of-the-State/Database-of-Political-Institutions-2017/938i-s2bw>
- [30] Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon and Richard Svensson. 2019. The Quality of Government Standard Dataset, version Jan19. University of Gothenburg: The Quality of Government Institute, <http://www.qog.pol.gu.se> [doi:10.18157/qogstdjan19](https://doi.org/10.18157/qogstdjan19)
- [31] Transparency International. (2019). Corruption perception index 2018. Retrieved from [http:// www.transparency.org/cpi](http://www.transparency.org/cpi) (Licensed under CC-BY-ND 4.0)
- [32] World Development Indicators, URL: <https://datacatalog.worldbank.org/>
- [33] World Health Organization. (2017). Global health observatory data repository. Retrieved from <http://www.who.int/gho/en/>
- [34] World Justice Project. (n.d.). Research and Data. Retrieved from World Justice Project: <https://worldjusticeproject.org/>
- [35] World Values Survey. (n.d.). Retrieved from World Values Survey: <https://www.worldvaluessurvey.org/wvs.jsp>

Appendix

Appendix A World Bank Doing Business Data Collection City Observations

Pre-2014 from the Doing Business Project, there were some city observations for certain countries versus a national average. The Doing Business Project carries this out to explore the heterogeneity in the business environment across countries (e.g. New York versus Los Angeles in the United States). Hence, we simply took the following major metropolises as being representative of the country itself:

- Dhaka – Bangladesh
- Sao Paolo – Brazil
- Shanghai – China
- Mumbai – India
- Jakarta – Indonesia
- Mexico City – Mexico
- Lagos – Nigeria
- Karachi – Pakistan
- New York – USA
- Tokyo – Japan

Appendix B Discrepancies in Low Inequality Dummy

Table B1: Discrepancies in Low Inequality Dummy

Old Dataset	New Dataset
Belarus	
Belgium	Albania
Bulgaria	Azerbaijan
Czech Republic	Belarus
Czechoslovakia	Belgium
Finland	Czech Republic
Hungary	Denmark
Latvia	Finland
Luxembourg	Hungary
Netherlands	Iceland
Poland	Iraq
Romania	Malta
Rwanda	Netherlands
Slovak Republic	Norway
Slovenia	Slovak Republic
Spain	Slovenia
Taiwan	Sweden
Ukraine	Ukraine
United Kingdom	

Note: red refers to those countries identified as having low inequality across both datasets.