

Pillars of Prosperity

The Political Economy of Development Clusters

Codebook

Accompanying Dataset
(Pillars.dta)

Tim Besley

Torsten Persson

STICERD and Department of Economics, London School of Economics

Institute for International Economic Studies, Stockholm University

February 12, 2013

Contents

1	Introduction	4
2	Measures of State Capacity	4
2.1	Measures of Fiscal Capacity	4
2.1.1	Measures of the Tax System	5
2.1.2	Introduction of VAT and Income Taxes	5
2.1.3	Measures of the Informal Economy	6
2.2	Legal Capacity	6
2.2.1	Government Antidiversion Policies	6
2.2.2	Doing Business Indexes	7
3	Measures of Violence	7
3.1	Civil War	7
3.2	Repression	8
3.3	Other measures	9
4	Measures of Model Parameters	9
4.1	Common Interests	9
4.2	Cohesive Institutions	10
4.2.1	Executive Constraints	10
4.2.2	Alternative Measures of cohesiveness	10
4.3	Political Stability	11
4.4	Legal Origins	11
5	Other Measures	12
5.1	Fragility Indexes	12
5.1.1	The Brooking Institution	12
5.1.2	Polity IV Fragility Index	12
5.2	GDP	12
5.3	Natural Disasters	13
5.4	Inequality	13
5.5	Private Credit	13
5.6	Corruption	14

5.7	Investment	14
5.8	Membership of Security Council	14
5.9	Aid	15

1 Introduction

This document describes the variables in *Pillars.dta* dataset accompanying Besley, Tim and Torsten Persson, 2011, *Pillars of Prosperity: The Political Economy of Development Clusters*, Princeton University Press. This dataset can be used in conjunction with the do files provided on the book's website to replicate the graphs and tables in the book. Note that in order to generate the maps (Figures 1.1, 1.2, 8.1, 8.2) in the book you need to download *world-c.dta* as well. 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 Measures of State Capacity

2.1 Measures of Fiscal Capacity

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

2.1.1 Measures of the Tax System

Measures of the tax system are taken from [Thomas Baunsgaard and Michael Keen, 2005, Tax Revenue and \(or?\) Trade Liberalization, IMF Working Paper \(WP/05/112\)](#). We take the last year for which a significant number of countries have data, 1999. There are 124 observations (countries) with nonmissing values for *taxrevenuegdp99* and *incometax_share99*. For *tradetax_share99* and the other two variables constructed from this variable we have 123 observations.

taxrevenuegdp99 Share of Taxes in GDP (1999)

The ratio of total tax revenue to GDP in 1999. This is directly taken from Baunsgaard and Keen (2005).

incometax_share99 Share of Income Taxes in Total Taxes (1999)

The share of revenue from income taxes in total taxes in 1999. This is constructed by dividing income tax revenue by tax revenue in 1999 from Baunsgaard and Keen (2005).

tradetax_share99 Share of Trade Taxes in Total Taxes (1999)

The share of revenue from trade taxes in total taxes in 1999. Dividing trade tax revenue by tax revenue in 1999 from Baunsgaard and Keen (2005) will generate this variable.

nontradetax_share99 Share of Non-Trade Taxes in Total Taxes (1999)

This variable is constructed from the above variable as: $1 - \text{tradetax_share99}$.

taxshare_bias99 Difference between income and trade tax shares (1999)

This is constructed by subtracting share of trade taxes from share of income tax in total tax revenue ($\text{incometax_share99} - \text{tradetax_share99}$).

2.1.2 Introduction of VAT and Income Taxes

These variables are collected by the authors.

vatdummy_ext VAT dummy

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

2.1.3 Measures of the Informal Economy

We use the variable Informal Economy in % of GNP 1999/2000 from [Friedrich Schneider, 2002, Size and Measurement of the Informal Economy in 110 Countries Around the World, mimeo](#). Effectively we are able to match 109 countries with our data and the observations are for 1999/2000.

minform Size of Informal Economy % GNP (1999/2000)

This is the original variable from Schneider (2002).

formal Size of formal Economy % GNP (1999/2000)

Size of the formal economy is defined as $1 - \frac{\text{minform}}{100}$.

2.2 Legal Capacity

2.2.1 Government Antidiversion Policies

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.

mgadp97 Property Rights Protection Index

This variable tries to measure the extent 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.

2.2.2 Doing Business Indexes

We use the following variables from [World Bank's Doing Business Project](#) in 2006. There are 172 countries with non missing observations in this year.

mcontracts06 Index of contract enforcement

This variable contains the country's normalized rank on the measure of enforcing contracts in 2006.

mproperty06 Ease of registering property(rank)

This variable contains the country's normalized rank on ease of registering property in 2006.

mbusiness06 Ease of doing business (rank)

This variable contains the country's normalized rank on the overall Doing Business indicator in 2006.

mcredit06 Ease of getting credit (rank)

This variable contains the country's normalized rank in the ease of access to credit in 2006.

3 Measures of Violence

3.1 Civil War

The following variables are all constructed from the measure of civil war incidence taken from [UCDP/PRIO Armed Conflict Dataset version 4-2007, 1946-2006](#) produced by peace research institutes in Oslo and Uppsala.

wardum civil war

This variable takes the value of one if there war incidence measure is equal to 1.

avg_civilwar00 Proportion years in civil war 1950-2000

This variable shows the proportion of years with civil war (where war incidence measure is equal to one) over the years without civil war over 1950-2000 for each country (excluding missing values).

avg_civilwar75 Proportion of years in civil war 1975-2006

This is constructed in the same way as the above variable but for 1975 - 2006 period.

countryavg_civilwar Proportion Years in Civil War 1950-2006

This is the same variable as *avg_civilwar00* except for the fact that it uses a longer time span.

yearavg_civilwar Prevalence of Civil War

This variable summarizes the fraction of countries in our sample that had civil war (based on incidence measure from ACD) in a given year. This is a time series variable and is defined from 1950 until 2006.

3.2 Repression

The following variable are all constructed from the number of purges variable (*purges*) taken from [Cross-National Time-Series Data Archive](#) .

avg_repression00 Proportion years in repression 1950-2000

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 – 2000 period. There are 195 countries with non missing observations for this variable.

avg_repression75 Average proportion of years in repression 1975-2006

This constructed in the same way as *avg_repression00* but for 1975-2006 period. There are 195 countries with non missing observations for this variable.

countryavg_repression Proportion Years in Repression 1950-2006

This is the same variable as *avg_repression00* except for the fact that it uses a longer time span. Again there are 195 observations.

yearavg_repression Prevalence of Repression

This variable summarizes the fraction of countries with positive purges in any given year using purges variable. This is a time series variable and is defined from 1950 until 2006.

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 incidence of war 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.

violence Repression or Civil War

This variable takes the value of 1 if the ordered variable (*repdum_purPRIO*) takes 1 or 2.

4 Measures of Model Parameters

4.1 Common Interests

avg_war00 Proportion 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 (or independence if later) until 2000. The two binary measures of interstate war and extrastate war from the [Correlates of War \(COW\) database](#) are used to see whether a country is in war with other countries. Specifically if any of these measure are showing a war in a given year that country-year is counted as having war and if both of the variables are nonmissing and zero the country-year has no war. Then the proportion of years in war is calculated as the number of years with war over the total number of non missing (with and without war) years. This variable is defined for 180 countries.

fe_etfra Ethnic Fractionalization

This is ethnic fractionalization index from [Fearon \(2003\)](#) (taken from [Quality of Government database](#)). This variable is defined for 157 countries.

fe_ethom Ethnic Homogeneity

This is defined as $1 - fe_etfra$. This variable is defined for 157 countries. 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 [Polij IV dataset](#).

mxconst00 Average executive constraints up to 2000

This measures the average value of the variable *xconst* from 1800 (or independence date if later) up to 2000. The average is taken over non missing values of *xconst* (values outside [1, 7] are treated as missing). This variable is normalized 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 159 non missing observations for this variable. This variable captures the parameter θ in the model.

goodex High executive constraints (Polity)

This variable is equal to one if the measure of executive constraint (*xconst*) is getting its maximum value (7) for each country and year. Note that goodex is equal to zero if *xconst* is different from 7 and *exconst* variable (from Polity IV) is not missing. This is a panel variable defined for a maximum of 159 countries over 1800-2000.

mgoodex High Executive Constraints (51 countries)

This variable shows the fraction of countries with the highest executive constraint score (*xconst*= 7) for all the countries that the PolityIV provides data on *xconst* from 1800 onwards (These are 51 countries). Note that this is a time series variable that spans 1800-2000.

4.2.2 Alternative Measures of cohesiveness

parl_dem Parliamentary Democracy

This variable measures whether a country has parliamentary democracy. It takes the value of one for country-year observations where *polity2* variable from PolityIV dataset is positive (democracy) and the country is not a presidential system. The variable takes zero otherwise (given *polity2* is non missing).

4.3 Political Stability

mgamma00 Average non-open executive recruitment up to 2000

This measures average values of the sum of *xropen* and *xrcomp* variables in [Polity IV dataset](#) from 1800 (or independence if later) to 2000. 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 normalize 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 158 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).

l_eng English Legal Origin

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

l_soc Socialist Legal Origin

This variable is equal to one if the country has a Socialist legal origin.

l_sca Scandinavian Legal Origin

This variable is equal to one if the country has a Scandinavian legal origin.

l_ger German Legal Origin

This variable is equal to one if the country has a German legal origin.

l_fre French Legal Origin

This variable is equal to one if the country has a French legal origin.

5 Other Measures

5.1 Fragility Indexes

5.1.1 The Brooking Institution

state_weakness Brookings Institution state weakness index (2008)

This is the overall score of state weakness from [Brooking Institution](#).

brookingtweak Quantile in Brooking Index of State Weakness (2008)

This contains the quantiles of state_weakness measure from Brooking Institution.

5.1.2 Polity IV Fragility Index

The variable *fragilityPolityIV* is from Marshall and Cole (2009). This variable contains the quantile of the fragility index for each country. The other variables listed below are dummy variables for each quantile of fragility.

fragilityPolityIV fragility: extreme=4, high=3, moderate=2, low=1, no fragility = 0

extremefragility Dummy=1 if country defined as extremely fragile, PolityIV(2008)

highfragility Dummy=1 if country defined as highly fragile, PolityIV(2008)

moderatefragility Dummy=1 if country defined as moderately fragile, PolityIV(2008)

lowfragility Dummy=1 if country has low fragility, PolityIV(2008)

nofragility Dummy=1 if country has no fragility, PolityIV(2008)

5.2 GDP

We have several variables that are constructed from *rgdpch63* variable taken from [Penn World Tables v.6.3](#). *rgdpch63* is “Real GDP per capita in constant 2005 international prices - chain series”.

loggdp Ln(rgdpch63)

This is natural logarithm of *rgdpch63* and is a panel variable. It is defined over 1950 - 2006 for a maximum of 186 countries.

mloggdp00 Log Income per capita in 2000

This takes the value of *loggdp* in year 2000. It is non missing for 186 countries.

mloggdp90 Log Income per capita in 1990

This takes the value of *loggdp* in year 1990. It is non missing for 173 countries.

mloggdp80 Log Income per capita in 1980

This takes the value of *loggdp* in year 1980. It is non missing for 163 countries.

5.3 Natural Disasters

nat_dum =1 if ext. temperature, flood, tsunamis (tidal waves) or landslides (avalanches)

This is a dummy variable generated from different disaster variables from [EM-DAT database](#). It is equal to one if a country had any of the following natural disasters in a given year: extreme temperature, floods, tsunamis or tidal waves, and landslides or avalanches. It takes zero otherwise.

5.4 Inequality

lowineq Low Inequality (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 per cent. The Gini coefficient is from [Deininger and Squire \(1996\)](#). We get this variable from Quality of Government Database (named *ds_gini*). We take the average of Gini coefficient for each country over all available years of data for that country. This average is used to construct lowineq variable. There are 111 countries with at least one year of observation for the Gini Index in [Deininger and Squire \(1996\)](#). Note that the earliest date for which a country has Gini data is 1947 while the latest date is 1996. Note also that there are only 43 countries with five or more years of observations for Gini in this dataset.

5.5 Private Credit

mpcy97 Private credit / GDP 1997

We constructed this variable by using the private credit as a share of GDP (*pcy*) variable from [Giovanni Favara, 2009, An Empirical Reassessment of the Relationship between Finance and Growth](#). Favara constructed this variable using the IMF's International Financial Statistics (IFS) database. *mpcy97* shows the normalized (divided by standard deviation of *pcy*) value of *pcy* for 1997 and is non missing for 95 countries.

5.6 Corruption

mcorr06 Corruption Perception Index 2006 (Transparency International)

This variable is the normalized value of the overall score of each country in Transparency International's [Corruption Perceptions Index 2006](#). We divide CPI score in 2006 by its standard deviation to obtain *mcorr06* which is non missing for 159 countries.

5.7 Investment

minvest06 Investment % GDP 2006 (PWT)

This is the variable *ci* from [Penn World Tables v.6.3](#) in 2006. *ci* is labeled as "Investment Share of Real Gross Domestic Product per Capita, current price". *minvest06* contains the value of *ci* divided by its standard deviation only for 2006.

5.8 Membership of Security Council

Data on the membership of security council is collected by the authors.

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 212 countries over 1946 - 2005 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 2005.

5.9 Aid

aidgni Aid / GNI

This variable is constructed by dividing current price aid by current price GNI (Atlas Method). Both of these variables are from [World Development Indicators \(WDI\)](#). However we get them from Quality of Government Database.

receivedaidtotalexcldebt Received aid total excl debt

This is total amount of aid received by each country excluding debt repayments. we use the variable "Official Development Assistance, Excl Debt (Constant Prices, 2007 USDmillions)" from the [OECD Development Database](#) on Aid from DAC Members (subset 2a).

References

- [1] Banks, Arthur S., 2011, Cross-National Time-Series Data Archive, Databanks International, Jerusalem, Israel. URL: <http://www.databanksinternational.com>
- [2] Baunsgaard, Thomas and Michael Keen, 2005, Tax Revenue and (or?) Trade Liberalization, IMF Working Paper (WP/05/112). URL: <http://www.imf.org/external/pubs/ft/wp/2005/wp05112.pdf>
- [3] Deininger, Klaus and Lyn Squire, "A New Data Set Measuring Income Inequality", The World Bank Economic Review, 10(3): 565-91, 1996.
- [4] Doing Business 2007: How to Reform, 2006, World Bank, Washington DC: The World Bank Group. <http://doingbusiness.org/>
- [5] EM-DAT: The OFDA/CRED International Disaster Database, Université catholique de Louvain, Brussels, Belgium. URL: www.emdat.net
- [6] Favara, Giovanni, 2009, An Empirical Reassessment of the Relationship between Finance and Growth, mimeo.
- [7] Fearon, James, 2003, "Ethnic and Cultural Diversity by Country", Journal of Economic Growth, vol. 8, 195-222.

- [8] Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand, 2002. Armed Conflict 1946–2001: A New Dataset, *Journal of Peace Research* 39(5): 615–637.
- [9] Harbom, Lotta, Halvard Buhaug, Joachim Carlsen, and Håvard Strand, 2007, UCDP/PRIO Armed Conflict Dataset Codebook: Version 4-2007, Uppsala Conflict Data Program (UCDP) and International Peace Research Institute, Oslo (PRIO).
- [10] Heston, Alan, Robert Summers and Bettina Aten, Penn World Table Version 6.3, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, August 2009.
- [11] International Country Risk Guide, © The PRS Group, Inc., East Syracuse, NY, 1980-present. URL: <http://www.prsgroup.com/>
- [12] La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny (1998), “Law and Finance,” *Journal of Political Economy*, vol. 106, 1113– 1155.
- [13] Marshall, Monty G. and Keith Jaggers. 2010. Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010. Version p4v2010 [Computer File]. College Park, MD: Center for Systemic Peace. URL: <http://www.systemicpeace.org/>.
- [14] Marshall, Monty G. and Benjamin R. Cole, 2009, Global Report 2009: Conflict, Governance, and State Fragility, Center for Systemic Peace and Center for Global Policy. URL: <http://www.systemicpeace.org/>.
- [15] Rice, Susan E. and Stewart Patrick, 2008, Index of State Weakness in the Developing World, The Brookings Institution. URL:<http://www.brookings.edu/>
- [16] Schneider, Friedrich, 2002, Size and Measurement of the Informal Economy in 110 Countries Around the World, mimeo.
- [17] Teorell, Jan, Marcus Samanni, Sören Holmberg and Bo Rothstein. 2011. The Quality of Government Dataset, version 6Apr11. University of Gothenburg: The Quality of Government Institute. URL: <http://www.qog.pol.gu.se>.

[18] World Development Indicators, URL: [http://data.worldbank.org/
data-catalog](http://data.worldbank.org/data-catalog)