CODEBOOK
Electoral systems – Design, context, and performance on the macro level
Johannes Raabe, University of Kiel

Contents

1  Case selection  2
2  Incomplete data and coding rules  2
3  Variables  3
  3.1  General (ID) variables  3
  3.2  Electoral system variables  4
  3.3  Context variables  8
  3.4  Performance variables  12
References  14
1 Case selection

The focus of this dataset is on mixed electoral systems – every effort was taken to include as many mixed-member electoral systems and PR systems based on multimember districts as possible. Plurality/Majority and pure PR systems were added as benchmark cases. Relying on existing data sources such as Dawn Brancati’s Global Elections Database (Brancati, 2015), Adam Carr’s Election Archive (Carr, 2015), and different volumes (co-)edited by Dieter Nohlen (Nohlen & Stöver, 2010; Nohlen, 1999, 2005) as well as official election statistics from the respective national institutions (typically the Electoral Commission or the Ministry of the Interior) we compiled complete election results for a large number of competitive elections. The dataset consists of 590 elections in 56 countries after 1945 – all elections that were conducted in a country that at least experienced a period of democracy and for which there was detailed and reliable enough data were included into the dataset.

The dataset contains election results from the following countries: Albania, Argentina, Australia, Austria, Belgium, Bolivia, Brazil, Bulgaria, Canada, Costa Rica, Croatia, Cyprus, Czech Republic, Denmark, El Salvador, Estonia, Finland, France, Georgia, Germany, Greece, Guatemala, Hungary, Iceland, Ireland, Israel, Italy, Japan, Latvia, Lesotho, Lithuania, Luxembourg, Macedonia, Malta, Mexico, Mongolia, Nepal, Netherlands, New Zealand, Norway, Paraguay, Poland, Portugal, Romania, Senegal, Slovakia, Slovenia, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Turkey, Ukraine, United Kingdom, Venezuela.

2 Incomplete data and coding rules

Every effort was made not to include an ‘others’ category into the dataset, however, the source material sometimes did not permit this for the vote data (seat data are fully disaggregated down to parties and individual/independent candidates for all elections in the dataset). The problem that most often occurs is that it is impossible to derive the vote numbers for individual independent candidates. In these cases, vote data are missing for these candidates. What has been done is that those elections where independent candidates played a significant role (won more than 5% of the seats in parliament) and/or where all parties/candidates not winning any seats have been summed up in an ‘other’ category have been marked by a dummy-variable (‘disaggr_problem_votes’). This dummy-variable may then later be used in order to identify these elections and either exclude the respective vote-based indices from the analysis or find a proper solution for these elections. The most prominent solution to this problem is Taagepera’s (1997) least components correction.

Overall these problems occur for 30 elections and thus roughly five percent of all elections in the dataset. Below is a list of the elections for which the vote data could not be disaggregated completely. It is important to note here that this problem is typically only of minor relevance and has seldom been found to influence substantial results (see Taagepera 1997; Gallagher & Mitchell 2005; Bormann & Golder 2013)

these are not independents, but very small minority parties which are guaranteed a seat by law; South Korea (1992, 1996, 2008); Ukraine (1998, 2002, 2012)


For mixed systems, in some cases the vote data for the majoritarian tier are missing or incomplete (this is the case for Bolivia 2002). If they are incomplete, the respective performance indices are coded as missing. The performance indices are based on the sum of all party/candidate seats and votes in the dataset, not on what is reported as the total number of valid votes (following Lijphart 1994). If voters were only able to vote for a coalition (not a specific party of that coalition), the coalition enters the dataset as a single-party. If voters vote for individual parties (who may be part of an umbrella coalition), these parties enter the dataset individually (as is the case, e.g., for Italy).

3 Variables

3.1 General (ID) variables

country
The name of the country.

year
The year of the election.

election_in_this_year
This variable signals whether an election was the first in the respective year (1), the second (2) etc.

region ('region2' from Bormann & Golder 2013): This is a categorical variable indicating the country's region of the world.
1. Sub-Saharan Africa
2. South Asia
3. East Asia
4. South East Asia
5. Pacific Islands/Oceania
6. Middle East/North Africa
7. Latin America
8. Caribbean and non-Iberic America
9. Eastern Europe/post-Soviet states
10. Western Europe

1 Where variables were taken as-is from other datasets, the respective descriptions of the original authors are also given here.
3.2 Electoral system variables

**legislative_type** (from Bormann & Golder 2013): This is a categorical variable that takes on one of three values indicating the basic type of electoral system used in the elections.
1. Majoritarian
2. Proportional
3. Mixed

- Updated through 2014 based on publicly available information on the respective electoral systems (IPU, Government sources).

**elecrule** (from Bormann & Golder 2013): This is a categorical variable that provides a more detailed indication of the type of electoral system used in the election.
1. Single-Member-District-Plurality (SMDP)
2. Two-Round System (TRS)
3. Alternative Vote (AV)
4. Borda Count (BC)
5. Block Vote (BV)
6. Party Block Vote (PBV)
7. Limited Vote (LV)
8. Single Nontransferable Vote (SNTV)
9. List Proportional Representation (List PR)
10. Single Transferable Vote (STV)
11. Mixed Dependent (or Mixed Member Proportional)
12. Mixed Independent (or Mixed Parallel).

- Updated through 2014 based on publicly available information on the respective electoral systems (IPU, Government sources).

**formula** (from Bormann & Golder 2013): This is a categorical variable that indicates the precise electoral formula used in an electoral tier.
1. Single-Member-District-Plurality (SMDP)
2. Two Round Majority-Plurality
3. Two Round Qualified Majority
4. Two Round Majority Runoff
5. Alternative Vote (AV)
6. Borda Count (BC)
7. Modified Borda Count (mBC)
8. Block Vote (BV)
9. Party Block Vote (PBV)
10. Limited Vote (LV)
11. Single Nontransferable Vote (SNTV)
12. Hare quota
13. Hare quota with largest remainders
14. Hare quota with highest average remainders
15. Hagenbach-Bischoff quota
16. Hagenbach-Bischoff quota with largest remainders
17. Hagenbach-Bischoff quota with highest average remainders
18. Droop quota
19. Droop quota with largest remainders
20. Droop quota with highest average remainders
21. Imperiali quota
22. Imperiali quota with largest remainders
23. Imperiali quota with highest average remainders
24. Reinforced Imperiali quota
25. D'Hondt
26. Sainte-Laguë
27. Modified Sainte-Laguë
28. Single Transferable Vote

- Updated through 2014 based on publicly available information on the respective electoral systems (IPU, Government sources).

**mixed_type** (from Bormann & Golder 2013): This is a categorical variable that indicates the precise type of mixed electoral system that is being used (Massicotte & Blais, 1999).
1. Coexistence
2. Superposition
3. Fusion
4. Correction
5. Conditional

- Updated through 2014 based on publicly available information on the respective electoral systems (IPU, Government sources).
- Mexico: Mexico was recoded to a conditional type mixed-member system since 1988.
- Venezuela 2010: Recoded to a mixed-member majoritarian electoral system (see Hidalgo 2011).

**no_ballots**
The number of ballots equals ‘1’ for all non-mixed-member electoral systems by default as this variable is not concerned with the option of preferential voting (see electoral formula variable).

**parl_size**
This variable denotes the overall total of seats in the parliament (allocated in the election).
**Maj_seats**
The number of seats contested in the majoritarian tier of an electoral system. For majoritarian electoral systems this number will equal the total number of seats (parliament size). Missing values are assigned for pure PR systems.

**PR_seats**
The number of seats contested in the PR tier of an electoral system. For pure PR electoral systems this number will equal the total number of seats (parliament size). Missing values are assigned for purely majoritarian systems. In the case of interdependent mixed-member systems (MMP variations), the sum of Maj_seats and PR_seats will be greater than the total number of seats due to the special allocation process in these systems where a subset of the seats are allocated in both electoral tiers.

**PR_threshold**
This variable denotes the height of the legal threshold for PR seat allocation. It is coded as missing/not applicable for majoritarian electoral system. If the legal threshold only applies in each district individually, this variable is coded as zero (e.g. Germany 1949 and all years for Spain). Most data come from Beck et al. (2001; updated through 2012). Additional data for Eastern Europe come from Shvetsova (1999).

- Sweden is a difficult case as a 4 percent threshold applies to all valid votes cast but there is an additional 12 percent threshold in each district. If the latter is passed while the former is not, a party is still able to win seats in the respective district. The PR threshold variable was coded as 4 since the district threshold is fairly high and unlikely to foster regionalized parties otherwise blocked by the national threshold.

**percent_SMDs**
This variable denotes the share of single-member districts for an election. In the case of SNTV or Block Vote systems, this variable is coded as the percentage of seats allocated under these plurality formulas (Japan before 1996; Mongolia 2004-2008; Thailand 1995-1996). This variable is insensitive to whether the electoral system works under a compensation mechanism or not (e.g., for Germany – using an MMP system – this variable takes on a value of .5 as there are 598 parliamentary seats and 299 single-member districts; the same is true for Lithuania, which uses an MMM system).

**PR_dm_mean**
This variable denotes the average district magnitude for districts in which seats are allocated based on PR rules. It is coded as missing/not applicable for majoritarian electoral system. For cases with more than two electoral tiers (e.g. Austria, Hungary, and Italy) the mean district magnitude is calculated for the PR tier on the lowest level. Most data for non-mixed cases come from Bormann & Golder (2013).
PR_topup
This is a dummy variable signalling whether there are additional PR top-up seats in a national tier of what is already a PR or mixed-member electoral system (e.g. Austria, Hungary). This variable is also coded as 1 if the top-up seats consist of remainder seats from the district allocation (e.g. Belgium before 1999). Most data come from Bormann & Golder (2013).

- Germany 2013: Since any disproportional effect of ‘Überhangmandate’ (overhang or surplus seats) will be corrected by additional PR (levelling) seats, this case is treated as having a top-up tier of additional PR seats, even if this tier will not necessarily come into force (it did in 2013).

Majority_bonus
This dummy variable indicates whether the party winning most seats in an initial allocation will receive a seat bonus aiming at creating a single-party majority in parliament.

- Italy: The majority bonus is not fixed but will close the gap between the number of seats won by the strongest party and 340 (which constitutes a 55 percent majority of parliamentary seats).
- Mexico: Between 1988 and 1997 (last applied for the 1994 elections), Mexico also employed a flexible majority bonus system (for the details see Molinar Horcasitas & Weldon 2001).
- Malta has a provision (since 1981) that ensures a party the majority of seats if it holds the majority of votes, this is, however, not a bonus arrangement and Malta is thus coded as having no majority bonus.

3.3 Context variables
regime (from Bormann & Golder 2013): This is a categorical variable indicating a country's regime type at the end of the given year.
1. Parliamentary democracy (0)
2. Semi-presidential democracy (1)
3. Presidential democracy (2)
4. Civilian dictatorship (3)
5. Military dictatorship (4)
6. Royal dictatorship (5)

- Switzerland: Recoded as parliamentary democracy (not presidential) since the Swiss president functions exactly as a prime minister in a parliamentary system (see Beck et al. 2001; updated through 2012).
**democ** (from Marshall et al. 2014)

Institutionalized Democracy: Democracy is conceived as three essential, interdependent elements. One is the presence of institutions and procedures through which citizens can express effective preferences about alternative policies and leaders. Second is the existence of institutionalized constraints on the exercise of power by the executive. Third is the guarantee of civil liberties to all citizens in their daily lives and in acts of political participation. Other aspects of plural democracy, such as the rule of law, systems of checks and balances, freedom of the press, and so on are means to, or specific manifestations of, these general principles. We do not include coded data on civil liberties. The Democracy indicator is an additive eleven-point scale (0-10).

This "institutional democracy" indicator follows a logic similar to that underlying the Polity I analyses. There is no "necessary condition" for characterizing a political system as democratic, rather democracy is treated as a variable. For example, the scale discriminates among Western parliamentary and presidential systems based on the extent of constraints on the chief executive. Charles de Gaulle as president of the French Fifth Republic operated within slight to moderate political limitations. Thus the early years of the Fifth Republic have lower Democracy scores than the United States or the Federal Republic of Germany, where constraints on the executive approach parity. Similarly, the onset of "cohabitation" in France during the second phase of the first Mitterrand presidency is marked by a shift toward parity on the Executive Constraints scale and a concomitant increase in France's Democracy score.

If the composite indicator of institutionalized democracy is inappropriate for some conceptual purposes, it can be easily redefined either by altering the constituent categories and weights, or by specifying some minimum preconditions. A mature and internally coherent democracy, for example, might be operationally defined as one in which (a) political participation is unrestricted, open, and fully competitive; (b) executive recruitment is elective, and (c) constraints on the chief executive are substantial.

- Iceland: Not part of the Polity IV dataset. Coded as perfect democracy for every year in the dataset.
- Malta: Not part of the Polity IV dataset. Coded as missing for every year in the dataset.

**autoc** (from Marshall et al. 2014)

Institutionalized Autocracy: "Authoritarian regime" in Western political discourse is a pejorative term for some very diverse kinds of political systems whose common properties are a lack of regularized political competition and concern for political freedoms. We use the more neutral term Autocracy and define it operationally in terms of the presence of a distinctive set of political characteristics. In mature form, autocracies sharply restrict or suppress competitive political participation. Their chief executives are chosen in a regularized process of selection within the political elite, and once in office they exercise power with few institutional constraints. Most modern autocracies also exercise a high degree of directiveness over social and economic activity, but we regard this as a function of political ideology and choice, not a
defining property of autocracy. Social democracies also exercise relatively high degrees of directiveness. We prefer to leave open for empirical investigation the question of how Autocracy, Democracy, and Directiveness (performance) have covaried over time. An eleven-point Autocracy scale is constructed additively.

The logic of this "institutionalized autocracy" scale is similar to that of the institutionalized democracy scale, below, and it is subject to the same kinds of operational redefinition to suit different theoretical purposes. Note that the two scales do not share any categories in common. Nonetheless many polities have mixed authority traits, and thus can have middling scores on both Autocracy and Democracy scales.

**polity** (from Marshall et al. 2014)
Combined Polity Score: The polity score is computed by subtracting the autocracy score from the democracy score; the resulting unified polity scale ranges from +10 (strongly democratic) to -10 (strongly autocratic).

**durable** (from Marshall et al. 2014)
Regime Durability: The number of years since the most recent regime change (defined by a threepoint change in the polity score over a period of three years or less) or the end of transition period defined by the lack of stable political institutions (denoted by a standardized authority score).

In calculating the ‘durable’ value, the first year during which a new (post-change) polity is established is coded as the baseline “year zero” (value = 0) and each subsequent year adds one to the value of the ‘durable’ variable consecutively until a new regime change or transition period occurs. Values are entered for all years beginning with the first regime change since 1800 or the date of independence if that event occurred after 1800.

- Germany: Recoded so that the reunification is not considered as a regime change.

**system** (from Beck et al. 2001; updated through 2012)
Parliamentary (2), Assembly-elected President (1), Presidential (0)

Systems with unelected executives (those scoring a 2 or 3 on the Executive Index of Political Competitiveness – to be defined below) get a 0. Systems with presidents who are elected directly or by an electoral college (whose only function is to elect the president), in cases where there is no prime minister, also receive a 0. In systems with both a prime minister and a president, we consider the following factors to categorize the system:

a) Veto power: president can veto legislation and the parliament needs a supermajority to override the veto.

b) Appoint prime minister: president can appoint and dismiss prime minister and / or other ministers.

c) Dissolve parliament: president can dissolve parliament and call for new elections.
d) Mentioning in sources: If the sources mention the president more often than the PM then this serves as an additional indicator to call the system presidential (Romania, Kyrgyzstan, Estonia, Yugoslavia).

The system is presidential if (a) is true, or if (b) and (c) are true. If no information or ambiguous information on (a), (b), (c), then (d). Consult Appendix for specific country examples.

Countries in which the legislature elects the chief executive are parliamentary (2), with the following exception: if that assembly or group cannot easily recall him (if they need a 2/3 vote to impeach, or must dissolve themselves while forcing him out) then the system gets a 1.

**fraud** (from Beck et al. 2001; updated through 2012)

Were vote fraud or candidate intimidation serious enough to affect the outcome of elections? This variable captures extra-constitutional irregularities, which are recorded only if mentioned in sources. 0 reported for countries where, for example, opposition parties are officially and constitutionally banned or where irregularities are not mentioned (although may still exist); “1” when opposition is officially legal but suppressed anyway. If not an election year, or if elected government has been deposed, refers to most recent election (i.e. the only way to get rid of a “1” is to hold a fair election). Recording is irrespective of whether only opposition claims that fraudulent elections have occurred or whether allegations are backed by independent international observers. Recorded also are any forms of boycotts carried out by important parties before or after parliamentary elections. In the cases where irregularities are mentioned in the text of the sources, they were recorded. However, there may have been instances of fraud/violence that were not reported, thus resulting in false negatives.

**author** (from Beck et al. 2001; updated through 2012)

Do the state/provinces have authority over taxing, spending, or legislating? If 1 for any of these, category gets a 1. Authority over ‘cultural affairs’, or ‘planning’ in Communist systems, does not qualify. This variable was extensively updated for this version, and as a result, the number of non-missing observations has increased from 38% to 42%.

**tensys** (from Beck et al. 2001; updated through 2012)

How long has the country been autocratic or democratic, respectively?

**ef** (Ethnic fractionalization; from Fearon 2003)

This is an index measuring the ethnic heterogeneity of a country: \( ef = 1 - \sum p_i^2 \) where \( p_i \) is the population share of group \( i \).

- Missing data for Iceland, Luxembourg and Malta.

**prespower1** (index of presidential power; from Doyle & Elgie 2015)
Ranging from 0 to 1; coded zero for countries without a president. Please refer to http://presidential-power.com/?page_id=2151 for a detailed explanation of how this index is calculated.

**prespower2 (index of presidential power; from Doyle & Elgie 2015)**
See above.

**pol_dcen (Political decentralization)**
This is a dummy variable signalling whether policy-making powers are vested in subnational governments/governing institutions. The codings are based on Brancati (2006, 2008; ‘dcen’ variable), Beck et al. (2001; updated through 2012; ‘author’ variable), Treisman (2002; ‘Decentralization’ and ‘auton’ variables) as well as the relevant variables from Hooghe et al. (2010). The policy scope variable from Hooghe et al. (2010) must at least be equal to or above 1 (implying that the subnational government at least has policy-making power in one substantial policy area) for pol_dcen to be coded as 1.

- Bolivia has undergone a process of decentralization, however, as this – until now – solely includes fiscal authority and implementation powers, it is still coded as a 0.
- Croatia: Although the policy scope is 2, Croatia is coded as 0.
- Israel coded as 0 since subnational governing bodies only function as implementers of policy, not as policy-makers.

**pres_concurrent (concurrent presidential elections)**
Coded zero for countries without a president and countries without a popularly elected president (codings based on dates given in Bormann & Golder 2013). Countries for which only a subset of parliamentary seats are elected concurrent to presidential elections (e.g. in Argentina) are also coded as holding concurrent elections.

### 3.4 Performance variables

If the majoritarian tier of an electoral system uses two-round rules, the calculation of vote-based indices is done using the votes from the first round. For mixed-member electoral systems all indices are calculated overall as well as for both tiers individually as the data permits.

**Disproportionality indicators**
In cases where there is missing information about the votes for independent candidates the disproportionality indices are based only on vote and seat shares of those parties and candidates for whom both values have been recorded correctly. This is the most reasonable approximation and potential problems can still be avoided by taking the ‘disaggr_problem’ variable into account as well as by using Taagepera’s (1997) least components approach. Furthermore, especially for independent candidates, the assumption that their over- and underrepresentation cancel out is a fairly reasonable assumption.
**LSI**

\[ \text{LSI} = \sqrt{\frac{1}{2} \sum^n_{i=1} (v_i - s_i)^2} \] where \( v_i \) and \( s_i \) are vote and seat shares of party \( i \) (Gallagher 1991).

**LSI\_maj**

This is the LSI calculated for the majoritarian tier.

**LSI\_pr**

This is the LSI calculated for the PR tier.

**LHI**

\[ \text{LHI} = \frac{1}{2} \sum^n_{i=1} |v_i - s_i| \] (Loosemore & Hanby 1971)

**LHI\_maj**

This is the LHI calculated for the majoritarian tier.

**LHI\_pr**

This is the LHI calculated for the PR tier.

**share\_wasted (Wasted votes percentage)**

This is the share of wasted votes based on dividing those votes that went to parties and candidates who could not gain a seat in the parliament by the total valid votes.

**share\_wasted\_maj**

This is the share of wasted votes calculated for the majoritarian tier.

**share\_wasted\_pr**

This is the share of wasted votes calculated for the PR tier.

**Concentration indicators**

For those cases where there is a substantial amount of successful independent candidates, fragmentation according to votes might actually be lower than according to seats. This phenomenon is based on the lack of disaggregated vote data discussed earlier and the respective elections can easily be identified by the disaggregation problem variable.

**ENPv**

\[ \text{ENPv} = \frac{1}{\sum^n_{i=1} v_i} \] where \( v_i \) is the vote share of party \( i \) (Laakso & Taagepera 1979).

**ENPv\_maj**

This is the ENPv calculated for the majoritarian tier.
This is the ENPv calculated for the majoritarian tier.

ENPv\_pr
This is the ENPv calculated for the PR tier.

ENPs
ENPs = \frac{1}{\sum_{i}^{n} s_i^2} where s_i is the seat share of party i (Laakso & Taagepera 1979).

ENPs\_maj
This is the ENPs calculated for the majoritarian tier.

ENPs\_pr
This is the ENPs calculated for the PR tier.
References


