Relative Political Capacity Dataset Documentation

Version 2.4

Last Update: August 2020

Principal Investigators:

Ali Fisunoglu, Kyungkook Kang,
Marina Arbetman-Rabinowitz, Jacek Kugler.

TransResearch Consortium

This document serves as a codebook for the Relative Political Capacity Dataset version 2.4, which consists of three measures: Relative Political Extraction (RPE), Relative Political Reach (RPR), and Relative Political Allocation (RPA) as well as a the newly developed measure of Absolute Political Extraction (APE). The conceptual definition and use of Relative Political Capacity is generally discussed in Organski and Kugler (1980), Arbetman and Kugler (1997), Kugler and Tammen (2012).

To acquire the datasets, components of RPE, RPR and RPA, and the detailed codebook, please visit:

http://www.transresearchconsortium.com/data or https://dataverse.harvard.edu/dataverse/rpc

The recommended citation for these datasets is: Ali Fisunoglu; Kyungkook Kang; Marina Arbetman-Rabinowitz; Jacek Kugler. (2020) Relative Political Capacity Dataset (Version 2.4), https://doi.org/10.7910/DVN/NRR7MB, Harvard Dataverse.

NOTE ON RELATIVE POLITICAL CAPACITY DATASET VERSION 2.4 (2020) UPDATE

The 2020 update for the Relative Political Capacity Dataset had four main purposes. First, whenever possible, researchers re-documented the source for the data points. Second, the data is extended from 2015 to 2018. Third, the set of countries that are covered is updated; we now estimate the political performance measures for countries with population above 350000 in 2018, although we keep providing the components for all countries. Finally, in some cases, the covered time span for the countries is improved.

Dataset Layout:

Variables for the rpc_2020:

- cowcode: Correlates of War Country code (for available countries)
- trccode: TransResearch Consortium Country code
- imfcode: International Monetary Fund Country code (for available countries)

- uncode: United Nations Country code (for available countries)
- ISO3: ISO 3166 alpha-3 code
- country: Country name
- year: Year of observation
- ape1: Absolute Political Extraction, calculated using Model 1 (see below)
- ape1n: Absolute Political Extraction, calculated using Model 2 (see below) with the adjusted non-resource taxes
- rpe_agri: Relative Political Extraction, calculated using Model 3 (see below)
- rpe_gdp: Relative Political Extraction, calculated using Model 4 (see below)
- rpe_gdp_nonres: Relative Political Extraction, calculated using Model 5 (see below) with the adjusted non-resource taxes.
- rpr_work: Relative Political Reach, calculated using Model 6 (see below)
- rpr_eap: Relative Political Reach, calculated using Model 7 (see below)
- rpa_full: Relative Political Allocation, calculated using Model 8 (see below)
- rpa_sub: Relative Political Allocation calculated using Model 8 (see below) with income partitions

VARIABLES	N	Mean	Min	Max	# of countries
ape1	8589	0.471	0.012	0.899	164
ape1n	8589	0.394	0.003	0.891	164
rpe_agri	8,660	1.026	0.011	3.832	170
rpe_gdp	8,660	0.995	0.012	3.769	170
rpe_gdp_nonres	8,660	1.001	0.007	3.473	170
rpr_work	8,163	1.000	0.259	1.875	161
rpr_eap	8,163	1.000	0.569	1.643	161
rpa_full	4,260	1.000	0.366	1.366	154
rpa_sub	4,260	1.000	0.326	1.372	154

COMPONENTS OF ABSOLUTE AND RELATIVE POLITICAL CAPACITY

Absolute Political Extraction:

APE uses Stochastic Frontier Analysis to directly measure the extractive capacity of nations. APE is estimated by multiplying these two different models with life expectancy:

All societies – General Sample (Model 1) (ape1):

$$\ln\left(\frac{\text{Tax}}{\text{GDP}}\right) = \alpha + \beta_1 \ln\left(\frac{\text{Mining}}{\text{GDP}}\right) + \beta_2 \ln\left(\frac{\text{Exports}}{\text{GDP}}\right) + \beta_3 \ln\left(\frac{\text{Social Contributions}}{\text{GDP}}\right) + \beta_4 \ln(\text{GDP per capita})$$
$$+ \beta_5(\text{Economically Active Population}) + \beta_6(\text{Education}) + \beta_7(\text{OECD}) + \beta_8(\text{Inclusion Dummy}) + \varepsilon$$

All societies – General Sample (Model 2) (ape1n):

$$\ln\left(\frac{\text{Non-Resource Tax}}{\text{GDP}}\right) = \alpha + \beta_1 \ln\left(\frac{\text{Mining}}{\text{GDP}}\right) + \beta_2 \ln\left(\frac{\text{Exports}}{\text{GDP}}\right) + \beta_3 \ln\left(\frac{\text{Social Contributions}}{\text{GDP}}\right) + \beta_4 \ln(\text{GDP per capita})$$
$$+ \beta_5(\text{Economically Active Population}) + \beta_6(\text{Education}) + \beta_7(\text{OECD}) + \beta_8(\text{Inclusion Dummy}) + \varepsilon$$

where

Tax / GDP = General Govt. Tax Revenues / GDP

Non-Resource Tax / GDP = General Govt. Tax Revenues excluding mineral revenues / GDP

Mining / GDP = Total mining revenues / GDP

Exports / GDP = Value of exported goods and services / GDP

Social Contributions = Social Contributions / GDP

GDP per capita = GDP per capita in 2010 constant US dollars

Economically Active Population = Labor Force Participation Rate

Education = Gross Secondary Enrollment Ratio

OECD = Dummy variable, which takes 1 if the country is an OECD member, 0 otherwise

Inclusion Dummy = Dummy variable, which takes 1 if the country is to be included in the regression, 0 otherwise.[1]

[1] Cases that are included are selected based on the population of the country and data availability and quality.

Relative Political Extraction:

RPE approximates the ability of governments to appropriate portions of the national output to advance public goals. Three equations are used to estimate RPE:

Developing societies (Model 3) (rpe_agri):

$$\frac{\text{Tax}}{\text{GDP}} = \alpha + \beta_1(\text{time}) + \beta_2\left(\frac{\text{Mining}}{\text{GDP}}\right) + \beta_3\left(\frac{\text{Agriculture}}{\text{GDP}}\right) + \beta_4\left(\frac{\text{Exports}}{\text{GDP}}\right) + \beta_5(\text{OECD}) + \beta_6(\text{Inclusion Dummy}) + \varepsilon$$

Developed societies – General Sample (Model 4) (rpe_gdp):

$$\frac{\text{Tax}}{\text{GDP}} = \alpha + \beta_1(\text{time}) + \beta_2\left(\frac{\text{Mining}}{\text{GDP}}\right) + \beta_3\left(\frac{\text{Exports}}{\text{GDP}}\right) + \beta_4(\text{GDP per capita}) + \beta_5(\text{OECD}) + \beta_6(\text{Inclusion Dummy}) + \varepsilon$$

Developed societies – General Sample (Model 5) (rpe_gdp_nonres):

$$\frac{\text{Non-Resource Tax}}{\text{GDP}} = \alpha + \beta_1(\text{time}) + \beta_2\left(\frac{\text{Exports}}{\text{GDP}}\right) + \beta_3(\text{GDP per capita}) + \beta_4(\text{OECD}) + \beta_5(\text{Inclusion Dummy}) + \varepsilon$$

where

Tax / GDP = General Govt. Tax Revenues / GDP

Non-Resource Tax / GDP = General Govt. Tax Revenues excluding mineral revenues / GDP

Mining / GDP = Total mining revenues / GDP

Agriculture / GDP = Total agriculture revenues / GDP

Exports / GDP = Value of exported goods and services / GDP

GDP per capita = GDP per capita in 2010 constant US dollars

OECD = Dummy variable, which takes 1 if the country is an OECD member, 0 otherwise

Inclusion Dummy = Dummy variable, which takes 1 if the country is to be included in the regression, 0 otherwise.[2]

[2] Cases that are included are selected based on data availability and quality.

Relative Political Reach:

RPR gauges the capacity of governments to mobilize populations under their control. The equations that is used to estimate RPR:

Model 6 (rpr_work):

$$\frac{\text{Activity Rate}}{\text{Population}} = \alpha + \beta_1(\text{time}) + \beta_2\left(\text{Education}\right) + \beta_3\left(\text{Young Population}\right) + \beta_4\left(\text{Social Security}\right)$$
$$+\beta_5(\text{Urbanization}) + \beta_6(\text{Populaton}) + \beta_7(\text{GDP per Capita}) + \beta_8\left(\text{Bureaucracy}\right) + \beta_9(\text{Inclusion Dummy}) + \varepsilon$$

Model 7 (rpr_eap):

$$\frac{\text{EAP}}{\text{Population}} = \alpha + \beta_1(\text{time}) + \beta_2\left(\text{Education}\right) + \beta_3\left(\text{Young Population}\right) + \beta_4\left(\text{Social Security}\right)$$
$$+\beta_5(\text{Urbanization}) + \beta_6(\text{Populaton}) + \beta_7(\text{Unemployment}) + \beta_8(\text{GDP per Capita}) + \beta_9\left(\text{Bureaucracy}\right)$$
$$+\beta_7(\text{Inclusion Dummy}) + \varepsilon$$

where

Activity Rate / Population = (Economically Active Population - Unemployment) / Population

EAP / Population = Economically Active Population / Total Population

Education = Secondary Education Attainment

Young Population = Population ages 0-14 / Total Population

Social Security = Social Security Taxes / GDP

Urbanization = Urban Population / Total Population

Population = Total Population

Unemployment = Unemployment rate

GDP per Capita = GDP per capita in 2010 constant US dollars

Bureaucracy = Expenditures in Government Wages / Total Expenditures

Inclusion Dummy = Dummy variable, which takes 1 if the country is to be included in the regression, 0 otherwise[1]

Relative Political Allocation:

Relative Political Allocation is a composite indicator to measure how public expenditures are prioritized in the government budget. The idea is that either over or under-allocating sectorial resources will hinder societies' path to prosperity. RPA measures the gaps between the actual percentage distribution of general government outlays and what the estimated percentages of the "best" expenditure portfolio to allocate to each functional area to maximize the living standards of the population. Such allocations echo the explicit choices governments make in distributing public resources and relate them to the operation of the economy at their levels of socioeconomic development.

Our conceptualization of political allocation is concerned with exploring the impact of four key government programs on aggregate productivity of the society. Following the classification of Saunders and Klau (1985) and Oxley et al. (1990), we first decompose general government expenditures into four key functional areas:

(A) Traditional Public Goods (or called "Traditional Domain"): administration of general public service, maintenance of public order, military defense of the territory, environmental protection, and other expenses.
(B) Economic Services (or "Mixed Economy"): government activities to assist the economic functioning of the private sector.
(C) Education Resources (or "Merit Goods of Welfare State"): socially desirable for improving productivity but under-produced and under-consumed by the market, which includes educational and cultural services.
(D) Social Protection Expenditure (or "Redistribution of Welfare State"): policy program designed to protect vulnerable populations by redistributing resources for housing, health, and income when they are unable to care for themselves.

How much a government spends in each budget category reflects their political choices and priorities to provide output value to society. Based on such functional area data, we estimate the optimal ratios of distribution that would maximize annual income per capita of each society given the level of development

(1) Setup:

For the first step, we define each country's distinct production function that specifies the output of an entire economy for all combination of productivity indexes, the net increase in physical assets, and the government expenditure outlays. This social production function denotes the political economic process of converting private capital inputs and public spending into total income of the society. Suppose a country's annual output y is produced using physical capital k and government functional outlays g. We simply express this formulation as be simply expressed as the following Cobb-Douglas combination (a lower-case letter denotes the per capita amount):

$$y = AMk^{\alpha} \prod_{j} g_j^{\beta_j} \quad , \qquad 0 < \alpha, \beta < 1$$

where A denotes the society's stock of technology and M is any other factors that influence productivity, both of which are not explained by levels of spending on k and g. Note j indexes four key government functions by attempting group public expenditures in line with the Classification of the Functions of Government (COFOG), developed by OECD. Specifically, our composition of government spending corresponds to each item of the following COFOG categories:

RPA Function (j)	COFOG categories (code)
Traditional Public Goods	General Public Spending (01), Defense (02), Public Order and Safety (03), Environmental Protection (05)
Economic Services	Economic affairs (04)
Education Resources	Recreation, Culture and Religion (08), Education (09)
Social Protection	Housing and Community Amenities (06), Health (07), Social Protection (10)

(2) Estimation:

The government budget that maximizes growth is estimated based on different combinations of these expenditure categories that vary with the level of development achieved by each society. According to Kang (2015), the first-order condition of profit maximization shows that a government can maximize the total output by allocating its public expenditures in accordance with the optimal ratios defined as:

$$\beta_j = \frac{g_j^*}{y}$$

We recognize that the public expenditure share that enters the function with higher power is more productive than others with lower powers. This implies that a government can improve its economic performance by adopting an alternative fiscal policy strategy to increase the expenditure of public sectors with the higher β coefficients while to limit public spending on other sectors with the lower β s. In other words, the estimated coefficient can yield an indication of the extent to which an optimal production input share deviates from actual input shares given production function.

Based on Battese and Coelli (1995)'s method, we estimate the frontier production function with optimal spending ratios reflected in beta coefficients:

Model 8 (rpa_full and rpa_sub):

$$\ln y_{it} = \ln A + \gamma \ln k_{it} + \sum_h \alpha_h m_{hit} + \sum_j \beta_j \ln g_{jit} + v_{it} - u_{it}$$

	Variable
y	Per capita Gross Domestic Product per capita (2010 prices –US dollars)
k	Per capita Gross Fixed Capital Formation
m	Secondary School Enrollment Gross Ratio
	Economic Openness as (Export+Import)/GDP
	OECD Membership as of 1990

g	Per capita general government expenditure on Traditional Public Goods
	Per capita general government expenditure on Economic Services
	Per capita government spending on Education Resources
	Per capita government spending on Social Protection

All y , k , g variables are measured at 2010 prices –US dollars.

(4) With obtained the coefficients of β , we construct a distance function as a "social allocative inefficiency score" for each country to measure difference between and the optimal public spending ratios. The social allocative inefficiency for each public sector j is defined as :

$$\left| \frac{\beta_j}{\sum \beta} - \frac{\text{public expenditure on } j}{\text{Total public expenditure}} \right|$$

By summing up all absolute distance values for every public sector, then we obtain the aggregate social inefficiency score for each country.

(5) For the final step, we then calculate the Relative Political Allocation (RPA) index from the obtained allocation inefficiency scores as follows (Model 9):

$$RPA = \left(1 - \frac{\text{Ineff. of Country } i \text{ – Average of Ineff. in the sample}}{\text{Max of Ineff. in the sample + Min of Ineff. in the sample}} \right)$$

The average RPA is scaled to 1. The value greater than 1 indicates the country allocates its public expenditure in a relatively efficient way, whereas the value smaller than means a relatively inefficient allocation.

We have two alternative estimations of RPA.

- rpa_full is calculated from a result from a whole sample regression (showing general relationships with overall comparisons).

- rpa_sub is calculated from 4 different results of income level specific group regressions – less than $1000 GDP per capita, $1000 to $4000, $4000 to $12000, and more than $12000 (controlling level of development and providing particular policy guidance when the characteristics of the comparison groups are similar).

GENERAL CONSIDERATIONS

Data Collection

Sources for RPC are chosen on the basis of consistent, reliable data that is regularly reported and in comparable units. Sources share some standardization in reporting format and methods (e.g., the IMF, World Bank, OECD). The World Bank used to be the depository of all data collection; however, since 1972 the IMF has assumed responsibility for the collection. The IMF relies on national reporting—countries fill out forms required by the IMF, and the IMF does not necessarily audit the data. Data can be reported and then updated by countries of origin, resulting in differences for IMF reporting for a country at different times. As is the case with any data reliant on national reporting, accuracy is not guaranteed; however, the data are internally consistent. Certainly, more transparent countries and those with established internal checks and balances tend toward greater accuracy in reporting. Political objectives such as external debt may contribute to inaccuracy in reporting for some countries, even countries that are considered developed, as well. Countries facing these problems may create ghost accounts in order to appear to be in better circumstances to external parties, resulting in inaccuracy in the most recent data. These reporting variances are typically adjusted with the passage of time. Updates in RPC require going back five to ten years to assess and correct all of these reporting errors. National Central Bank reports and figures can also be used to fill holes in IMF reporting.

In some instances, different institutions stop reporting particular items or alter reporting mechanisms. Tax structures can dramatically change, as can currencies. When these changes occur, transition years can contain reporting under each structure or both currencies, making it difficult to look at the measure consistently.

Methodology for Overlapping Time Series

Comparisons of the rate of change are an initial starting point for making comparisons between series. Series often reflect different measurement units or methods. Consistency in the variance in observations contained in both series can be a basis for some confidence that the series are capturing the same variables. There has been significant improvements in data quality in recent years due to technological advances and improved training programs offered to bureaucrats and civil servants by international organizations throughout the world. Relying on the most recent and reliable data points and applying the rate of change from older sources prevents artificial introduction of variance created by differing measurement techniques or reporting agencies. Historical sources must be referenced in order to assure that steady change is occurring, and fluctuations are not a result of changing data sources, measurement techniques, differing conversion factors, or introduction of new components. The occurrence of dramatic or drastic change such as a military coup, financial crisis, or natural disaster can massively affect the rate of revenue collection and national expenditure.

Quality Control

As in any other data collection situation, we need to consider quality control to ensure highest reliability and validity. The accuracy of the sources and the quality of the transcription of the data are of utmost importance. But there are also problems related to the level of change from year to year of the data that cannot be accounted for by a normal trend. For example, countries that suffer from high level of inflation might report data at different times of the year that are not comparable. Countries might change their currencies, and the same problem is present: those two pieces of information are not comparable. These problems are usually highlighted when we work with ratios. Ratios that are very different raise a red flag to check for the reason for the outlier and permit us to ‘‘translate’’ all the information to common denominator. Another problem that often presents is the change of accounting systems. Not all countries were able to adapt to the revisions to the System of National Accounts in 1993 and 2008, which creates higher variations around those years, but working with rations minimizes the potential discrepancies in the data related to the revisions. So, for quality control, working with absolute numbers in developing countries might lead us to ignore problems of the data itself. Of course, not all variation is a

result of data quality, problems related to politics or natural disasters have their own explanations.

EMPIRICAL CHALLENGES

Demographic and economic data can suffer from poor reporting and variance in reporting, particularly in countries without strong institutions that collect information on regular basis, or does not allocate sufficient resources to sampling or civil servant preparedness. Countries experiencing high levels of conflict may experience long periods of time without a complete or even partial census taking place. An additional problem evident with census data is that predictable but infrequent intervals in census data require some filling of the data to achieve a complete cross-sectional time-series dataset. Although labor data gets adjusted with sampling and there are constant modelling and adjustments between the censuses, the interpolations can cause biases or miss unexpected temporary variations. Some secondary sources are available to fill in additional years; however, some variance is lost in the process. In the current estimation, we used linear estimation techniques on a country by country basis.

Tax revenues can suffer from poor collection and variations in reporting, particularly in countries without strong financial institutions that collect information on regular basis. Even in countries characterized by strong institutions and transparency, changes in accounting, which is not always accompanied by good documentations, can contribute to the distortion of revenue figures. Inflation and the timing of reporting and collection that do not always coincide with active inflation also contribute to the distortion of revenues and the allocation of expenditures and makes the reporting date an critical issue, although monthly deflators correct these fluctuations.

Social security data may also suffer from poor reporting, and the differences between and within countries (e.g., the elimination or implementation of a social security program, changes in the structure of collection and disbursement) compound the difficulty in accurate measurement across societies. At the same time, some countries have semi-privatized or privatized systems, making inter-country comparisons more challenging. For the same reason, health expenditures also are difficult to measure comparatively across societies. Countries have different mixes of private and public healthcare systems that drastically alter the level of expenditures on healthcare

by individual governments. These same countries change the degree of government involvement in healthcare provision over time, and different budget line items may be included under the umbrella of healthcare in some countries that may fall under other social services in others (e.g., hospice care and nursing homes in Scandinavian countries are included in healthcare expenditures—these are frequently private expenses in other countries). Although health expenditures have also been attached to developed countries' array of policy choices (private vs. public), more and more developing countries are making similar decisions both in social security and health.

For the education data, although we use a cross-national base for the data, we do rely on domestic sources for large amounts of the remaining required data. National governments can be politically motivated to inflate or address these education measures, a problem that occurs with the more historical data. Increased transparency and governance in data collection efforts, in addition to more widespread survey research, conducted cross-nationally, provide triangulation points for much of the current data. Fortunately, we are able to identify where historical numbers are grossly inflated by evaluating country by country trends over time. Another issue is the varying definitions used by different countries. For example, work force is counted from different ages, which makes cross country data not comparable in some cases. A final issue occurs with the definition of secondary education. While cross-national sources do contain consistent definitions for this measure, some national governments have different divisions for primary, secondary, and tertiary education levels. Countries structure their education differently and the duration of each segment consists of different number of years, again making cross country comparisons inadequate.

Another challenge we face is the occasional collapse or split of countries, creation of new countries, and changes of borders. In these instances, we try to historically restructure the RPC considering the new borders. Nevertheless, RPC shows great variance around the years when new countries are established. This is partly due to the political instability that often accompanies the establishment of countries and partly due to the weak recording of data and statistical capacity as the institutions that are responsible for data collection are just being established. Therefore, users should be cautious for the years around which a country is established.

Sample Size

There is an indirect trade-off between the expansion of the sample to include all countries and accuracy and comparability of the data. For some countries, particularly extremely poor developing countries, it is diffucult to obtain accurate reporting of the components of RPC for the duration of the series. Inaccuracies in the data and gaps in reporting influence whether data within a given country is reliable. Ex-Soviet and Eastern European countries reported net material product until the mid-1980s, and those series are difficult to convert back to the 1960s, when our data collection starts, to GNP or GDP. Mainly for this reason, but also because of the fundamental differences in general economic structure, we started adding these countries after 1991, at the cost of having shorter time series.

At the same time, due to its relative nature, the models require a large sample to be accurate. Without a sufficiently large sample, the establishment of a relative model does not yield results that reflect the capability of a country in comparison to others in the international system. A goal of data collection and extension of this dataset should be to expand the sample to cover more countries -all countries, if possible- in the world for the duration of the series. Improvements in transparency, reporting, and information access greatly facilitate this goal. The smaller and the more homogeneous the sample, the less variance, resulting in predicted values that are not that different from actual values. Although we control for these clusters, this is a problem, especially when the availability of data for developed countries is greater than for underdeveloped countries. The need for data for developing countries should not be dismissed, and often it is better to make assumptions and accept lower quality data than to reject those series altogether. For these reasons, linear interpolations, moving averages, or best guesses based on the existing data and other indicators are used in some cases. These data points are reported in detail in the attached country tables.

Also in line with the requirement to have a large sample, the RPC estimations that are reported in the dataset are carried out with the largest sample possible, including all the countries and years with sufficient data quality. However, the provision of components of RPC enables the users to construct alternative RPCs as their specific research interests and questions require. Alternative RPCs can be estimated using a smaller set of countries and additional variables. For example, one can calculate the RPC just for Latin American or OECD countries, or exclude

countries with a population less than 500000. Alternatively, one can include additional control variables such as healthcare or education spending. We consider the flexible nature of RPC a major advantage over other measures of state capacity.

MAJOR DATA SOURCES

Relative Political Performance dataset has been a work in progress for over 30 years. Relative capacity of governments was first used by Organski and Kugler (1980) with the argument that the ability of a government to mobilize the resources of its population represents a critical component in the ability to effectively wage war. Since then, the RPC went through some significant updates. This update relies heavily on these previous versions. Specifically, three previous datasets are used as the foundations for the RPE part of this update. The first one is the BEST dataset, which consists of the original collections of Jacek Kugler, and was used in Organski et al. (1984), *Birth, Deaths, and Taxes*. The second one is the 1995 update to RPC, which is mainly used in Arbetman and Kugler (1997), *Political Capacity and Economic Behavior*. The third one is the most recent update, which is described in more detail in Arbetman and Johnson (2008a). Arbetman and Johnson (2008b) served as the foundation dataset for the update for RPR. In addition to these, the aforementioned general sources are used to update and extend the dataset. Description and the primary sources used for each variable are explained in the next sections.

COMPONENTS AND PRIMARY SOURCES FOR RELATIVE POLITICAL EXTRACTION

Tax Ratio

Whenever possible, The IMF Statistics Department's Government Finance Statistics Manual 2014 (GFSM 2014) definition of general government tax revenues is used for tax revenue data collection. Thus, tax revenues are “compulsory, unrequited amounts receivable by to the general government sector from institutional units. Taxes can be receivable in cash or in kind. By its nature, only a government unit can receive revenue in the form of taxes. When an institutional unit other than a government unit collects taxes, the tax should be attributed in

accordance with tax attribution guidelines. Tax revenue is considered to be unrequited because the government provides nothing directly to the individual unit in exchange for the payment. Governments may use the tax revenue to provide goods or services to other units, either individually or collectively, or to the community as a whole. Certain compulsory receivables, such as fines, penalties, and most social security contributions, are not considered taxes. These types of revenue have, under certain conditions, an element of exchange and are therefore not classified as taxes." (International Monetary Fund, 2014). Alternatively, in older datasets, "Tax Revenue = Total Revenue – Total Social Security Revenue – Total Nontax Revenue" is used to calculate the tax revenues. Tax ratios are calculated using dividing tax revenues by GDP.

The main source used for tax ratio data is IMF's Government Finance Statistics. When available, general government tax data is taken from the latest version of Government Finance Statistics (GFS) Online (July 2020 version). We also benefited from IMF Article IV Staff Reports for years after 1995 based on availability. These reports were especially useful to improve the accuracy of the data by showing its breakdowns, especially for countries with significant mineral production. As alternatives, we used the OECD data from OECD's Revenue Statistics (2020) both for member countries and Latin American, African, and Asian countries. We also benefited from World Development Indicators (World Bank, 2020), the International Center for Tax and Development and the United Nations University World Institute for Development Economics Research Government Revenue Dataset (ICDT/UNU WIDER, 2020).

There are differences in classification of some categories of subcomponents of taxes. The differences of classification between GFS and 2008 System of National Accounts (SNA) can be omitted because the data we are using uses the total taxes. On the other hand, OECD Revenue Statistics treats compulsory social security contribution as taxes, which creates a problem when we combine it with GFS data. To ensure standardization, we deducted social security contributions from taxes whenever possible.

For years where data for the central government was available whereas data for the general government was not, we used the sum of central, state, and local taxes from GFS when possible. When this was not possible, we used the rate of changes of central government data to extend the data back. The previous RPC datasets are also used to extend the data back. Among these, Arbetman and Johnson (2008a) used IMF GFS for years 1972-2005, World Bank for

1960-1972, and overlapping series and national sources for the cases where data was missing. Arbetman and Kugler (1997) used GFS Yearbooks from various years. In line with these, Organski et al (1984) data consists mainly of World Bank and OECD data from 1975-1977. Another report we used was IMF and World Bank country reports or countries' internal national accounts publications. We also utilized World Bank (2020) as a source of last resort incorporating their rates of change to our base lines. When no other data was available, we used the rate of changes of revenue statistics from Banks and Wilson (2020) and taxes on production and imports divided by GDP from United Nations Statistics Division National Accounts Official Country Data (2020a),[3] as they are rough representations of general tax revenues. Linear interpolations are used for gap years where we had data points before and after missing years. Best guesses or 3 year moving averages are used when necessary for extrapolations.

In addition to the general government tax revenues, this version of the RPC dataset distinguishes the total tax revenues with total non-resource tax revenues, which excludes taxes originated in the mining industry. Using the non-resource tax revenues, we calculate the rpe_gdp_nonres, which is theoretically more valid, especially for countries with significant resource revenues, since it captures the capability of governments to directly extract taxes from their populations. This variable is very elastic for many minerals and responds to changes in commodity prices as well as production and discovery of new resources. We mainly used the IMF Article IV Staff Reports, IMF's historical statistical appendices, and ICDT/UNU WIDER (2020) to calculate the difference between total taxes and non-mineral taxes. Since the data before 1990 was unavailable in neither of these resources, we used the rate of changes to extend it back to 1960. We are aware that using the rate of changes to expand the non-mineral taxes is not necessarily a valid way since non-mineral revenues are not highly correlated with mineral revenues for the specific set of countries this variable targets. For this reason, users should be cautious using this variable especially during major commodity crises, such as the oil crisis of 1973. In the future iterations of the RPC dataset, we aim to improve the validity by going through historical data to obtain the actual raw numbers.

[3] This data was available in national currencies, and divided by current GDP in national currency from the same source (United Nations Statistics Division, 2020b).

Agriculture

Share of agriculture in total economy "corresponds to International Standard Industrial Classification (ISIC) divisions 1-5 and includes forestry, hunting, and fishing, as well as cultivation of crops and livestock production. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by the ISIC, revision 3. Note: For VAB countries, gross value added at factor cost is used as the denominator" (World Bank, 2020).

The main source used for agriculture is the UN National Accounts Main Aggregates (United Nations Statistics Division, 2020a). For cases not available in this dataset, we used UN National Accounts Official Country Data (United Nations Statistics Division, 2020d),[4] World Bank World Development Indicators (2020), Arbetman and Kugler (1997), which relied on World Bank World Tables, Arbetman and Johnson (2008a), which used IMF GFS for years 1972-2005, World Bank for 1960-1972, and overlapping series and national sources for the cases where data was missing, and Organski et al (1984), which mainly relied on World Bank data from 1975-1977. These alternative sources are often used for rate of changes to extend the data back. Linear interpolations are used for gap years where we had data points before and after missing years. Best guesses or 3 year moving averages are used when necessary for extrapolations. Before doing any interpolation or extrapolations, we checked with the national sources to ensure there were not any major droughts, other agriculture-related crises, or even changes of main crops.

Mining

Share of mining in total economy corresponds to "the extraction of minerals occurring naturally as solids (coal and ores), liquids (petroleum) or gases (natural gas). Extraction can be achieved by different methods such as underground or surface mining, well operation, seabed mining etc." as defined by ISIC Rev. 4 Code B (United Nations Statistics Division, 2020e).

[4] This data was available in national currencies, and divided by current GDP in national currency from the same source (United Nations Statistics Division, 2020b).

The main source used for mining is UN National Accounts Official Country Data (United Nations Statistics Division, 2020d) [5]. In this version of RPC, we updated the data from 2005 using the last version of UN National Accounts Official Country Data (United Nations Statistics Division, 2020b) to make sure the adjustments in the data for the last years are taken into account. For cases not available in this dataset, World Bank World Development Indicators (2020) (coal, mineral, natural gas, and oil rents / gdp), Arbetman and Kugler (1997), which relied on World Bank World Tables, Arbetman and Johnson (2008a), which used IMF GFS for years 1972-2005, World Bank for 1960-1972, and overlapping series and national sources for the cases where data was missing, Organski et al (1984), which mainly relied on World Bank data from 1975-1977, and (mining, manufacturing and utilities – manufacturing) calculation from UN National Accounts Main Aggregates (United Nations Statistics Division, 2020c) . These alternative sources are often used for rate of changes to extend the data back. Linear interpolations are used for gap years where we had data points before and after missing years. Best guesses or 3 year moving averages are used when necessary for extrapolations.

Exports

Share of exports of goods and services in total economy "represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments." (World Bank, 2020).

The main source used for exports is UN National Accounts Main Aggregates (United Nations Statistics Division, 2020a). For cases not available in this dataset, we used World Bank World Development Indicators (2020), Arbetman and Kugler (1997), which relied on World Bank World Tables, Arbetman and Johnson (2008a), which used IMF GFS for years 1972-2005, World Bank for 1960-1972, and overlapping series and national sources for the cases where data was missing, and Organski et al (1984), which mainly relied on World Bank data from 1975-

[5] This data was available in national currencies, and divided by current GDP in national currency from the same source (United Nations Statistics Division, 2020b).

1977. These alternative sources are often used for rate of changes to extend the data back. Linear interpolations are used for gap years where we had data points before and after missing years. Best guesses or 3 year moving averages are used when necessary for extrapolations.

GDP Per Capita

"GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources." (World Bank, 2020).

The GDP per capita is measured in constant 2010 dollars. The base series used is from the World Bank (2020). We also provided the purchasing power parity (PPP) converted GDP per capita in 2011 constant international dollars, taken from Penn World Tables 9.0 (Feenstra, Inklaar, and Timmer, 2015) despite we did not directly use the PPP measure in our calculations. Theoretically, GDP per capita at PPP can also be used to calculate RPC. In order to complete the series for years in which the data was not available, overlapping series including the World Development Indicators (2020), Maddison (2008) and national sources, as well as rate of changes of constant and PPP values used for extrapolations.

COMPONENTS AND PRIMARY SOURCES FOR RELATIVE POLITICAL REACH

Economically Active Population

"The economically active population comprises all persons of either sex who furnish the supply of labour for the production of economic goods and services as defined by the United Nations Systems of National Accounts and balances during a specified time-reference period. According to these systems the production of economic goods and services includes all production and processing of primary products whether for the market for barter or for own consumption, the production of all other goods and services for the market and, in the case of households which produce such goods and services for the market, the corresponding production for own consumption" (International Labour Organization, 1982).

We use economically active population / total population for the calculation of the ratios. Economically Active Population data primarily comes from Arbetman and Johnson (2008b), which relies on Taylor and Jodice (1982) and International Labour Organization, specifically the ILO modeled estimates, with overlapping series from World Development Indicators, United Nations Demographic Yearbooks (particularly UN Population Projections 2019), National Census, and Banks and Wilson (2020). Updates were implemented using rate-of change extrapolations from International Labour Organisation (2011), United Nations Food and Agriculture Organization (2020), and World Bank (2020) for missing data years. Best guesses or 3 year moving averages are used when necessary for extrapolations.

Unemployment

Unemployment basically refers to "share of the labor force that is without work but available for and seeking employment" (World Bank, 2020). Although the definition can change from country to country, we tried to implement the general international definition (International Labour Organization, 2020). Unemployment data is appropriated from International Labour Organization (1982), specifically the ILO modeled estimates, and using rate-of-change extrapolations from Central Intelligence Agency (2001-2020), International Monetary Fund (2020), and World Bank (2020). Linear interpolations are used for gap years where we had data point before and after missing years. Best guess or 3 year moving averages are used when necessary for extrapolations.

Activity Rate

We calculated activity rate ("work" in the dataset) by subtracting unemployment rate from economically active population rate, for years which both of these are available. When at least one of these was unavailable, we considered activity rate as missing. Ideally we should have subtracted the ratio of part time workers to total population, but since this data was unavailable for most of the cases, and is often very small, we did not include it in our calculations at this time.

Young Population

Young population refers to "population between the ages 0 to 14 divided by the total population. Population is based on the de facto definition of population" (World Bank, 2020).

Young population is taken from World Bank (2020). Linear interpolations are used for gap years where we had data point before and after missing years. Best guess or 3 year moving averages are used when necessary for extrapolations.

Social Security

Social Security Contributions "are actual revenue receivable by social security schemes organized and operated by government units, for the benefit of the contributors to the scheme. These contributions are classified by the source of the contribution, which may be the employers or the household sector (separated according to whether they are employees, self-employed, or unemployed). Employee contributions are either payable directly by employees or deducted from employees' wages and salaries and transferred on their behalf by the employer. Employer contributions are payable directly by employers on behalf of their employees. Self-employed or unemployed contributions are paid by contributors who are not employees. Unallocable contributions are those contributions whose source cannot be determined. Amounts payable by general government employers are not eliminated by consolidation when the paying and receiving units are in the same sector or subsector" (International Monetary Fund, 2014).

The data for social security contributions primarily comes from ICDT/UNU WIDER (2020). The missing years are completed using Arbetman And Johnson (2008a,b) which uses on International Monetary Fund GFS for years 1972-2005, World Bank for 1960-1972, and overlapping series and national sources for the cases where data was missing, the most recent GFS available (International Monetary Fund, 2020) and IFS (International Finance Statistics) (International Monetary Fund, 2020). Linear interpolations are used for gap years where we had data points before and after missing years. Best guesses or 3 year moving averages are used when necessary for extrapolations.

Education

The education measure we used is the gross secondary school enrollment ratio for all programs per thousand people as defined by "Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown. Secondary education completes the provision of basic education that began at the primary level, and aims at laying the foundations for lifelong learning and human development, by offering more subject- or skill-oriented instruction using more specialized teachers." (World Bank, 2020).[6]

Our primary source for gross secondary school enrollment ratio is UNESCO Institute for Statistics (2020b). The missing years are completed using Arbetman and Johnson (2008b), which collected their data mainly from Taylor and Jodice (1982), and supported this with data from the World Development Indicators and National Sources to supplement data where necessary. Linear interpolations are used for gap years where we had data points before and after missing years. Best guesses or 3 year moving averages are used when necessary for extrapolations.

Urban Population

Urban population refers to "Urban population refers to people living in urban areas as defined by national statistical offices. The data are collected and smoothed by United Nations Population Division." (World Bank 2020).[7]

Urban population is primarily taken from World Bank World Development Indicators (2020). Best guess or 3 year moving averages are used when necessary for extrapolations.

Population

Population is "based on the de facto definition of population, which counts all residents regardless of legal status or citizenship. The values shown are midyear estimates." (World Bank, 2020).

[6] A similar definition can be found in UNESCO Institute for Statistics (2020a).
[7] To see details about different definitions of statistical offices, please see https://blogs.worldbank.org/sustainablecities/how-do-we-define-cities-towns-and-rural-areas

Population data comes from World Bank (2020). United Nations Population Projections (2019) is used for countries not included in World Bank (2020). Linear interpolations were used for missing data years. Best guess or 3 year moving averages are used when necessary for extrapolations.

GDP Per Capita

"GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources." (World Bank, 2020).

The GDP per capita is measured in constant 2010 dollars. The base series used is from the World Bank (2020). We also provided the purchasing power parity (PPP) converted GDP per capita in 2011 constant international dollars, taken from Penn World Tables 9.0 (Feenstra, Inklaar, Timmer, 2015) despite we did not directly use the PPP measure in our calculations. Theoretically, GDP per capita at PPP can also be used to calculate RPC. In order to complete the series for years in which the data was not available, overlapping series including the World Development Indicators (2020), Maddison (2008) and national sources, as well as rate of changes of constant and PPP values used for extrapolations.

Bureaucracy

We measure the size of bureaucracy by looking at the share of government wages in total general government expenses. "Wages and salaries are compensation of employees payable in cash and/or in kind, except for social contributions payable by employers. Wages and salaries exclude amounts connected with own-account capital formation. They include amounts withheld from wages and salaries by the employer for administrative convenience or other reasons, such as social contributions, income taxes, and other deductibles, payable by the employee. These deductibles are oft en paid directly to social insurance schemes, tax authorities, etc., on behalf of the employee. Wages and salaries may be payable in various ways, including goods or services

provided to employees as remuneration in kind instead of, or in addition to, remuneration in cash. Including remuneration in kind allows GFS to measure the full cost of labor employed." (International Monetary Fund, 2014).

Bureaucracy and expenditure data is taken from GFS (International Monetary Fund, 2020). The missing years are from Arbetman and Johnson (2008b), whose data come primarily from the Government Finance Statistics published by the International Monetary Fund. We also use data from the World Development Indicators, Europa Yearbooks, and National Sources. When the exact wages divided by total general government expenses could not be found, we relied on the rate of changes of wages divided by central government expenses, or wages divided by GDP for extrapolations. Linear interpolations are used for gap years where we had data points before and after missing years. Best guesses or 3 year moving averages are used when necessary for extrapolations.

COMPONENTS AND PRIMARY SOURCES FOR RELATIVE POLITICAL ALLOCATION:

Dependent Variable:

(1) Gross Domestic Product (GDP) per capita at 2010 prices –US dollars
Source: World Development Indicators by World Bank

Total Factor Productivity Control Variables:
(2) School enrollment, secondary (% gross)
Source: World Development Indicator by World Bank
(3) Economic openness. (Export+Import)/GDP
Source: National Accounts Estimates of Main Aggregates by United Nations Statistics Division (UNSD)
(4) OECD Membership as of 1990. 1 if the country is an OECD member, 0 otherwise
Source: OECD

Physical Capital Input Independent Variable:

(5) Gross Fixed Capital Formation per capita at 2010 prices – US dollars

Source: National Accounts Estimates of Main Aggregates by United Nations Statistics Division (UNSD)

The Public Expenditure Shares Independent Variables:

(6) General government expenditure on Traditional Public Goods per capita at 2010 prices – US dollars

(7) General government expenditure on Economic Services per capita at 2010 prices – US dollars

(8) General government expenditure on Education Resources per capita at 2010 prices – US dollars

(9) General government expenditure on Social Protection per capita at 2010 prices – US dollars

Source:

UNSD: United Nations Statistics Division. Global Indicator Database-National Accounts and Industrial Production. General Government Final Consumption Expenditure by Function.

IMF:

(a) Before 1995: International Monetary Fund. Government Finance Statistics. ICPSR08624-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-07-26. doi:10.3886/ICPSR08624.v2

(b) After 1995: International Monetary Fund. Government Finance Statistics. June 2020.

Bibliography

Arbetman, Marina. 1990. ‘‘The Political Economy of Exchange Rate Fluctuations.’’ PhD diss., Vanderbilt University.

Arbetman, Marina. 1994. *The Concept of Political Penetration*. In Proceedings of the Conference on Political Capacity and Economic Behavior, organized by the Center for Politics and Economics. Claremont: Claremont Graduate School.

Arbetman, Marina, Jacek Kugler, and A. F. K. Organski. 1997. ‘‘Political Capacity and Demographic Change.’’ Marina Arbetman and Jacek Kugler, eds., *Political Capacity and Economic Behavior*. Boulder, CO: Westview Press.

Arbetman, Marina & Kristin Johnson. 2008a. Relative political capacity: Empirical and theoretical underpinnings. Paper presented at the 49th Annual Convention of the International Studies Association, San Francisco, CA, 26–29 March.

Arbetman, Marina, and Kristin Johnson. 2008b. ‘‘Power Distribution and Oil in the Sudan: Will the Comprehensive Peace Agreement Turn the Oil Curse into a Blessing?’’ *International Interactions* 34 (4): 382–401.

Banks, Arthur S., Wilson, Kenneth A. 2020. *Cross-National Time-Series Data Archive*. Databanks International. Jerusalem, Israel

Battese, G.E., and T.J. Coelli. 1995. “A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data.” *Empirical Economics*, 20 (2).

Central Intelligence Agency. The World Factbook. Washington, DC: Central Intelligence Agency, Various Years.

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.

ICTD/UNU-WIDER. 2020. *Government Revenue Dataset, 2020*, https://www.wider.unu.edu/project/government-revenue-dataset

International Labour Assocation. 1982. ILO Resolutions Concerning Economically Active Population, Employment, Unemployment and Underemployment Adopted by the 13th International Conference of Labour Statisticians, October 1982, para. 5. Available from http://www.ilo.ch/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_087481.pdf

International Labour Association. 2020. ILOSTAT SDG Labour Market Indicators. Available from https://www.ilo.org/shinyapps/bulkexplorer7/?lang=en&segment=indicator&id=EAP_DWAP_SEX_AGE_RT_A

International Monetary Fund. 2009-2020. *Government Finance Statistics* (Online and CD-Rom). Washington, DC: International Monetary Fund.

International Monetary Fund. 2014. *Government Finance Statistics Manual 2014*. Washington, DC: International Monetary Fund.

International Monetary Fund. Government Financial Statistics Yearbooks, Various Years. Washington, DC: International Monetary Fund.

International Monetary Fund. Official Country and Staff Reports, Various Years. Washington, DC: International Monetary Fund. (Specific citations available in each country report).

Kang, K. 2015. "Measuring the Allocative Political Performance of Nations." Paper presented at the Annual Midwest Political Science Association (MPSA) meeting, Chicago, IL.

Maddison, Angus. 2008. Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD. Available from http://www.ggdc.net/MADDISON/oriindex.htm

Oxley, H., Maher M., Martin, J.P., Nicoletti, G, and Alonso-Gamo, P. 1990. "The Public Sector: Issues for the 1990s." *OECD Working Paper*, No. 90, OECD Department of Economics and Statistics, Paris: OECD.

Organization of Economic Cooperation and Development. 2020. Revenue Statistics – OECD Member Countries: Comparative Tables, Latin American Countries, African Countries, Asian Countries, OECD Tax Statistics (database). Available from http://stats.oecd.org/Index.aspx?QueryId=21699

Organski, A. F. K., and Jacek Kugler. 1980. *The War Ledger*. Chicago: University of Chicago Press.

Organski, A. F. K., Jacek Kugler, Timothy Johnson, and Youssef Cohen. 1984. *Births, Deaths and Taxes*. Chicago: University of Chicago Press.

Saunders, P. and Klau, L. 1985. "The Role of the Public Sector: Causes and Consequences of the Growth of Government." *OECD Economic Studies*, 4, 11-239.

Taylor Charles Levis Taylor and David A. Jodice. 1982. World Handbook of Political and Social Indicators III: 1948-1982. ICPSR, Study Number: 7761.

UNESCO Institute for Statistics. 2020a. Glossary. Available from http://glossary.uis.unesco.org/glossary/en/home

UNESCO Institute for Statistics. 2020b. Data Centre Custom Tables, Gross Enrollment Ratio, Secondary, All Programmes. Available from http://data.uis.unesco.org/index.aspx?queryid=142 .

United Nations Demographic Yearbook. New York: United Nations; Various Years.

United Nations. 1993 & 2008. *System of National Accounts*. Washington, DC: United Nations.

United Nations Department of Economic and Social Affairs, Population Division. 2019. World Population Prospects: The 2019 Revision. Available from https://population.un.org/wpp/.

United Nations Statistics Division. 2020a. National Accounts Official Country Data. Available from http://data.un.org/Data.aspx?d=SNA&f=group_code:401;item_code:10&c=2,3,4,6,7,8,9,10,11,12,13&s=_cr_engNameOrderBy:asc,fiscal_year:desc,_grIt_code:asc&v=1

United Nations Statistics Division. 2020b. National Accounts Official Country Data. Available from http://data.un.org/Data.aspx?d=SNA&f=group_code:401;item_code:8&c=2,3,4,6,7,8,9,10,11,12,13&s=_cr_engNameOrderBy:asc,fiscal_year:desc,_grIt_code:asc&v=1

United Nations Statistics Division. 2020c. *National Accounts Main Aggregates Database, Percentage Distribution of GDP*. Available from http://unstats.un.org/unsd/snaama/dnlList.asp

United Nations Statistics Division. 2020d. National Accounts Statistical Aggregates Database, Official Country Data. Available from http://data.un.org/Data.aspx?d=SNA&f=group_code%3a201%3bitem_code%3a1%2c16%2c19%2c2%2c22&c=2,3,4,6,7,8,9,10,11,12,13&s=_cr_engNameOrderBy:asc,fiscal_year:desc,_grIt_code:asc&v=1

United Nations Statistics Division. 2020e. Detailed Structure and Explanatory Notes, ISIC Rev.4. Available from https://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27

United Nations Food and Agriculture Organization. 2020. Total Economically Active Population Data. Available from: http://www.fao.org/faostat/en/#data/OA

World Bank. 2007-2020. *The World Development Indicators*. Washington, DC: World Bank.