Electoral Systems: Assessing the Cross-Sectional Time-Series Data Sources

Jan Teorell and Catharina Lindstedt

Abstract

In this paper we compare and assess four freely available cross-sectional time-series data sets in terms of their information on the ballot structure, district structure and formula of the electoral system in use for lower house and, if relevant, upper house and presidential elections. The four datasets evaluated are Golder (2005), the Database of Political Institutions (Beck et al. 2001; Keefer 2005), Persson and Tabellini (2003) and Johnson and Wallack (2006). We find that the choice of data source matters for conclusions drawn on the consequences of electoral systems for both party systems and corruption, but that no data source can be given prominence over the other on methodological grounds. Students of electoral systems must thus, in the future, make their results sensitive to the choice of data source.

Key words: electoral systems, data evaluation, party systems, corruption.

Jan Teorell
Department of Political Science,
Lund University
jan.teorell@svet.lu.se

Catharina Lindstedt
Department of Political Science
University of Gothenburg
catharina.lindstedt@pol.gu.se

Introduction

Few political scientists would deny that the study of electoral systems lies at the heart of our comparative endeavor. A long line of research has established the importance of electoral systems for understanding the workings of democracy (e.g., Duverger 1954; Taagepera and Shugart 1989; Lijphart 1994; Cox 1997; Powell 2000; Norris 2004), and electoral engineering has proven to be a critical issue even under autocracy (Mozaffar and Vengroff 2002; Posusney 2002; Lust-Okar 2006). Yet there is no agreed-upon authoritative source of cross-national data on electoral systems across the globe. Surprising as it may seem, we lack a single political science dataset on a shared set of properties of electoral systems that allows systematic comparisons across countries and over time. The reason for this could hardly be that the necessary theoretical framework for understanding electoral systems is missing. Although the specificities asked for vary depending on one's exact purpose of study, most comparativists seem to agree on some general dimensions along which electoral systems may vary, most notably the ballot structure, the district structure and the electoral formula (Rae 1969; Blais 1988; Cox 1997; Farrell 2001).

The purpose of this paper is to compare and assess four large-*n* datasets in terms of their information on these three dimensions of the electoral system in use for lower house and, if relevant, upper house and presidential elections. The four datasets to be evaluated are Golder (2005), the Database of Political Institutions (Beck et al. 2001; Keefer 2005), Persson and Tabellini (2003) and Johnson and Wallack (2006). These are the largest and most ambitious time-series cross-sectional datasets on electoral systems freely available in spreadsheet format.

Two kinds of criteria will guide our assessment. The first we call *internal* criteria, which pertain to the quality of the descriptive information available. The second criterion we

call *external*, namely whether it matters for substantial conclusions on external outcome variables, what data source on electoral systems we use. We perform this second assessment by exploring the extent to which the different data sources exert different effects in two replication studies of the consequences of electoral systems for party systems and the level of perceived corruption in a country. We find that the choice of data source does matter for conclusions drawn in the external assessment, but that no data source can be given prominence over the other on internal, methodological grounds. Students of electoral systems must thus, in the future, make their results sensitive to the choice of data source.

The paper is organized as follows. In the following section we provide a brief overview of the main theoretical dimensions along which electoral systems may vary. We then give an initial presentation of the four datasets to be reviewed, followed by the internal and external assessments, respectively. We conclude by discussing the implications of our findings.

Dimensions of Electoral Systems

There is fairly strong agreement in the literature, going back to Rae's (1967) seminal contribution, that electoral systems may vary along three generic dimensions: the ballot structure, the district structure, and the electoral formula (Blais 1988; Farrell 2001, 6).[1] Our presentation of these dimensions mostly rely on Cox (1997, chap. 3), although with some simplifications and modifications.

---

[1] Election laws more generally also include several administrative dimensions of elections, such as voter registration and electoral governance (for en encompassing overview, see Massicotte et al. 2004). In this paper, however, we only focus on electoral systems proper, that is, "rules that govern how votes are cast and seats allocated" (Blais and Massicotte 2002, 40).

The *ballot structure* determines how citizens cast their votes and how their votes are counted. A first distinction here regards the entities for which citizens may vote: for party lists (or cartels of lists) only, for individual candidates only, or for a mixture of the two. A second distinction refers to the number of votes that may be cast for these entities (candidates and/or lists). A third distinction concerns whether multiple votes cast are categorical, indicating a preference or not for each candidate/list, or ordinal, requiring a rank-ordering of the candidates and/or lists. A fourth distinction is between single- and multi-ballot systems, the former consisting of only one electoral round whereas the latter might entail more than one round. And a fifth and final distinction concerns what seat-relevant vote totals are affected by the votes cast.[2]

The *district structure* refers to the number, hierarchy and magnitude of the electoral districts used in the system, "where an electoral district is defined as a geographical area within which votes are aggregated and seats allocated" (Cox 1997, 48). There may be one single national electoral district or many. With multiple districts, those may be allocated to a single tier or organized hierarchically into multiple lower and upper tiers. In multi-tier systems, seats are usually allocated within districts of the lower tier first, and then remaining seats within the higher tier or tiers. District magnitude, finally, refers to the number of seats returned from a district.

The *electoral formula* determines how votes are translated into seats. The most basic and well-known formulas are the classical tripartite distinction between plurality, majority and proportional representation (PR) systems. Since formulas are applied for each vote cast at

---

[2] There are many nuanced subdistinctions in this last category, but the most relevant ones concern whether (a) single votes are exclusive or may be transferred or pooled to affect the vote totals of other candidates or lists than the ones voted for, and (b) whether multiple votes need all be cast, need all be cast for a single party, or need all be cast for different candidates (Cox 1997, 41-44).

the district level, it follows that different formulas may be applied in different districts, or more commonly, for different tiers of districts. Moreover, rules may allow cartels of lists to form alliances (such as *apparentement*), and PR systems vary in the way seats are allocated within districts (according to largest remainders or quotas), as well as whether electoral thresholds and bonus seats occur. All these complexities in the process of translating votes into seats are captured by Cox's (1997, 60-64) concept of the "formulaic structure" of an electoral system. This structure operates both at the level of entities (cartels, lists and candidates) and at the level of tiers, with the possibility of having different formulas and other allocation requirements in different parts of the system.

There are some notable features of this three-dimensional approach to describing electoral systems. One is that it is based on those features of a system that are regulated by law. This means that, for example, demographic peculiarities of a country, such as how many people live in different electoral districts, are kept out of the picture, as are procedures for intra-party candidate selection, since there is large variation among countries in terms of whether these nomination rules are regulated by law or not (cf. Gallagher and Marsh 1988).

A second noticeable feature is that the much discussed concept of disproportionality produced by the system is not included. The reason for this is that disproportionality is usually regarded as one of several *consequences* of the electoral system. It is thus left for empirical scrutiny how this relationship plays out in practice.

A third point of notice concerns the focus on dimensions along which electoral systems may vary, as opposed to how combinations of characteristics in different dimensions produce different types of electoral systems taken as a whole. This alternative, typological approach to describing electoral systems is the major and even more "classical" alternative in the literature on electoral systems. Blais and Massicotte (2002) and Farrell (2001), for example, make a typology of electoral systems based on the electoral formula used—

plurality, majority, PR, or mixed—while taking into account district and ballot structure in describing variations among the larger types.

Fourth, and finally, both the dimensional and typological approaches share an ambition to establish a generic description of electoral systems, which, in principle, could be used for any theoretical or empirical purpose. A conceivable alternative, however, is to develop a theory of electoral systems that is specifically tailored to explaining one particular outcome, such as Carey and Shugart's (1995) seminal piece on "the incentives to cultivate a personal vote".

Four Sources of Data on Electoral Systems

The four datasets selected for scrutiny all: (a) include a temporal as well as a cross-sectional dimension, (b) are recorded in a machine-readable or spreadsheet format, and (c) are freely accessible. The first criterion rules out ambitious but purely cross-sectional data collection efforts on electoral systems, such as the one provided by IDEA (Reynolds et al. 2005) or the Comparative Study of Electoral Systems (www.cses.org). The second criterion rules out data handbooks with all information provided in printed tables, such as the seminal handbook series edited by Dieter Nohlen (Nohlen et al. 1999; Nohlen et al. 2001; Nohlen 2005). To the best of our knowledge, this basically leaves us with four datasets: Golder, the Database of Political Institutions (DPI), Persson and Tabellini (PT), and Johnson and Wallack (JW). [3] Let us take a closer look at each in turn.

---

[3] The only two exceptions, known to us, are Gerring et al. (2005) and Huber et al. (2004), both of which however only provide one single variable on electoral institutions. Moreover, in the Gerring et al. case this variable is arguably of little import for other purposes than the study of "centripetalism" (for example, majoritarian and open-list PR systems are grouped in the same category, without allowing any distinctions

Golder

Golder (2005) provides electoral system data for 867 legislative elections to national lower houses and 194 presidential elections in countries all across the globe during democratic periods between 1946 (or independence) and 2000.[4] His criterion for selecting democratic periods is based on Przeworski et al.'s (2000) dichotomous measure of democracy, which he has systematically updated to cover the entire time period.

Golder basically follows the classification scheme of Blais and Masicotte (2002), but adds an additional category, multi-tier systems with a single formula, which can thus include both traditional "PR" and some plurality systems. He treats plurality systems as a subcategory of the larger "majoritarian" family, in which his data also distinguishes among absolute majority, qualified majority, limited vote, alternative vote, single non-transferable vote (SNTV), and the "Modified Borda" system (only in use in Nauru). Within the PR family, Golder distinguishes between single transferable vote (STV) and list systems, while the latter group covers the more exact nature of the formula being used, including both highest average (D'Hondt, Saint Laguë and Modified Saint-Laguë) and quota (Hare, Droop and Imperiali) systems.[5] His classification of mixed systems follows Massicotte and Blais's (1999) five-fold typology, which is based on the distinction of whether the two systems work independently or not.

Golder's data are thus fairly rich in terms of the formula in use, which also applies to his data on presidential elections. He includes several indicators of district structure: the

---

among the two). In the Huber et al. case, only 18 countries are covered and the information is for all practical purposes cross-sectional (only two country year observations score a change in the electoral system in use).

[4] Golder's original data may be accessed at: http://homepages.nyu.edu/~mrg217/elections.html.

[5] Blais and Massicotte (2002, 42-54) provide an excellent overview of differences among plurality, majority and PR systems.

number of seats in the legislature, the number of electoral districts, the median and average district magnitude in the lowest tier, as well as the number of seats allocated in the upper tier. Apart from what may be indirectly inferred from knowing the type of system, however, there is very little independent information on ballot structure. The major exception is an indicator for both legislative and executive elections of whether a runoff (dual ballot) is held or not.

The Database of Political Institutions

The Database of Political Institutions (DPI) provides, among other things, data on electoral systems in use for both lower and upper houses in a global sample of countries (democratic or not) in 1975-2004 (Beck et al. 2001; Keefer 2005).[6] The DPI electoral data is like Golder's mainly concentrated to the formulaic structure. One dummy variable registers whether plurality, another whether PR, is used to select any candidate in any house. Another indicator, one for each house, codes the formula that governs the majority of the house seats (with a special code for tied houses, where 50% of the candidates are selected by each rule). The minimum threshold in use in PR systems is coded, as well as a dummy for whether the d'Hondt formula is in use.

There is again scarce information on ballot structure, the exception this time being an indicator for whether closed lists (where no preferences for candidates may be expressed) are in use under PR. The district structure is covered by variables recording the number of seats as well as the average district magnitude for both lower and upper houses. These average district magnitudes are, where information is available, constructed as "the weighted average of the number of representatives elected by each constituency size" (Keefer 2005, 16).

---

[6] The original data of the DPI may be accessed at: http://go.worldbank.org/2EAGGLRZ40.

Persson & Tabellini

Person and Tabellini (2003) have put together a dataset on (among other things) rules in use for elections to the lower house, covering at best 85 democracies in 1960-1998.[7] They select democracies by applying thresholds in the Polity and Freedom House scores of a country. This is not a multi-purpose dataset; it has been consciously designed to test a specific theoretical argument: how electoral (and other political) institutions affect the government's provision of public vs. targeted goods and rent extraction through the choice of economic policy. This proviso must be kept in mind when judging the quality of their data.

All three dimensions of electoral systems are presented in Person & Tabellini's data, although in terms of very few indicators. The ballot structure is captured by two variables, one reflecting the share of legislators elected by plurality rule via a vote on individuals, the other gauging the share of legislators elected as individuals (by plurality rule) *or* on open lists. Although these two variables thus take both formula and ballot structure into account, they tap more into how votes are cast than how they are translated into seats. The district structure is (apart from the number of seats in the lower house) captured by two inverted district magnitude variables: the usual average district magnitude, but where the number of districts are divided by the number of seats rather than the opposite, and a weighted average, "where the weight on each district magnitude in a country is the share of legislators running in districts of that size" (Persson and Tabellini 2003, 92). The electoral formula, finally, is captured by a single indicator of whether a country elects its lower house exclusively through

---

[7] Persson and Tabellini's original data may be accessed at: http://www.igier.uni-bocconi.it/whos.php?vedi=1169&tbn=albero&id_folder=177.

plurality (or majority) rule.[8] PT provide their electoral formula dummy in annual format, covering the time period from 1960-1998. The other electoral system variables are provided as country averages across the time period 1990-1998 only.

Johnson & Wallack

The explicit purpose of Johnson and Wallack's (2006) data on electoral rules is to provide measures for the variables, theorized by Carey and Shugart (1995), to provide incentives to cultivate a personal rather than party reputation.[9] The first, called *ballot*, reflects the degree of control party leaders exercise over party endorsement and, in list systems, ballot rank. The second, called *pool*, measures "whether votes cast for one candidate of a given party also contribute to the number of seats won in the district by the party as a whole" (Carey and Shugart 1995, 421). The third, called *vote*, captures whether votes are cast for a party or a candidate, and whether there are one or multiple votes and/or ballots. The fourth is simply the *district magnitude*. The data covers both upper and lower houses in a global sample of countries (democratic or not) in 1978-2005.

The lion share of the variables thus covers the ballot structure under Carey and Shugart's (1995) three conceptual headings: ballot, pool and vote. Up to six variables for each of these headings are coded per country, including separate measures for single-member districts (SMDs) and multi-member districts (MMDs) for the lower and upper house, and average measures (for both houses, where applicable). The data also contains two summary

---

[8] In the book, Persson & Tabellini (2003, 83) also include an indicator of whether a mixed formula is used. Unfortunately, however, this indicator never made it to their published data.

[9] Johnson and Wallack's original data may be accessed at: http://dss.ucsd.edu/~jwjohnso/espv.htm. This dataset expands and (in some cases) corrects the previous data provided by Wallack et al. (2003).

rank-orderings in terms of the incentives to cultivate a personal vote, as well as indicators for whether dual ballots are held and whether voters are given one vote per tier.

The district structure variables include the size of the legislature, the average and weighted average district magnitude, the presence of multiple tiers, and the proportion of legislators elected in a national tier, in single-member (SMDs) or multi-member districts (MMDs)—all for both houses (where applicable). The one dimension of electoral systems hardly covered this time is the electoral formula. One single variable may be interpreted to represent this dimension, namely an indicator of whether multiple tiers are elected in parallel or compensatory fashion.

Internal Assessment: The Quality of Descriptive Information

We shall begin our evaluation of these four data sources according to internal assessment criteria, i.e. criteria that pertain to the quality of descriptive inferences on electoral systems that may be drawn from the data. How useful are these data sources for understanding the electoral systems in use across the world? Given the lack of generally agreed upon standards on which to base such an assessment (cf. Herrera and Kapur 2007), we concentrate on four internal assessment criteria: coverage, conceptual relevance, clarity and concurrence.[10]

Coverage

Coverage includes two different aspects. Dimensional coverage concerns the extent to which the three dimensions of electoral systems are covered by a data source. We have already noted how the four data sets differ in this respect. Couched in terms of the number of variables supplied for each dimension, the patterns are summarized in the upper panel of Table 1.

Whereas Golder and the DPI primarily convey information about the electoral formulas in use, paying minor attention to ballot structure, JW do the exact opposite: they concentrate on ballot structure and almost completely ignore the formula. PT are more balanced in this regard, but in relative terms convey fairly little information on both ballot structure and electoral formula. With respect to district magnitude all give the most basic information, but JW have an edge in terms of providing the most extensive information. In terms of the elected bodies covered, finally, Golder alone provides information on the formulas used in presidential elections, whereas the DPI and JW are alone in providing information on electoral rules for upper houses.

Empirical coverage instead concerns the spatial and temporal scope of the data (see the lower panel of Table 1). In terms of a "core" variable of each data source,[11] we observe fairly substantial differences among the datasets in terms of spatial outreach. The DPI fares far best in this regard, with a total of 165 countries covered at some time point, whereas PT, at the other end of the spectrum, include only 61 countries at best. This, of course, reflects the restrictive selection criterion for being a democracy they apply. Golder, also using a democratic selection criterion, falls somewhere in between those extremes with a total of 124

---

[10] Our assessment is facilitated by the fact that all four datasets on electoral systems have already been included in the freely available Quality of Government Dataset (Teorell et al. 2008): http://www.qog.pol.gu.se/.

[11] Since the empirical coverage varies considerably from one variable to the other within each dataset, we selected a core variable from each dataset in order to make a fair comparison. In the case of Golder, this is the variable classifying each system according to its general formula (plurality, PR, multi-tier or mixed). In the case of the DPI, we chose the indicator of whether plurality rule is used or not, which is the one with the best coverage. Persson & Tabellini only have one variable with annual time coverage: the plurality/majority dummy. Johnson & Wallack, finally, have the largest number of variables in their dataset, but, from a theoretical point of view, one of them clearly stands out: the summary rank-ordering of electoral systems in terms of their incentives to cultivate a personal vote.

countries covered. Interestingly, the figure for JW is similar (127), despite the fact that they lack an explicit selection criterion.

With respect to time, Golder covers the longest historical period, going back all the way to 1946, whereas the DPI and JW have more recent starting points in the 1970s. PT fall in between, although it should again be noted that they only provide annual time coverage for one single variable.[12]

Taken together, the time and spatial coverage expressed in terms of country years is most extensive for the DPI, followed by Golder and then (at approximately similar levels) PT and JW. What does this tell us about the representativeness of the countries covered by each dataset? Table 2 provides the means in various standard cross-national indicators, both globally and among the cases for which there is valid information on the core variable of each data source. The larger discrepancy between the mean per data source and the global mean, the less representative the source. Not surprisingly, rich countries are overrepresented in all datasets, although this difference decreases if we look at the Human Development index rather than at levels of income. Larger countries are also overrepresented, but a more important discrepancy concerns the nature of the political regime. Whatever measure of democracy is used, data on electoral systems cover democratic countries to a far larger extent than non-democratic countries. At first sight, one might think this is because elections are not held in autocracies, hence there are no electoral systems. But this is not the case. Only a handful of countries in the world hold no elections whatsoever, although a large number of countries stage seriously flawed and rigged elections (Golder 2005, 105). As they do hold

---

[12] We thank one of the anonymous reviewers for pointing out that none of these datasets take the *timing* of electoral reform into account. All four sources code the implementation of election laws in election years, but none of them code separately when these reforms were enacted in the first place. This could for example be of potential importance when the design of electoral systems is to be the dependent variable.

elections, they also must have an electoral system. Particularly in authoritarian systems with some degree of multiparty competition, reform of electoral rules (usually rigged to the incumbent's favor) is generally high on the political agenda (Mozaffar and Vengroff 2002; Posusney 2002; Lust-Okar 2006). It is thus not the case that electoral systems are merely ephemeral creatures in authoritarian settings. Hence, these authoritarian electoral systems remain an untapped source of information about the workings of elections in the world.

[Table 2 about here]

There is however one data source on electoral systems that does not exclude poor, less developed, small and undemocratic countries to the same extent as the others: the DPI. Overall, then, the DPI provides the best empirical country coverage with the most globally representative selection of countries. Since most research on electoral systems have been driven by a focus on at least minimally democratic elections, one could of course argue that coverage should be assessed with this narrower population of countries in mind. Using that comparative yardstick, Golder provides the broadest and most representative coverage.[13]

Conceptual relevance

Our second assessment criterion concerns how well particular variables map into the conceptual language used in the theoretical literature on electoral systems. Whereas JW are exemplary in this regard (implementing, as they do, a singular theoretical approach), we have

---

[13] This follows from the observation that the DPI covers fewer countries and a shorter time period among the observations for which Golder, using a minimal democracy criterion, has data. It is also the case that the DPI provides data for a slightly less representative set of democratic countries than does Golder, particularly in terms of GDP per capita and population size.

found some deficiencies in the other three datasets. The major problem with the DPI concerns the way in which it aggregates potentially varying electoral formulas used within the same country. This task is performed in two different ways. On the one hand through two dummy variables: one indicator for whether plurality, the other for whether PR, is the formula applied for selecting candidates *to any house*. When both these variables score 1, one would thus be inclined to infer that a mixed system is in operation. That need not be the case, however, since a proportional formula could be in use for the lower, but plurality for the upper house, as in Spain, Bolivia and Brazil. The DPI's twin set of indicators are thus not applied in a way that is consistent with the standard definition of mixed systems, which requires different formulas to be in use for elections *to a single body* (Massicotte and Blais 1999, 345; cf. Shugart and Wattenberg 2001, 10; Golder 2005, 111). The second strategy applied by the DPI, to code whether the majority of seats in each house is selected by plurality or not, unfortunately fares no better, since a legislature such as the one in New Zealand, where 54 % of the seats are allocated by plurality and 46 % by PR, is coded the same as a legislature where all seats are allocated by plurality. Save for the special case where the split between plurality and PR seats is even (in which case this indicator is scored .5), this again means that there is no consistent way to discern mixed systems in the DPI data.

The DPI also has another problem, that it shares with PT, namely that no distinction is made between plurality and majority formulas. Since the literature has pointed to a closer affinity of majoritarian systems to PR than to plurality in some regards (Blais and Masicotte 2002), this is potentially a major drawback.

The conceptual problem in Golder's dataset concerns the category of "multi-tier" systems. In his coding scheme these are all systems with more than one electoral tier but where a single formula is being used. To the best of our knowledge, however, there has never been any theoretical proposition in the literature on electoral systems that pay attention to this

category. The absolute majority of Golder's "multi-tier" systems are pure proportional systems, and hence for theoretical purposes should better be treated as such. The fact that they are multi- and not single-tier must be a subsidiary distinction, which anyway is already independently measured by another variable in Golder's data. The problem is that some multi-tier systems are actually of the plurality type. As it stands, then, Golder's "multi-tier" variable confuses the basic typology of electoral systems, which he purports to measure.[14]

Clarity

Rather than being concerned with theoretical relevance, we deal under this heading with several issues concerning how easily interpretable the data are for a potential user. A first issue concerns the general appearance of the data sets and the structure and consistency of their measurements. By and large we find all dataset to perform satisfactory in this regard, with the partial exception of Golder. To begin with, Golder has disaggregated most of his classifications so that one variable appears for each final node in the classification tree, leaving no information as to its preceding branches. For example, rather than having one variable each containing his ninefold classification of PR systems, sevenfold subdivision of majoritarian and fivefold typology of mixed systems, Golder provides 21 dummy variables, one for each type. In the codebook these variables are then listed in alphabetical order together with all other variables (including, e.g., district structure and presidential systems), without any indication as to which variable belongs to what family in the classification tree. It takes a great deal of time and effort before a potential user can make sense of this data

---

[14] Of course this is no insurmountable obstacle: since Golder's rich codebook clearly specifies the two single countries that apply a majoritarian formula in multiple tiers (Papua new Guinea and Mauritius), the Quality of Government Dataset (Teorell et al. 2008) has provided an alternative variable for Golder's basic typology, where the multi-tier countries are being reclassified as being either of the PR or majoritarian type.

structure. A second point of confusion in Golder's dataset concerns the consistency in how to supply annual measurements. Since electoral rules in these datasets are only coded to the extent they are implemented in election years, the natural unit of observation for these data is the election-country-year rather than the country-year. The question then arises what to do with observations for non-election years. They could be either left missing, or the values for the last election could be repeated until the next election year. The latter strategy has been used in all the other datasets under scrutiny, but Golder unfortunately uses a combination of this latter strategy for some variables, and a clearly inferior alternative for others: to replace the in-between election year observations by trailing zeros. This causes problems when the zero is a code with substantial meaning.[15]

A second issue concerns the clarity and transparency of the coding rules applied. For most of the four datasets this causes minor problems. Where the variables include more complex operations, such as how to compute the average district magnitude, crisp definitions are key. Both Golder and PT are strong in this regard, and the DPI fares relatively well despite some ambiguities with respect to when the district magnitude is weighted and when not. For JW, however, this is a serious obstacle. Recall that their task is to translate real world systems into graded measures of the importance of personal reputations. This involves a complex mapping process, which unfortunately is not always conveyed with exemplary clarity. Moreover, as will become clear below, JW are not consistent in the way they compute the average district magnitude in multitier systems.[16]

---

[15] Golder's data are again not beyond repair in either of these respects, and the Quality of Government Dataset have addressed both problems in order to present a more tractable version of his original election year data.

[16] A third and related issue concerns the documentation of which sources have been used for coding electoral systems. Such reference lists are provided by all four sources, although at a level of detail which is somewhat larger in the cases of Golder and JW.

Concurrence

We now turn to our fourth and final internal assessment criterion: the extent to which the differing data sources concur in their classification of electoral systems.[17] First and foremost, we must take note of the fact that there is surprisingly little information in these four datasets that is strictly comparable, in the sense that a pair of sources have attempted to measure exactly the same phenomenon. Even when there is at first appearance a general agreement on what aspect of the electoral system is being measured, as we shall see, there might be nuanced differences occurring in practice.

Table 3 presents all variables that intend to measure roughly the same aspect of the electoral formula in use. In all instances we are dealing with dichotomous indicators. We have therefore reported the proportion of observations (election years) falling along the main diagonal in a cross-tabulation of each pair of data sources, together with Scott's (1955) inter-coder reliability measure, which also takes into account the agreement that may emerge out of pure chance.[18] Among the variables purporting to classify electoral systems as being purely majoritarian/plurality or not, there is actually only one pair of data sources that agree perfectly in terms of what they *intend* to measure. That is Golder and PT, which both exclusively focus on elections to the lower house, and both provide measures that explicitly leave out all mixed systems from the measure.[19] Not surprisingly, then, the largest concurrence applies to this

---

[17] This criterion is "internal" in the sense that it still pertains to the quality of data on electoral systems themselves, rather than to what may be inferred about the relationship between electoral systems and other, "external", country characteristics.

[18] Scott's π = (observed concurrence – expected concurrence) / (1 – expected concurrence).

[19] In Golder's case this equivalence was however achieved only after reclassifying his "multi-tier" category, as stated above.

paired comparison: 94 percent of their classifications of majority/plurality rule are in agreement with each other, corresponding to a reliability score of .86.

[Table 3 about here]

One could, of course, still ponder how it is possible to disagree at all about such a basic feature of the electoral system as the formula in use. Looking at what countries account for the discrepancies between Golder and PT's classifications, however, makes clear that not even this undertaking is simply a matter of converting hard facts into numbers. Take Chile, for example, having PR in two-member districts, but with the plurality rule determining which candidate on each party's list is being elected. Is this a PR or a plurality system? Nominally, of course, Chile is a PR system, and this is how Golder has classified it. But the logic of its operation is more akin to a plurality system, which is how it is treated by PT. Japan is another delicate case. The SNTV system for the lower house, in use until the electoral reform in the 1990s, basically implies plurality rule in multi-member districts, with each voter casting a single vote. Again this is nominally a plurality system (Golder's coding), but in terms of its logic of operation it has so many similarities with other large district systems (i.e. PR) that it sometimes in the literature is accorded the label "semi-proportional" (e.g. Farrell 2001, 45-47). PT thus have some theoretical rationale for *not* classifying it as a plurality system.

The following four comparisons of the majority/plurality measures are all with the DPI, which—as stated above—is somewhat ambiguous in the way it measure the formula in use. The indicator, constructed for comparative purposes, scores a case as majority/plurality when this formula in the DPI is coded as being in use for selecting any candidate to any house, while at the same time the PR indicator states that PR is not in use (to any house). The

18

problem with this indicator (as mentioned) concerns bicameral systems, where PR might be in use for the upper house only, whereas all lower house representatives are selected by majority/plurality. Not surprisingly, the classification based on DPI concurs to a lesser extent with both Golder and PT, especially when the possibility of merely random concurrence is taken into account (according to Scott's π). When attention is restricted to unicameral systems, however, these sources all concur very well in terms of their classification of the plurality/majority formula in use.

The comparisons indicate less concurrence, when we want to classify the electoral formula as PR or not. The DPI indicator now only concurs with Golder's classification in 84 percent of the cases, leading to a meager reliability score of .69. This situation improves when we restrict attention to unicameral systems, but is still far below the concurrence for the majority/plurality vs. others distinction. A noteworthy fact is also that DPI and Golder are in rather sharp disagreement over whether the d'Hondt formula is in use for allocating seats within districts. However, the worst concurrence in terms of the reliability score pertains to DPI's and Golder's classification of whether a system is mixed or not. This is a troubling result, given that these two are the only data sources that include codes for mixed systems. Among the mixed systems, on the other hand, there is great concurrence (although for a limited set of observations) between Golder and JW as to whether the upper tier is being used to compensate for distortions in the lower tier or not.

Turning to pairwise concurrence in average district magnitude, it should first be noted that there is not one instance in which two sources apply exactly the same definition of average district magnitude. Golder does not weight his average district magnitude measure, while the DPI sometimes does, depending on the availability of information. PT and JW provide both weighted and unweighted measures. Moreover, Golder only gives estimates for the lower tier, whereas the other sources average across multiple tiers, albeit in different ways.

Given the complexity and potentially different outcomes involved, perhaps the most sensible solution would have been to refrain from computing one overall "average district magnitude measure" in multi-tier systems. This is however not how the DPI, PT or JW has approached this problem. The DPI appears to take the average of the average district magnitudes across tiers, whereas PT divide the total number of seats by the number of districts summed across all tiers. This could lead to substantially different conclusions for multitier systems. In the Russian electoral system in use until the 2004 elections, for example, 225 seats of the State Duma were elected in single member districts in the lower tier, while the remaining 225 seats were elected on party lists in one national district in the upper tier. According to Golder, then, the average district magnitude in this system was 1, only taking account of the lower tier. According to the DPI, on the other hand, the average district magnitude was 113; that is, the average of 1 and 225, the latter being the district magnitude in the upper, national tier. PT, finally, designate the figure 1.99 to the average district magnitude of the Russian system; that is, 450 (the total number of seats) divided by 226 (the total number of districts in any tier, or 225+1). Unfortunately JW alternate between these three different ways of measuring district magnitude.[20]

These caveats notwithstanding, the pairwise concurrence with respect to average district magnitude is fairly strong. Figure 1 presents the scatter plot matrix together with the correlation coefficients for the three data sources that provide time-series information on district magnitude; the latter is measured by its inverse in order to ameliorate the influence of

---

[20] In the Russian case JW follow PT, whereas in Germany they apply the DPI formula; in yet other instances, such as in Sweden, they disregard the upper tier entirely (and thus follow Golder's strategy). There may be some logic to these alternations, but this is not explicitly stated.

extreme outliers. The DPI correlates at .95 with JW's weighted measure.[21] The main exception is Golder's measure, which only correlates at around .80 with the other sources, but this can easily be attributed to the treatment of multi-tier systems. If attention is restricted to single-tier electoral systems (results not shown), Golder's district magnitude correlates at .96-.99 with the DPI and JW's measures. A similar conclusion goes for the congruence between PT's two inverted measures of district magnitude in a cross-section of countries averaged over the period 1990-1998 (results not shown). The correlation between their unweighted measure and the inverted version of all other sources however still range between .89 and .96, although their weighted measure correlate only at .63-.87.

[Figure 1 about here]

The only information on ballot structure that is comparable are PT's two variables on the share of legislators that are elected individually by plurality rule or on open lists, which have rough correspondents in JW's data. Particularly the ballot version that does not incorporate the open vs. closed list distinction correlates fairly high (at .87), despite the fact that being selected by plurality is conceptually not the same as being selected in a single-member district (the latter is what JW measure). The second version concurs to a lesser extent across the two data sources (at .72), which probably reflects the problems involved in singling out open-list PR candidates from JW's ballot variable.

Overall, then, the different data sources do not concur perfectly in their measures of electoral systems, but this stems particularly from the fact that they do not purport to measure

---

[21] When comparing average district magnitudes in the upper house, for which figures are provided by the DPI and JW, the overall correlations are weaker: the DPI measure correlate at .82 and .77 with the two JW measures.

exactly the same properties. Given this, does the choice of a particular source matter for explaining real-world outcomes? This is the question we raise in the following section.

External Assessment: Consequences for Party Systems and Corruption

We term this an external assessment, since what is now at stake is the relationship between electoral systems and other country attributes external to this system. The two particular studies we have chosen to replicate are Clark and Golder's (2006) reassessment of Duverger's (1954) "law" and Persson et al.'s (2003) study of the effects of electoral rules on corruption. There are several reasons for this choice. To begin with, both studies rely on one of the data sources we are assessing (Golder and PT).[22] Second, the study of how electoral rules shape party systems is one of the most established research programs in comparative politics (e.g., Taagepera and Shugart 1989; Lijphart 1994; Norriss 2004), whereas the consequences of electoral systems for corruption is an emerging field wrought with conflicting findings (e.g., Kunikova and Rose-Ackerman 2005; Chang and Golden 2006; Treisman 2007). Examining the robustness of previous findings across these vastly opposing settings provides, as it were, an "extreme bounds" assessment of the potential consequences of choosing one data source over another. Third, and finally, the Clark and Golder (2007) and Persson et al. (2003) studies are models of scholarship in their conceptual clarity, theoretical sophistication, and empirical thoroughness. Their setup is thus easily followed and replicated.

---

[22] This is at least almost the case: there are actually some minor discrepancies between the electoral systems data in Persson et al. (2003), generously provided to us by the authors, and the published Persson and Tabellini (2003) data. Clark and Golder's (2007) replication data are available at: http://hdl.handle.net/1902.1/10477.

Testing the Mechanical Effects of Electoral Systems

Clark and Golder (1997), themselves replicating and improving on a previous study by Amorim Neto and Cox (1997), provide a straightforward test for what is usually termed the "mechanical" effect of electoral systems according to Duverger (1954): the way they influence the translation of votes into seats. More specifically, they test the extent to which larger district magnitude affects the correspondence between the effective number of parties in the electorate and in the legislature. It is widely held that electoral systems with single-member districts should incur a shrinkage of the number of parties elected to parliament, as compared to the number of parties receiving votes in the electorate. The larger the number of candidates returned from each district, however, the better should the partisan composition of legislature reflect how votes are cast.[23]

In Table 4, we replicate this test based on Clark and Golder's (2007) original data, supplemented with the two other data sources on (the natural logarithm of) average district magnitude – the DPI and JW – merged from the Quality of Government Dataset (Teorell et al. 2008). In all assessments the same model specification as in the original analysis is being used, including the same control variables and estimation methods. This means that we also take into account the extent to which an upper compensatory tier ameliorates the disproportionality effect of single-member systems, although Clark and Golder (2007) does not find any systematic evidence for this. We also compare findings restricted to the same set

---

[23] This is the first component of "Duverger's law". The second, strategic or "psychological", effect instead pertains to the tendency among voters to avoid casting wasted votes for parties affected by the mechanical effect. This effect is also tested by Clark & Golder (2006) by studying the extent to which the effect of ethnolinguistic fragmentation on the number of parties in the electorate is affected by the size of districts. Since their findings in this regard is less robust even to begin with, we refrain from replicating them here.

of observations across datasets, in order to rule out selection bias as the source of possible differences in results.

[Table 4 about here]

Model (1) replicates Clark and Golder's (2006) main result for the entire sample of democratic election years. The coefficient of .56 for *ElectoralParties* is the marginal effect of the effective number of electoral parties on the effective number of legislative parties in single-seat districts (since ln[1]=0) with no upper tier seats. As expected, this coefficient is smaller than 1, thus providing evidence for "shrinkage" (or disproportionality). The positive and significant coefficient of 0.07 for the interaction term with ln(*Magnitude*), however, indicates that in elections with larger districts on average, there is less shrinkage (or more proportionality). This thus confirms the mechanical effect hypothesis. What happens when we replace the average district magnitude measure by approximate equivalents from the other data sources? As can be seen in model (2), when the DPI is used as data source the interaction term decreases to 0.02 and is no longer statistically significant, despite being based on exactly the same set of observations. The same is true for the JW data in model (4). In both cases, single-member district systems create disproportionality, but larger districts do not on average compensate for this.

How can we explain these surprising and contradictory results? Given what we noted before about the different strategies these data sources use to measure average district magnitude in multitier systems, our first suspicion would be that results should be more robust when these systems are excluded. This also turns out to be the case for the DPI data, whose coefficient for the interaction term is positive and significant once we restrict the analysis to single-tier systems (results not shown). In the case of JW, however, the coefficient

24

for the interaction term only passes marginal thresholds for significance (p=.075) even in singe-tier systems. The robustness problem thus cannot be fully explained by the problem of how to measure average district magnitude in multi-tier systems.

A second possibility would be to suspect nascent democracies to cause conflicting findings, given that they have relatively immature party systems. For this reason models (5)-(8) of Table 4 are restricted to what Clarke and Golder (2007) term "established democracies", in effect excluding all countries that did not become democracies until the 1990s. As can be seen, the mechanical effect of having larger district holds up more robustly in this setting. The coefficient for the interaction term is very similar in magnitude when Calrk and Golder's variable is replaced by both the DPI (0.06) and JW (0.05) as data sources.[24]

In sum, even for a well-established research proposition, such as the mechanical effect of electoral systems, the choice of data source may affect results at the margin. This appears to occur in particular when studying average district magnitude in multi-tier systems and in nascent democracies. We now turn to our assessment of the difference our choice of data source makes when a newer proposition is put to the test.

Electoral Rules and Corruption

Persson et al. (2003) most conclusively muster support for two hypotheses relating electoral rules to corruption. The first predicts that countries with larger district magnitudes have lower barriers to entry into politics. This enlarges the pool of honest candidates, helps voters oust dishonest incumbents, and thus should overall lead to less corruption. The second hypothesis instead predicts that elected officials concerned about their future careers have stronger

incentives to avoid abuse of office when held individually accountable at the polls. This implies that corruption should be decreasing in the number of candidates elected on an individual ballot by plurality rule (or, although less robustly, on open lists by PR).

In Table 5, we replicate Persson et al.'s key results based on the author's original data from 80 democracies in the 1990s, again holding model specification and the set of sample observations constant. The measure of corruption is from the World Bank Institute Governance Indicators as of around 1998 (Kaufman et al. 1999), rescaled so that higher values indicate more corruption, whereas the electoral systems data is entered as averages over the 1990-98 period.

[Table 5 about here]

Model (1) perfectly replicates the predicted negative effect of the share of legislators elected by plurality rule via a vote on individuals (*PINDP*), and the positive effect of the inverse of district magnitude (*MAGN*). When replaced by the inverse of the average district magnitude from the DPI data in model (2), however, these findings are disconfirmed. Both the magnitude measure, and the original *PINDP* measure, are incorrectly signed and far from statistically significant. If we instead turn to JW's data in model (4), where we can find a substitute also for *PINDP*, the same pattern emerges. As shown in model (3), where the original results hold in the somewhat reduced sample of countries, this is not due to selection bias. Moreover, if we exchange the *MAGN* and *PINDP* measures one at a time with the JW correspondents (results not shown), we find that both are to blame for this result. That is, the

---

[24] The fact that the coefficient in the JW case is only marginally significant cannot be explained by the use of data, since on the same 291 election year observations it is only marginally significant for CG as well.

effects on corruption do not hold when either one of them is replaced. With respect to Golder's district magnitude measure, however, we find in model (6) that the original findings hold.

This time "nascent democracies" (those having transited after 1989) cannot explain divergences in findings. Again we may however expect countries with multi-tier systems to account for them. It also turns out that when the JW measure of district magnitude replaces the original measure on a restricted sample of countries with a single tier, the original result again holds up (despite the reduced sample size of only 46 countries). The discrepancies with respect to the DPI, however, cannot only be explained in terms of how multi-tier systems are treated, since both Singapore and Cyprus are countries influencing the difference in results despite having single tiers. Moreover, the fact that multi-tier systems have to be excluded in order to sustain their findings is a concern for the robustness of Persson et al.'s (2003) results, since no one can claim to measure district magnitude in these systems "better", or more "correctly", than any one else. It is simply the case that it can be done in different ways, and what way is chosen matters for the results.

With respect to *PINDP*, there is a single country that can explain the entire difference in results: Chile. As noted above, Chile is an ambiguous case, but PT concentrate on its plurality features (scoring *PINDP*=1). Since no one is elected in SMDs in Chile, however, the JW version comes to the opposite conclusion (scoring *PINDP*=0). By reversing this single decision, the original results again reappear. Again, however, this is not an issue of being "right or wrong" about Chile. Either decision could be defended on theoretical grounds. The fact that this single decision matters for the outcome thus casts additional doubt on the robustness of Persson et al.'s (2003) findings.

To sum up, neither of Persson et al.'s (2003) hypotheses are robust to the choice of data source for measuring electoral systems. This can again partly (but not fully) be explained

by the problem of how to compute average district magnitude in multi-tier systems, but also by the coding choice for one particularly influential outlier (Chile).

Conclusion

This paper has assessed four cross-sectional time-series sources of data on electoral systems. In terms of our internal assessment criteria, we have not been able to single out any particular data source as being better than all the others on all dimensions. Most data sources have an unbalanced coverage of the three theoretical dimensions of electoral systems, with JW being heavily skewed towards covering the ballot structure, and the DPI and Golder concentrating primarily on the formula. PT cover all dimensions, but in far less detail. The DPI is strongest in terms of global coverage, but suffers from problems of classifying electoral systems as unambiguously based on the plurality, majority, PR or mixed formula. Golder is strongest in terms of spatial and temporal coverage of democracies. However, Golder's data are not ideally structured, and his peculiar "multi-tier" category causes conceptual concerns. PT's data are clearly the narrowest, both in terms of country coverage and the level of detail, and the single indicator for which they provide annual over-time information does not distinguish among plurality and majority formulas. JW, finally, provide extensive information at a large level of detail for a substantial number of countries; but mostly on one singular dimension of electoral systems, that of the ballot structure. Their coding rules are a also a cause for concern.

An assessment of the extent to which different data sources concur in their information of the same countries and time periods is only possible for some very basic characteristics of electoral systems. The reason for this is that different data sources rarely attempt to measure exactly equivalent properties of electoral systems even to begin with. When it comes to the most basic of all distinctions, that between PR and plurality/majority systems, the degree of concurrence is fairly satisfactory, although not perfect. The concurrence in terms of average

district magnitude is somewhat troublesome in multi-tier systems. Moreover, the absence of perfect concurrence across data sources does have implications for studying the consequences of electoral systems. In our external assessment, we find that the choice of data source influences conclusions drawn on the effect of electoral systems on both party systems and corruption.

The conclusion we draw from these findings is that comparativists must be sensitive in their choice of data source on electoral systems in the future. Since no particular data source has prominence on methodological grounds, only strong theoretical claims can justify choosing one source over the others. We expect that most of the time, the most cautious strategy is to use more than one different source for studying each particular aspect or dimension of the electoral system, and assess the robustness of one's result across sources. True, a qualification to this recommendation is in place. Since our replication studies show that results are less sensitive when confined to established research propositions (such as "Duverger's law") in established democracies with "pure" single-tier systems, researchers oriented towards these settings may thus expect less cause of alarm in their choice of dataset. However, with complex multi-tier mixed electoral systems on the rise even among advanced industrialized democracies (Golder 2005), studies thus restricted may end up with theories and results confined to a shrinking part of the globe.

In the future, more theoretical work has to be done on how to gauge the district structure in multi-tier systems, and why the choice of one particular gauge may matter for certain outcomes (such as corruption). More needs to be learned on electoral systems in authoritarian or semi-competitive elections. More detailed information on the ballot structure is also required. Hopefully this will also lead to improvement in the existing data sources over the course of time.
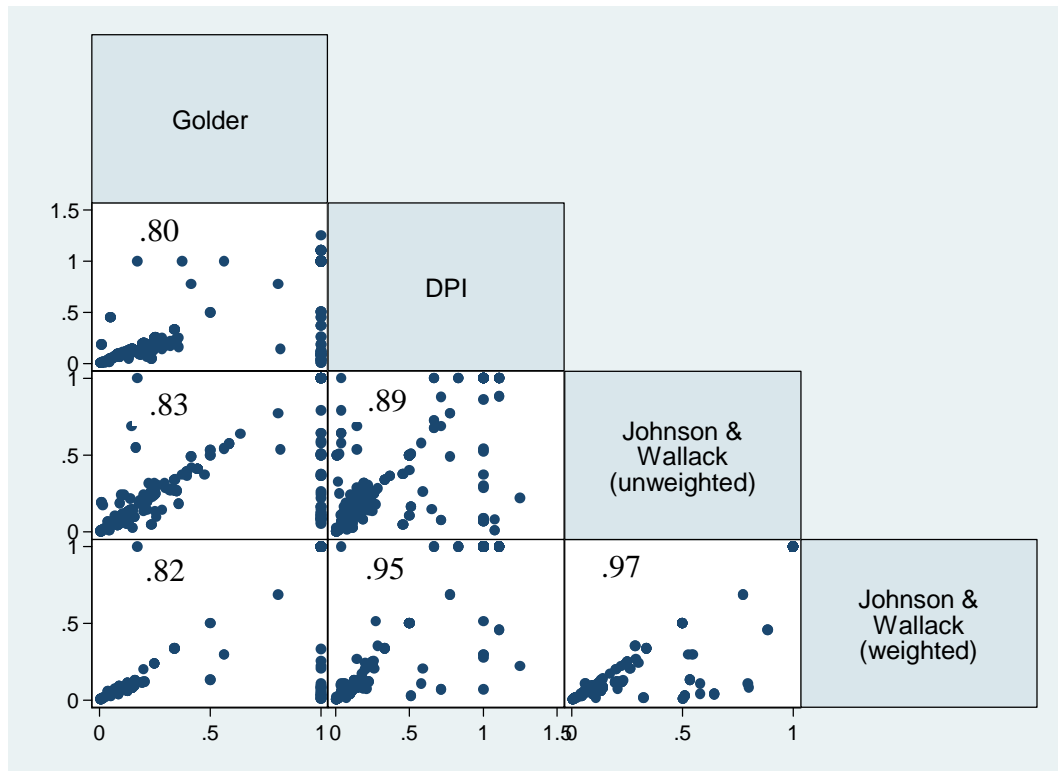
References

Amorim Neto, Octavio and Gary Cox (1997) "Electoral Institutions, Cleavage Structures, and the Number of Parties", *American Journal of Political Science* 41(1): 149–174.

Beck, Torsten et al. (2001) "New Tools in Comparative Political Economy", *World Bank Economic Review* 15(1): 165–176.

Blais, André (1988) "The Classification of Electoral Systems", *European Journal of Political Research* 16(1): 99–110.

Blais, André and Louis Massicotte (2002) "Electoral Systems", in LeDuc, L., R. Niemi, and P. Norris, eds. *Comparing Democarices 2*. London. Sage.

Carey, John ad Mathew Soberg Shugart (1995) "Incentives to Cultivate a Personal Vote", *Electoral Studies* 14(4): 417–439.

Chang, Eric and Miriam Golden (2006) "Electoral Systems, District Magnitude and Corruption", *British Journal of Political Science* 37: 115–137.

Cheibub, Antonio and Jennifer Ghandi (2004) "Classifying Political Regimes", Paper presented at he Annual Meeting of the American Political Science Association.

Clark, William Roberts and Matt Golder (2006) "Rehabilitating Duverger's Theory: Testing the Mechanical and Strategic Modifying Effects of Electoral Laws", *Comparative Political Studies* 39(6): 679–708.

Clark, William Roberts and Matt Golder (2007) "Replication data for: Rehabilitating Duverger's Theory", hdl:1902.1/10477.

Cox, Gary (1997) *Making Votes Count.* Cambridge: Cambridge University Press.

Duverger, Maurice (1954) *Political Parties*. New York: Wiley.

Farrell, David (2001) *Electoral Systems.* Houndmills: Palgrave.

Gallagher, Michael and Michael Marsh, eds. (1988) *Candidate Selection in Comparative Perspective*. London: Sage.

Golder, Matt (2005) "Democratic Electoral Systems around the World", *Electoral Studies* 24: 103–21.

Herrera, Yoshiko and Davesh Kapur (2007) "Improving Data Quality", *Political Analysis* 15(4): 365–286.

Huber, Evelyn, Charles Ragin, John Stephens, David Brady, and Jason Beckfield (2004) *Comparative Welfare States Data Set*. Northwestern University, University of North Carolina, Duke University and Indiana University.

Johnson, Joel and Jessica Wallack (2006) "Electoral Systems and the Personal Vote", mimeo., San Diego: University of California.

Kaufman, Daniel, Art Kraay, and Pablo Zoido-Lobaton (1999) "Aggregating Governance Indicators", World Bank Working Paper 2195, New York.

Keefer, Philip (2005) "DPI2004 Database of Political Institutions", mimeo., The World Bank Development Research Group.

Kunikova, Jana and Susan Rose-Ackerman (2005) "Electoral Rules and Constitutional Structures as Constraints on Corruption", *British Journal of Political Science* 35: 573–606.

Lijphart, Arendt (1994) *Electoral Systems and Party Systems*. Cambridge: Cambridge University Press.

Lust-Okar, Ellen (2006) "Elections under Authoritarianism", *Democratization* 13(3): 456–71.

Maddison, Angus (2003) *The World Economy*. Paris: OECD Development Center.

Massicotte, Louis and André Blais (1999) "Mixed Electoral Systems", *Electoral Studies* 18(3): 341-366.

Massicotte, Louis, André Blais, and Antoine Yoshinaka (2004) *Establishing the Rules of the Game: Election Laws in Democracies*. Toronto: University of Toronto Press.

Mozaffar, Shaheen and Richard Vengroff (2002) "A 'whole systems' approach to the choice of electoral rules in democratizing countries", *Electoral Studies* 21(4): 601–616

Nohlen, Dieter, Bernard Thibaut, and Michael Krennerich, eds. (1999) *Elections in Africa*. Oxford: Oxford University Press.

Nohlen, Dieter, Florian Grotz, and Christof Harman, eds. (2001) *Elections in Asia and the Pacific*. Oxford: Oxford University Press.

Nohlen, Dieter, ed. (2005) *Elections in the Americas*. Oxford: Oxford University Press.

Norris, Pippa (2004) *Electoral Engineering*. Cambridge: Cambridge University Press.

Persson, Torsten and Guido Tabellini (2003) *The Economic Effects of Constitutions*. Cambridge: The MIT Press.

Persson, Torsten, Guido Tabellini, and Francesco Trebbi (2003) "Electoral rules and corruption", *Journal of the European Economic Association* 1(4): 958–989.

Posusney, Marsha Pripstein (2002) "Multi-Party Elections in the Arab World", *Studies in Comparative International Development* 36(4): 34–62

Powell, Bingham (2000) *Elections as Instruments of Democracy*. New Haven & London: Yale University Press.

Przeworski, Adam, Michael Alvarez, José Antonio Cheibub and Fernando Limongi (2000) *Democracy and Development*. Cambridge: Cambridge University Press.

Rae, Douglas (1967) *The Political Consequences of Electoral Laws*. New Haven: Yale University Press.

Reynolds, Andrew, Ben Reilly, and Andrew Ellis (2005) *Electoral System Design*. Stockholm: International Institute for Democracy and Electoral Assistance.

Scott, William (1955) "Reliability and content analysis: the case of nominal scale coding", *Public Opinion Quarterly* 19: 321–325.

Shugart, Mathew Soberg and Martin Wattenberg (2001) "Mixed-Member Electoral Systems", in Shugart M. and Wattenberg, M., eds. *Mixed-Member Electoral Systems*. Oxford: Oxford University Press.

Taagepera, Rein and Mathew Soberg Shugart (1989) *Seats and Votes*. New Haven: Yale University Press.

Teorell, Jan, Sören Holmberg and Bo Rothstein (2008) "The Quality of Government Dataset," version 7Jan08. University of Gothenburg: The Quality of Government Institute, http://www.qog.pol.gu.se.

Treisman, Daniel (2007) "What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?", *Annual Review of Political Science* 10: 211–44.

Wallack, Jessica Seddon, Alejandro Gaviria, Ugo Panizza, and Ernesto Stein (2003) "Political Particularism around the World", *World Bank Economic Review* 17(1): 133–143.

Figure 1. Concurrence in measuring (inverse) average district magnitude

Note: Entries are the Pearson's correlation coefficients for each pairwise comparison. Observations (election years) are included if either source codes a year as an elections year. Number of observations range from 263 to 582. All estimates are from the Quality of Government Dataset (Teorell et al. 2008).

Table 1. Coverage of Datasets on Electoral Systems

	Golder	DPI	Persson & Tabellini	Johnson & Wallack
Dimensional coverage:				
Variables on ballot structure	2	1	2	23
Variables on district structure	5	6	3	14
Variables on electoral formula	6	6	1	1
Total number of variables	13	13	6	38
Empirical coverage:				
Number of countries	124	165	61	127
Mean number of countries/year	52	114	56	81
Time period	1946-2000	1975-2004	1960-1998	1978-2005
Mean number of years/country	23	21	36	18
Number of country-years	2831	3423	2179	2267

Note: Entries with respect to empirical coverage pertain to the following variables in the Quality of Government Dataset (Teorell et al. 2008): gol_est (Golder), dpi_pluralty (DPI), pt_maj (Persson & Tabellini), and jw_persr (Johnson & Wallack).

Table 2. Global Representativeness of Datasets on Electoral Systems

	Golder	DPI	Persson & Tabellini	Johnson & Wallack	Globally
GDP/capita (G-Kh. $'s)	7822	6518	7950	8378	4510
HDI (0 - 1)	.77	.71	.76	.76	.67
Population (1000')	32 307	29 298	34 321	32 746	24 186
Polity score (–10 - +10)	7.9	3.4	6.1	6.3	–.11
FH Civil Liberties (1 - 7)	2.2	3.4	2.4	2.6	3.9
FH Political Rights (1 - 7)	1.7	3.2	2.1	2.4	3.9
Proportion dictatorships	.02	.41	.23	.18	.58

Note: Entries are mean values for all valid observations of the variables mentioned in the table note of Table 1. Data on GDP/capita (in Geary-Khamis dollars) and population are from Madison (2003). HDI is the Human Development Index from the UNDP. Proportion dictatorships is based on the classification by Cheibub and Ghandi (2004). All estimates are from the Quality of Government Dataset (Teorell et al. 2008).

Table 3.Concurrence in Classifying Electoral Formulas

	Observed concurrence	Expected concurrence	Scott's π	No. of elections
Majority/Plurality:				
Golder–PT	.94	.56	.86	471
DPI–Golder	.92	.56	.81	439
Unicameral: DPI–Golder	.97	.55	.93	340
DPI–PT	.93	.57	.83	321
Unicameral: DPI–PT	.97	.55	.94	240
PR:				
DPI–Golder	.84	.49	.69	434
Unicameral: DPI–Golder	.90	.50	.80	335
d'Hondt:				
DPI–Golder	.75	.50	.49	216
Mixed:				
DPI–Golder	.79	.67	.36	433
Unicameral: DPI–Golder	.91	.79	.56	334
Compensatory upper tier:				
Golder–JW	.94	.51	.87	31

Note: Entries are the proportion of observations (election years) that concur in their classification across each pair of data source. Concurrence occurs both when both sources code a zero (0) and a when both sources code a one (1). Observations are included if either source codes a year as an elections year. PT refers to Persson and Tabellini, JW to Johnson and Wallack. All estimates are from the Quality of Government Dataset (Teorell et al. 2008).

Table 4. Replication Estimates I: District Magnitude and the Effective Number of Parties

Regressor	Whole Sample				Established Democracies			
	(1) CG	(2) DPI	(3) CG	(4) JW	(5) CG	(6) DPI	(7) CG	(8) JW
ElectoralParties	0.56***	0.67***	0.61***	0.64***	0.61***	0.62***	0.65***	0.65***
	(0.08)	(0.07)	(0.09)	(0.07)	(0.07)	(0.07)	(0.09)	(0.08)
ln(Magnitude)	–0.07	0.10	–0.02	0.02	–0.05	–0.06	–0.01	–0.01
	(0.09)	(0.11)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.08)
ElectoralParties × ln(Magnitude)	0.07**	0.02	0.06**	0.04	0.06**	0.06**	0.05*	0.05*
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
UppertierSeats	0.02*	0.02*	0.02**	0.02**	0.01	0.01	0.02	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
ElectoralParties × UppertierSeats	–0.00	–0.01	–0.00	–0.01*	0.00	–0.00	–0.00	–0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.59**	0.30	0.45*	0.38*	0.45**	0.44**	0.31	0.33
	(0.25)	(0.21)	(0.25)	(0.20)	(0.21)	(0.21)	(0.24)	(0.23)
Observations (election-years)	339	339	359	359	289	289	291	291
Adjusted R-squared	0.84	0.83	0.86	0.85	0.90	0.89	0.89	0.88

* significant at the .10-level, ** significant at the .05-level, *** significant at the .01-level.

Note: Entries are regression coefficients, with robust standard errors clustered by country within parentheses. Grey areas indicate that the original data (CG) is being replaced: by the DPI or by Johnson and Wallack (JW). We exactly follow the specification of Clark and Golder (2006), Table 1, "Pooled Analysis".

Table 5. Replication Estimates II: Ballot Structure, District Magnitude, and Corruption

Regressor	(1) PTT	(2) DPI	(3) PTT	(4) JW	(5) PTT	(6) Golder
PINDP	−1.48**	0.24	−1.17*	−0.06	−1.64**	−1.28**
	(0.62)	(0.45)	(0.61)	(0.59)	(0.65)	(0.49)
MAGN	2.07***	−0.25	1.68**	0.40	2.30***	1.66***
	(0.77)	(0.57)	(0.75)	(0.75)	(0.80)	(0.52)
Observations	80	80	76	76	67	67
Adj. R2	0.85	0.83	0.86	0.85	0.87	0.88

* significant at the .10-level, ** significant at the .05-level, *** significant at the .01-level.

Note: Entries are Weighted Least Squares regression coefficients, with standard errors within parentheses, and the inverse of the estimated error variance in the corruption perceptions measure used as weight. Grey areas indicate that the original data (PTT) is being replaced: by the DPI, Johnson and Wallack (JW), or Golder. We exactly follow the specification of Persson, Tabellini and Trebbi (2003), Table 2, including income, age, the Freedom House scores, level of education, a dummy for OECD countries, trade openness, ethno-linguistic fractionalization, and the share of protestants and Confucians as controls.