

The Economic Effects of Constitutions

by Torsten Persson and Guido Tabellini

Data Appendix

The two data sets used in the book are called respectively (both are Stata data sets):

85cross_7nov.dta: This is the data set for the cross section of 85 countries. Observations are averaged over the period 1990-98 (or the subperiod for which data are available).

60panel_26maj.dta: This is the data set for the panel of 60 countries. Observations are yearly data in the period 1960-98 (or the subperiod for which data are available) After the book went to press, we discovered some mistakes in some variables in this data set. The version posted on our web site has corrected those mistakes. Because of this, some of the tables in our book cannot be reproduced with this dataset. The corrected tables (reproducible with these data) are contained in the file "errata corrige.pdf".

Variables are called in the same way in all three datasets, though some variables are present only in some data sets, depending on how they were used in the book. Precise definitions are provided in Chapters 3, 4 and 5 of the book, available from the authors upon request.¹

Cross sectional observations are identified by the variable *ctrycd* (the IMF country code) or by their name (the variable *country* in the datasets). Yearly observations in the panel are also identified by their year (the variable called *year*).

¹Two variables were incorrectly coded in the book, for one observation. Specifically, the variables *PIND* and *PINDO* for Guatemala were incorrectly coded as 0 in the book, while the correct values are reported in this dataset. For this reason, the tables of the book where these two variables appear cannot be exactly reproduced with this dataset (the differences are negligible, though).

List of variables:

AFRICA: regional dummy variable, equal to 1 if a country is in Africa, 0 otherwise.

AGE: age of democracy, defined as: $AGE = (2000 - DEM_AGE)/200$ and varying between 0 and 1, with US being the oldest democracy (value of 1). Source: see *DEM_AGE*.

ASIAE: regional dummy variable, equal to 1 if a country is in East Asia, 0 otherwise.

AUTOC: indicator of institutionalized autocracy, derived from codings of the competitiveness of political participation, the regulation of participation, the openness and competitiveness of executive recruitment, and constraints on the chief executive.

Source: Polity IV Project (<http://www.cidcm.umd.edu/inscr/polity/index.htm>).

AVELF: index of ethnolinguistic fractionalization, approximating the level of lack of ethnic and linguistic cohesion within a country, ranging from 0 (homogeneous) to 1 (strongly fractionalized) and averaging 5 different indexes. Source: La Porta et al. (1998). For Central and Eastern Europe countries computations follow Mauro (1995) with data from Quain (1999).

CATHO80: percentage of the population belonging to the Roman Catholic religion in 1980. Source: La Porta et al. (1998).

CCG_NET_0: consolidated central government net domestic debt as a percentage of gross national disposable income, in the first year for which a value of *SPL* is available. The Consolidated Central Government (CCG) is defined as follows: budgetary central government plus extra budgetary central government plus social security agencies. This definition of the central government is equivalent to that of general government minus local and regional governments. Source: World Savings Database

CGEXP: central government expenditures as a percentage of GDP, constructed using the item Government Finance - Expenditures in the IFS, divided by GDP at current prices and multiplied by 100. Source: IMF - IFS CD-Rom and IMF - IFS Yearbook.

CGREV: central government revenues as a percentage of GDP, constructed using the item Government Finance - Revenues in the IFS, divided by GDP at current prices and multiplied by 100. Source: IMF - IFS CD-Rom and IMF - IFS Yearbook.

CLIST : indicator for closed party lists. Sources: see *LIST* and *SEATS*.

COL_ESP: dummy variable, equal to 1 if the country is a former colony of Spain or Portugal, 0 otherwise. Source: Wacziarg (1996).

COL_ESPA: Spanish colonial origin, discounted by the years since independence (*T_INDEP*), and defined as $COL_ESPA = COL_ES * (250 - T_INDEP) / 250$. Source: Wacziarg (1996).

COL_OTH: dummy variable, equal to 1 if the country is a former colony of a country other than Spain, or Portugal, or the UK, 0 otherwise. Source: Wacziarg (1996).

COL_OTH_A: defined as $COL_OTH * (250 - T_INDEP) / 250$. See also *COL_ESPA*. Source: Wacziarg (1996).

COL_UK: dummy variable, equal to 1 if the country is a former UK colony, 0 otherwise. Source: Wacziarg (1996).

COL_UK_A: defined as $COL_UK * (250 - T_INDEP) / 250$. See also *COL_ESPA*. Source: Wacziarg (1996).

CON2150: dummy variable for the period in which the current constitutional features originated, equal to 1 if either *YEARELE* or *YEAREG* falls in the period between 1921 and 1950, 0 otherwise. Source: see *YEARREG* and *YEARELE*.

CON5180: dummy variable for the period in which the current constitutional features originated, equal to 1 if either *YEARELE* or *YEAREG* falls in the period between 1951 and 1980, 0 otherwise. Source: see *YEARREG* and *YEARELE*.

CON81: dummy variable for the period in which the current constitutional features originated, equal to 1 if either *YEARELE* or *YEAREG* falls in the period after 1981, 0 otherwise. Source: see *YEARREG* and *YEARELE*.

CONFU: dummy variable for religious tradition, equal to 1 if the majority of population is Confucian/Buddhist/Zen, 0 otherwise. Source: Wacziarg (1996), CIA-The World Factbook 2000.

CPI9500: corruption perception index, measuring perceptions of abuse of power from public officials. Average of the CPI Index over the period 1995-2000, which ranges from 0 to 10, with higher values denoting more corruption. Source: Transparency International (www.transparency.de) and Internet Center for Corruption Research (www.gwdg.de/~uwvw).

COUNTRY : Name of the country.

CTRYCD: IMF country code that identifies countries.

DEM_AGE: first year of democratic rule, corresponding to the first year of an uninterrupted string of positive yearly values of the variable *POLITY*

(see below) until the end of the sample, given that the country was also an independent nation. Does not count foreign occupation during WWII as an interruption of democracy. Source: See *POLITY*.

DEMOC: institutionalized democracy index, derived from codings of the competitiveness of political participation, the regulation of participation, the openness and competitiveness of executive recruitment, and constraints on the chief executive.

Source: Polity IV Project (<http://www.cidcm.umd.edu/inscr/polity/index.htm>).

DCGEXP: first difference of *CGEXP*. Source: see *CGEXP*.

DISTRICTS: the number of electoral districts in a country (including the number of primary as well as secondary and tertiary if applicable). Sources: Quain (1999), Kurian (1998), and national sources.

DSSW: first difference of *SSW*. Source: see *SSW*

DU60CTRY: dummy variable that identifies which of the 85 countries is also included in the 60-country data set.

DU_NOEC: dummy variable for oil exporter.

EDUGER: total enrolment in primary and secondary education, as a percentage of the relevant age group in the population. Computed dividing the number of pupils (or students) enrolled in a given level of education regardless of age by the population of the age-group which officially corresponds to the given level of education, and multiplying the result by 100. Source: UNESCO - Education Indicator - Category Participation. Available on www.unesco.org

ELEX: dummy variable for executive elections, equal to 1 in a year when the executive is elected, and 0 otherwise. Takes into consideration both presidential elections and legislative elections.

Source: <http://www.ifes.org/eguide/eleguide.htm> plus other national sources.

ELLEG: dummy variable for legislative elections, equal to 1 in the year the legislature is elected, independently from the form of government. Source: <http://www.ifes.org/eguide/eleguide.htm> plus other national sources.

$EL_MAJ = MAJ * ELEX$. Source: see *ELEX* and *MAJ*.

$EL_MAJPAR = (1 - PRES) * MAJ * ELEX$. Source: see *ELEX*, *PRES* and *MAJ*.

$EL_MAJPRE = PRES * MAJ * ELEX$. Source: see *ELEX*, *PRES* and *MAJ*.

$EL_PAR = (1 - PRES) * ELEX$. Source: see *ELEX* and *PRES*.

$EL_PRE = PRES * ELEX$. Source: see *ELEX* and *PRES*.

$EL_PRO = (1 - MAJ) * ELEX$. Source: see *ELEX* and *MAJ*.

$EL_PROPAR = (1 - PRES) * (1 - MAJ) * ELEX$. Source: see *ELEX*, *PRES* and *MAJ*.

$EL_PROPRES = PRES * MAJ * ELEX$. Source: see *ELEX*, *PRES* and *MAJ*.

ENGFRAC: the fraction of the population speaking English as a native language. Source: Hall and Jones (1999).

EURFRAC: the fraction of the population speaking one of the major languages of Western Europe: English, French, German, Portuguese, or Spanish. Source: Hall and Jones (1999).

FEDERAL: dummy variable, equal to 1 if the country has a federal political structure, 0 otherwise. Source: Adserà, Boix and Paine (2001).

FRANKROM: natural log of the Frankel-Romer forecasted trade share, derived from a gravity model of international trade that only takes into account country population and geographical features. Source: Hall and Jones (1999).

GADP: index of government's anti-diversion policies, measured over the period 1986-95. It is an equal-weighted average of these five categories: i) law and order, ii) bureaucratic quality, iii) corruption, iv) risk of expropriation and v) government repudiation of contracts (each of these items has higher values for governments with more effective policies towards supporting production) and ranges from zero to one. Source: Hall and Jones (1999).

GASTIL: average of indexes for civil liberties and political rights, where each index is measured on one-to-seven scale with one representing the highest degree of freedom and seven the lowest. Countries whose combined averages for political rights and for civil liberties fall between 1.0 and 2.5 are designated "free", between 3.0 and 5.5 "partly free" and between 5.5 and 7.0 "not free". Source: Freedom House, Annual Survey of Freedom Country Ratings.

GDP: gross domestic product at current price. Source: IFS CD-Rom and IFS Yearbook.

GINI_8090: Gini index on income distribution, computed as the average of two data points: the observation closest to the 1980 and the observation closest to the 1990. When only one of the two years year is available, only that year is included. Source: Deininger and Squire (1996).

GOVEF: point estimate of "Government Effectiveness", the third cluster of the Kaufmann et al.(1999a) governance indicators. It combines perceptions of the quality of public service provision, the quality of the bureaucracy,

the competence of civil servants, the independence of the civil service from political pressures, and the credibility of the government's commitment to policies into a single grouping. It ranges from around 0 to around 10 (lower values correspond to better outcome). Sources: Kaufmann et al. (1999a.), available at <http://www.worldbank.org/wbi/gac>.

GRAFT: point estimate of "Graft", the sixth cluster of Kaufmann et al.'s governance indicators, focusing on perceptions of corruption. It ranges from around 0 to around 10 (lower values correspond to better outcome). Sources: Kaufmann et al. (1999a.), available at www.worldbank.org/wbi/gac.

LAAM: regional dummy variable, equal to 1 if a country is in Latin America, Central America or the Caribbeans, 0 otherwise.

LAT01: rescaled variable for latitude, defined as the absolute value of *LATITUDE* divided by 90 and taking values between 0 and 1. Source: Hall and Jones (1999).

LATITUDE: distance from the equator (in degrees), ranging between -90° to 90° . Source: Hall and Jones (1999).

LCGEXP: one-year lag of *CGEXP*. Source: see *CGEXP*.

LCGREV: one-year lag of *CGREV*. Source: see *CGREV*.

LEGOR(UK, FR, GE, SO, SC) : dummy variables for the origin of the legal system, classifying a country's legal system into Anglo-Saxon Common Law (UK), French Civil Law (FR), German Civil Law (GE), Socialist Law (SO), or Scandinavian Law (SC). Source: La Porta et al. (1998).

LELEX: One year lag of *ELEX*. Source: see *ELEX*.

LELMAJ: One year lag of *ELMAJ*. Source: see *ELMAJ*.

LELMAJPAR: One year lag of *ELMAJPAR*. Source: see *ELMAJPAR*.

LELMAJPRE: One year lag of *ELMAJPRE*. Source: see *ELMAJPRE*.

LELPRO: One year lag of *ELPRO*. Source: see *ELPRO*.

LELPROPAR: One year lag of *ELPROPAR*. Source: see *ELPROPAR*.

LELPROPPE: One year lag of *ELPROPPE*. Source: see *ELPROPPE*.

LIST: number of lower-house legislators elected through party list systems. Sources: Quain (1999) and Kurian (1998).

LOGA: natural log of total factor productivity, measured in 1988. Source: Hall and Jones (1999).

LOGYL: natural log of output per worker, measured in 1988. Source: Hall and Jones (1999).

LPOP: natural log of the total population (in millions). Source: World Bank

LSPL:

LSSW: It represents the one-period lagged series of *SSW*. Source: IMF - GFS Yearbook 2000 and IMF - IFS CD-Rom.

LYP: natural log of per capita real GDP (*RGDPH*). *RGDPH* is defined as real GDP per capita in constant dollars (chain index) expressed in international prices, base year 1985. Data through 1992 are taken from the Penn World Table 5.6 (variable named *RGDPC*), while data on the period 1993-98 are computed from data taken from the World Development Indicators, the World Bank. These later observations are computed on the basis of the latest observation available from the Penn World Tables and the growth rates of GDP per capita in the subsequent years computed from the series of GDP at market prices (in constant 1995 U.S. dollars) and population, from the World Development Indicators. Sources: Penn World Tables - mark 5.6 (PWT), available on <http://datacentre2.chass.utoronto.ca/pwt/docs/topic.html>. The World Bank's World Development Indicators; www.worldbank.org.

MAGN: inverse of district magnitude, defined as *DISTRICTS* over *SEATS*. Sources: see *DISTRICTS* and *SEATS*.

MAJ: dummy variable for electoral systems. Equals 1 if all the lower house is elected under plurality rule, 0 otherwise. Only legislative elections (lower house) are considered. Sources: Cox (1997), International Institute for Democracy and Electoral Assistance (1997), Quain (1999), Kurian (1998), and national sources.

$MAJ_BAD = MAJ * GASTIL$. Source: see *MAJ* and *GASTIL*

$MAJ_GIN = MAJ * GINI_8090$. Source: see *MAJ* and *GINI_8090*

$MAJ_OLD = MAJ * AGE$. Source: see *MAJ* and *AGE*

$MAJPAR = MAJ * (1 - PRES)$. Source: see *MAJ* and *PRES*.

$MAJPRES = MAJ * PRES$. Source: see *MAJ* and *PRES*.

MINING_GDP: share of mining sector over GDP. Source: UN National accounts.

MIXED: dummy variable for electoral systems, equal to 1 if the electoral formula for electing the lower house is neither strict plurality rule nor strict proportionality, 0 otherwise. Semi-proportional (or mixed) electoral rule identifies those electoral systems characterized by both proportional and first-past-the-post representation for allocating seats (for example Bolivia, Germany, Italy after the reform of 1993, etc.). The share of the total number of seats allocated under the Proportional rule can be greater or smaller than the complementary plurality-allocated share. Only legislative elections are considered. Sources: Cox (1997), International Institute for Democracy and Electoral Assistance (1997), Quain (1999), and Kurian (1998) and national

sources.

NEGYG: negative values of *YGAP*, 0 if *YGAP* is positive. Source: see *YGAP*.

OECD: dummy variable, equal to 1 for all countries that were members of OECD before 1993, 0 otherwise, except for Turkey coded as 0 even though it was a member of OECD before the 1990s.

OIL: price of oil in US dollars. Source: Datastream.

OIL_EX: *OIL* times a dummy variable that equals 1 if net exports of oil are positive, 0 otherwise. Source: See *OIL*

OIL_IM: *OIL* times a dummy variable that equals 1 if net exports of oil are negative, 0 otherwise. Source: See *OIL*

PIND: continuous measure of the ballot structure, defined as $1 - \frac{LIST}{SEATS}$. Source: see *LIST* and *SEATS*.

PINDO: continuous measure of the ballot structure defined as $1 - \frac{LIST}{SEATS} CLIST$. Source: see *LIST* and *SEATS*.

POLITY: score for democracy, computed by subtracting the *AUTO*C score from the *DEMOC* score, and ranging from +10 (strongly democratic) to -10 (strongly autocratic). Source: Polity IV Project (<http://www.cidcm.umd.edu/inscr/>)

POLITY_GT: interpolated version of *POLITY*, rescaled with the same units of *GASTIL* (i.e. higher values denote worse democracies). Computed as the forecasted value obtained by regressing the rescaled values of *POLITY* on *GASTIL*. Source: see *POLITY* and *GASTIL*.

POSYG: positive values of *YGAP*, 0 if *YGAP* is negative. Source: see *YGAP*.

PRES: dummy variable for forms of government, equal to 1 in presidential regimes, 0 otherwise. Only regimes where the confidence of the assembly is not necessary for the executive (even if an elected president is not chief executive, or if there is no elected president) are included among presidential regimes. Most semi-presidential and premier-presidential systems are classified as parliamentary (see the text in Chapter 4 for further discussion and clarification). Source: Shugart and Carey (1992) and national sources).

$PRES_BAD = PRES * GASTIL$. Source: see *PRES* and *GASTIL*

$PRES_GIN = PRES * GINI_8090$. Sources: see *PRES* and *GINI_8090*

$PRES_OLD = PRES * AGE$. Source: see *PRES* and *AGE*

PROP1564: percentage of population between 15 and 64 years old in the total population. Source: World Development Indicators CD-Rom 1999.

PROP65: percentage of population over the age of 65 in the total population. Source: World Development Indicators CD-Rom 1999.

$PROPAR = (1 - MAJ) * (1 - PRES)$. Source: see *MAJ* and *PRES*.

$PROPRES = (1 - MAJ) * PRES$. Source: see *MAJ* and *PRES*.

PROT80: percentage of the population in each country professing the Protestant religion in 1980. Source: La Porta et al. (1998).

SDM: district magnitude (i.e., as seats over districts), computed as a weighted average, where the weight on each district magnitude in a country is the share of legislators running in districts of that size. Source, Seddon et al. (2001). Relative to the original variable in Seddon et al. (2001), this variable is divided by 100 so that it takes values comparable to those of *MAGN*.

SEATS: number of seats in lower or single chamber for the latest legislature of each country. It is also related to the number of districts in which primary elections are held. Source: International Institute for Democracy and Electoral Assistance (1997), Quain (1999), Kurian (1998) and national sources.

SPL: central government budget surplus (if positive) or deficit (if negative), as a percentage of GDP, constructed using the item Government Finance - Deficit and Surplus in the IFS, divided by the GDP at current prices and multiplied by 100. Source: IMF - IFS CD-Rom and IMF - IFS Yearbook.

SPROPN: share of legislators elected in national (secondary or tertiary) districts rather than sub-national (primary) electoral districts. Source: Seddon et al. (2001)

SSW: consolidated central government expenditures on social services and welfare as percentage of GDP, as reported in GFS Yearbook, divided by GDP and multiplied by 100. Source: IMF - GFS Yearbook 2000 and IMF - IFS CD-Rom.

T_INDEP : years of independence, ranging from 0 to 250 (the latter value is used for all non-colonized countries). Source: Wacziarg (1996).

TRADE: sum of exports and imports of goods and services measured as a share of GDP. Source: The World Bank's World Development Indicators CD-Rom 2000.

YEAR : Year to which the panel observations refer.

YEARELE: the year when the current electoral rule, as coded by *MAJ*, was first introduced, or the first year of democratic rule, whatever came last.

YEARREG: the year when the current form of government, as coded by *PRES*, was first introduced, or the first year of democratic rule, whatever came last.

YRSOPEN: index for openness to international trade, compiled by

Sachs and Werner (1995), measuring the fraction of years during the period 1950-1994 that the economy has been open and ranging between 0 and 1. Source: Hall and Jones (1999).

YGAP : deviation of aggregate output from its trend value in percent, computed as difference between the natural log of real GDP in the country and its country-specific trend (obtained, using the Hodrick-Prescott filter). Source for real GDP: World Bank.