Codebook for joint WP5 and WP11 dataset at regional level

The following presents a description of the data and sample for all regional level data collected and used in the WP5 and WP11 ANTICORRP project to date. The primary source paper for each variable is included by the variable name where applicable. The dataset is included in Excel format and can be easily transferred to a statistical program such as STATA, SPSS or R.

Sample and Units of analysis

The units of analysis are NUTS[1] regions in 22 countries in Europe, including all EU28 countries with at least 2 NUTS regions as well as Serbia and Turkey.

Countries with regional data at the <u>NUTS 1</u> level here are:

-Belgium, Germany, Greece, Hungary, Sweden, Turkey and United Kingdom

Countries with regional data at the <u>NUTS 2</u> level are:

-Austria, Bulgaria, Czech Republic, Denmark, Finland, France, , Ireland, Italy, Netherlands, Portugal, Poland, Romania, Slovak Republic, Spain and Serbia,

Variables

NUTS region – the NUTS code for each region in the dataset. Codes with one number following the two letters indicate that the region is at the NUTS 1 level, while those with two numbers indicate the NUTS 2 level.

reg_name – lists the name of each region in the dataset.

---

[1] NUTS stands for ' Nomenclature of territorial units for statistics' and is used to identify countries and regions for statistical purposes. For more information, see:
http://ec.europa.eu/eurostat/web/nuts/overview

NUTS_c – NUTS code of the country for each region in the dataset.

c_num – the code number for each contry according to the 2013 EQI survey (Charron, Dijkstra and Lapuente 2015)

capital – a dummy variable that equals '1' if the region is a capital region and '0' if otherwise.

polrev – a variable that equals '1' if the NUTS region in question has at policy relevance, which is defined as having an elected parliament at the NUTS level in question and some political, fiscal and/or administrative control over education, health care or law enforcement policy areas.

eqi2013 – Each region's EQI ('European Quality of Government') estimate for the 2013 round of the data. The EQI is an index of 16 'quality of government' (QoG) related question about the extent to which regional public sector services – education, health care and law enforcement – are allocated with quality, impartiality and without corruption. The data is standardized so that the sample mean is '0' with a standard deviation of '1'. -*Source* (and for more information): Charron, Dijkstra and Lapuente 2015

eqi2010 - Each region's EQI ('European Quality of Government') estimate for the 2010 round of the data.

Source (and for more information): Charron, Dijkstra and Lapuente 2014

lcri_euc1_r – a measure of corruption risk based on EU procurement data averaged from 2009-2013. The variable captures high-level corruption risk at the regional level via measuring deliberate restriction of open competition for government contracts in order to benefit a well-connected company. This measure is the percentage of single-bidder contracts awarded in all the awarded in all EU procurement contracts by region from 2009-2012.

-Source (and for more information): Fazekas, Tóth and King 2013

lcri_neu_r: a measure of corruption risk based on EU procurement data averaged from 2009-2013. The variable captures high-level corruption risk at the regional level via measuring deliberate restriction of open competition for government contracts in order to benefit a well-connected company. This measure is an index of several indicators of corruption risk that include:

i) a type of restricted, non-open tendering procedure; ii) the use of subjective, non-price related assessment criteria; iii) a very short advertisement period; and iv) a quick evaluation of bids and v) the percentage of single-bidder contracts awarded in all the awarded in all EU procurement contracts, by region from 2009-2012.

-Source (and for more information): Fazekas, Tóth and King 2013

Pub_merExp: a measure intended to capture the extent to which success in the public sector in each region is based merit vs. connections and luck based on the perceptions of public sector workers themselves. The measure is taken from the following survey question in the latest EQI survey:

"Which statement comes closer to your own views? 1 means you agree completely with the statement on the left; 10 means you agree completely with the statement on the right; and if your views fall somewhere in between, you can choose any number in between:

1 (In the public sector most people can succeed if they are willing to work hard)

10 (Hard work is no guarantee of success in the public sector for most people— it's more a matter of luck and connections)"

The question is then aggregated by region using only a sample of public sector workers.

Source (and for more information): Charron, Dahlstrom and Lapuente 2015

PubMe_se – the standard error of the variable **Pub_merExp**. The standard error is calculated as the standard deviation 's' over the square root of the number of observations by region 'n', or: $s/\sqrt{n}$.

-Source (and for more information): Charron, Dahlstrom and Lapuente 2015

Pub_merPer - a measure intended to capture the extent to which success in the public sector in each region is based merit vs. connections and luck based on the perceptions non-public sector workers, making the measure a more perceptions-based measure than Pub_merExp. The question and method used to calculate the estimate by region is the same as Pub_merExp, however, only non-public sector workers are used in aggregating each region's score.

-Source (and for more information): Charron, Dahlstrom and Lapuente 2015

PubMp_se – the standard error of the public sector merit measure based on the sample of non-public sector workers (**Pub_merPer**).

TRUST – a measure of generalized trust by region based on the following survey question from the 2013 EQI survey data:

"Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people in your area?"

the response "Most people can be trusted" is coded as 1 and "Can't be too careful" is coded as 0 and the individual level responses are aggregated to each region so that the regional measure captures the proportion of '1' responses.

-Source (and for more information): Charron and Rothstein 2014

pop12 – the population of each region from 2012 Eurostat data.

Popwgt_s – the proportion of the population that the region in question contributes to the sample based on latest population statistics from Eurostat. Can be used for weighting in statistical analysis.

Popwgt_c – the proportion of the population that the region in question contributes to each country based on latest population statistics from Eurostat. Can be used for weighting in statistical analyses or to aggregate regional data to the national level.

Sources:

Charron, N., Dijkstra, L., & Lapuente, V. (2014). Regional governance matters: quality of government within European Union member states. *Regional Studies*, *48*(1), 68-90.

Charron, N., Dijkstra, L., & Lapuente, V. (2015) Mapping the Regional Divide in Europe: A Measure for Assessing Quality of Government in 206 European Regions. *Social Indicators Research*, *122* (2), 315-346. Doi: 10.1007/s11205-014-0702-y

Charron, N., C. Dahlstrom and V. Lapuente. 2015. Measuring Meritocracy in the Public Sector in Europe: a New National and Sub-National Indicator. QoG Working Paper Series 2015: 8.

Charron, Nicholas & Bo Rothstein: 2014. Social Trust, Quality of Government and Ethnic Diversity: An Empirical Analysis of 206 Regions in Europe. QoG Working Paper Series 2014: 20.

Fazekas, M., Tóth, I. J., & King, L. P. (2013). Anatomy of grand corruption: A composite corruption risk index based on objective data. *Corruption Research Center Budapest Working Papers No. CRCB-WP/2013, 2.*