Rationalizing Regime Classification:
A Structured Review of the ACLP Political and Economic Database

Åse Gilje Østensen
The Norwegian Social Science Data Services
ase.ostensen@nsd.uib.no

Paper prepared for the ECPR Joint Sessions: 11 - 16 April, 2008, Workshop 17, The Numbers We Use, the World We See: Evaluating Cross-National Datasets in Comparative Politics
1 Introduction

The study of democracy and democratization has constituted one of the most important areas of research in political science for several decades, especially since the start of the “third wave” of democratisation in the mid-1970s (Munck and Snyder 2007: 9). As part of this research tradition, several instruments for measuring the extent of democracy have been developed, and indices of democracy are now widely used to examine the causes and consequences of democratization. Most existing democracy indicators have been challenged for the way they define, operationalize and measure central political concepts, and there is little academic agreement to the preference of any one democracy index. The choice of such an index should hence be subject to close consideration, and users should take note of several issues that may affect their comparability and quality, and ultimately the interpretations that stem from the analysis.

This paper focuses on the ACLP Political and Economic Database, which original version was compiled by Alvarez, Cheibub, Limongi, and Przeworski (1996) and remain among the most widely used democracy indices in the field. As pointed out by Sousa, however, extensive use is not necessarily a guarantee of validity or reliability, and a closer examination of the ACLP is thus warranted (Sousa 2003: 28). This paper offers an assessment of key features of the ACLP Political and Economic Database and discusses the particularities of the dataset that make up the strengths and weaknesses of this particular resource. The following section of this paper provides a brief descriptive account of the dataset and introduces its most important conceptualizations and operationalizations, section 3 discusses these features in light of the central polemic on democracy indices, the forth section briefly considers other variables and sources that make up the dataset, and section 5 addresses other methodological traits. Finally, section 6 draws some conclusions and seeks to sketch out some of the substantive implications derived from the use of the ACLP index.
2 Description

The ACLP Political and Economic Database was originally produced by Michael Alvarez, José Antonio Cheibub, Fernando Limongi and Adam Przeworski for the project Democracy and Development, a project which sought to examine the emergence and durability of democratic regimes.

2.1 Coverage

The dataset includes a sample of 135 states observed between 1950 (or the year of independence or the first year when economic data are available) and 1990 (or the last year for which data are available), a total of 4,318 observations. The unit of analysis is a given country during a particular year. The team found 224 regimes, 123 that did not qualify as democracies, and 101 that did. The project further observed 50 transitions to democracy, and 40 transitions to dictatorship (Przeworski, Alvarez, Cheibub and Limongi 1996: 39).

The original dataset contains 105 variables (including lagged variants and country identifiers). The core of the dataset is constituted by the regime variables, which classify political regimes according to several criteria. The most important variable is a dichotomous measure of democracy, which will be paid particular attention in this paper. In addition, the dataset contains economic indices, social variables, religious measures and other political variables. The countries covered by the original dataset include 28 European, 48 African, 29 American, 23 Asian, and seven countries belonging to Oceania.

In 2004 Cheibub and Gandhi launched an updated version of the dataset, which covered more countries (199) and extended the temporal coverage (1946-2002). This version thus contains a total of 7,880 country-years of which 58.5 percent were classified as dictatorships and 41.5 percent were democracies (Cheibub and Gandhi 2004: 2).
The original version of the dataset is available online, free of charge. The Cheibub-Gandhi update is not available online, but some of the regime variables can be found in other freely available datasets, such as the Quality of Government Institute’s QoG Time-Series Data and Pippa Norris’s Democracy Indicators Cross-National Time-Series Dataset (Rydland et al. 2008: 107).

2.2 Conceptualization

The concept of democracy remains a contested notion and there is disagreement over its definition, content, measurement and input to measurement. Illustrative of this lack of consensus, the European Union consistently avoids defining the term (Landman and Häusermann 2003: 2). While several authors stress the multidimensional character of democracy (Coppedge, Alvarez and Maldonado 2007, Hadenius and Teorell 2005 etc.), Przeworski, Alvarez, Cheibub and Limongi advocate a minimalist definition of the concept and deliberately focus on the smallest possible number of attributes that are still seen as giving a viable semantic standard for the term (Collier and Levitsky 1997: 433). According to the authors, democracy has too often been taken to encapsulate “almost all normatively desirable aspects of political life, and sometimes even social and economic life” (Przeworski et al. 2000: 14). The ACLP definition of democracy thus depends on the mere presence of certain institutions, without taking their particular effects into account (Cheibub and Gandhi 2004: 6). Such a definition draws on Dahl’s classic *Polyarchy*, and is closely related to definitions used by several other indices of democracy (Przeworski et al. 1996: 39). Democracy is defined simply as a regime in which “governmental offices are filled as a consequence of contested elections”. Only if the opposition is allowed to compete, win, and assume office is a regime democratic (ibid: 50). The definition thus consists of two elements: offices and contestation.

Offices:

The definition is mainly concerned with the offices of the chief executive and the seats in the effective legislative body and does not imply that all offices in a regime should be filled as a consequence of elections (ibid).
Contestation:
“Contestation occurs when there exists an opposition that has some chance of winning office as a result of elections” (ibid). Przeworski et al. further identify three core features of the concept of contestation: (i) *ex ante* uncertainty, (ii) *ex post* irreversibility, and (iii) repeatability. *Ex ante* uncertainty refers to the probability that the incumbent party, coalition, or at least a member of the ruling coalition can lose office as a result of an election. According to the authors then, uncertainty is not to be confused with unpredictability as the outcome of elections frequently can be predicted to some extent, e.g. by opinion polls. The term refers to the mere *possibility* that the incumbent party could lose (ibid). The second feature of contestation, *ex post* irreversibility, covers the assurance that any one party winning an election will be allowed to assume office. Finally, “repeatability” assumes that the winning faction cannot “use office to make it impossible for the competing political forces to win next time” (ibid).

The authors distinguish three types of democracies based on the rules that specify who may dismiss the government and classify democracies as parliamentary, mixed or presidential regimes (ibid:52). The updated version of the dataset relies on the original definition of democracy, but in addition to the threefold democratic classification, also differentiates among three types of dictatorships: monarchical dictatorships, military dictatorships and civilian dictatorships (Cheibub and Gandhi 2004: 23). Yet, the new classification still differentiates sharply between the two main types of regimes (ibid: 19).

### 2.3 Operationalization and coding

The ACLP authors stress that democracy and dictatorship are mutually exclusive categories, and democracy is an “all or nothing affair”. Initiatives that code countries in this way have a minimum threshold for classifying a country as a democracy. Przeworski et al. (2000:18-29) apply a simple set of four decision rules, all of which must be met in order to categorize a regime as democratic:

1. The chief executive must be elected.
2. The legislature must be elected.
3. There must be more than one party competing in elections.
4. An alternation in power under identical electoral rules must have taken place (Summarized in Cheibub and Gandhi 2004: 3).

If a regime fails on any of these four terms, it is subsequently coded as a dictatorship or autocracy (used interchangeably). While the first three requirements are relatively straightforward, rule four is somewhat more complicated. As pointed out by the authors, it is difficult to distinguish regimes where incumbents only hold elections because they know they will not lose them, from the regime where incumbents never lose power due to their popularity. Although these two categories are substantially very different, they are observably equivalent and it is thus merely impossible to distinguish among them. Cheibub and Gandhi 2004 take Botswana as an example of a case that is impossible to code according to the above rules. However, instead of excluding such cases, they are coded as dictatorships, but identified through a variable called “Type II”. In this way users have the cases of doubt singled out, and can recode these cases as democracies or choose to exclude them from the analysis (ibid: 4). As a result of these regimes where lack of information obscured the coding process, new information can lead the coding to change over time. Using the four rules, Cheibub and Gandhi 2004 assert that both versions of the dataset have unambiguous codes for 92 percent of the observations, i.e. eight percent of the observations were classified as dictatorships based solely on the fourth rule.¹

The coding rules for the classification of democracies can be summarized in a sequence of three questions or steps: Is there an independently (directly or indirectly) elected president? (Identifies parliamentary systems). Is the government responsible to the assembly? (Identifies presidential systems). Is the government responsible to the president? The classification is derived from the rules prescribed in the country’s constitution (ibid:16-17) and summarized in figure 1.

¹ This does not mean that there has not been recodings as new information became available, merely that the amount of undistinguishable cases were the same in the two versions (Cheibub and Gandhi 2004: 4).
Figure 1. The ACLP procedure for classification of democratic regimes

(Cheibub and Gandhi 2004: 27)

These subtypes of democracy represent a move down Sartori’s “ladder of generality”, meaning they have more defining attributes than the overarching concept (here “democracy”). The subtypes are nevertheless definitely fully democratic, they represent types of democracies, but not different degrees of democracy (Collier and Levitsky 1997: 435).

Similarly, Cheibub and Gandhi discern dictatorships according to the institution likely to remove the dictator from power. Regimes where the executive’s power basis is family and kin networks are classified as monarchies, in military dictatorships the
dictator relies on the armed forces, while all other dictatorships are classified as civilian (ibid: 23).

3. Contested minimalism

Whether democracy and autocracy should be understood and measured along a single continuum, or treated as isolated phenomena is a, as mentioned, a hotly debated issue which have spurred numerous arguments from both camps. The overall rationale for the dichotomy that Przeworski, Alvarez, Cheibub, and Limongi advocate is based on both a generic claim about how democracy should be conceptualized and on a more specific claim about measurement and the empirical distribution of cases (Collier and Adcock 1999: 549). This paper discusses these claims in that same order.

3.1 Conceptualization

Alvarez et al. cite Robert Dahl’s 1971 classic *Polyarchy* in establishing their theoretical base for their indicator. Dahl’s definition of democracy focuses on two dimensions: (i) Contestation and (ii) Participation (Dahl 1971: 1-6). Munck and Verkuilen identify the recognition of these two dimensions of democracies as the “first step in the construction of an index” (2002). However as seen above, ACLP diverges from Dahl’s definition by excluding participation as a direct input to their indicator. They base the exclusion on e.g. the rationale that they are concerned with gathering data for the Post World War II period, and that suffrage is to be taken for granted in this period (Alvarez et al. 1996: 5). As pointed out by Munck and Verkuilen however, there exist a range of informal potential restrictions to voting, which all are ignored in the coding of the ACLP dataset (2002: 11).
Munck and Verkuilen are, however, far from the only scholars that have expressed concern with the validity of the ACLP’s narrow definition of democracy. Hadenius and Teorell e.g. denote the democracy criteria as insufficient, and argue that the index omits important aspects of the concept by focusing solely on the electoral aspects of democracy such as political freedoms, thus leaving out aspects of liberal democracy (Hadenius and Teorell 2005: 22). Other scholars have expressed concern about the lack of reference to qualitative aspects of government (Mainwaring, Brinks and Pérez-Liñan), or of its failure to adequately distinguish between subtypes of democracy (O’Donnel 1995).

Such criticism is, nevertheless, far from idiosyncratic to the ACLP index and both Vanhanen and Polity receive criticism on similar terms. Freedom House in contrast, receive partial praise for its degree of construct validity, although this apparently can be considered more of a maximalist approach including components, which appear redundant (Hadenius and Teorell 2005: 17). Foweraker and Krznaric, claims that all key democracy indices to some degree suffer from the a narrow institutional focus which neglects aspects such as protection of civil rights, property rights, and minority rights (Foweraker and Krznaric 2000: 774-776).

The conceptual argument further hinge on the definition’s compatibility with most research topics concerning regimes in some way. Whether regimes are treated as a variable in studies concerning economic development or conflict/ war, the occurrence of contested elections, it is argued, is the mechanism that links political regimes to the respective outcomes (Cheibub and Gandhi 2004: 7). The authors also explain in detail the conceptual rationale for excluding social and economic aspects, accountability,
responsibility, responsiveness, representation, political freedom, civil rule over the military (see Przeworski et al. 2000: 33-36). In the case of the exclusion of participation, perhaps the most contested omission, Przeworski et al. stress their concern with identifying regimes in which “some, but not necessarily all, conflicting interests contest elections” and do see suffrage as a guarantee of this (Przeworski et al. 2000: 34).

Furthermore, the authors also confirm to have more pragmatic or methodological reasons to adopt a minimalist definition. One reason for excluding participation from their definition thus relies on their intention to test theories about the effects of participation on the durability of the democracy (ibid), an analysis which necessarily needs to treat the concepts separately. Altogether, the argument that the more dimensions that are considered in classifying political regimes, the more difficult it is to distinguish the causal mechanisms that link regime and the outcome studied, appears to bear quite some weight in the choice of the minimalist approach. The authors prefer to study empirically the relationship between democracy and economic growth, social and economic equality, effective exercise of political rights, widespread participation, control by citizens over politicians etc., rather than to have it decided definitionally (ibid: 36). This way, the criticism that points to ignorance of qualitative aspects of regimes is answered much by turning this aspect into an asset since a procedural definition will avoid blurring the boundaries between political regimes and other political phenomena, political entities or aspects of the state (Cheibub and Gandhi: 8, Muck and Verkuilen 2002: 54) and enable extensive empirical analysis of causal relationships. Such a definition, it is argued, will also
have the benefit of being compatible with different variants of social and political organization (Cheibub and Gandhi 2004: 7).

Moving beyond the face definition of democracy, a concept can typically be disaggregated to a number of attributes and components with different level of abstraction (Munck And Verkuilen 2002). In order to assure meaningful operationalizations, and in turn assist in measurement and aggregation procedures, the different attributes of the term should be clearly stated. However, in order to avoid a mere “checklist of vaguely related aspects of democracy”, the attributes should also be organized vertically according to their levels of abstraction and their relation to the overarching concept. Muck and Verkuilen thus introduce a hierarchically structured approach to such organization, which they term “a concept tree”.

**Figure 2. The concept tree format applied to the ACLP’s definition of democracy**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Democracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
<td>Contestation</td>
</tr>
<tr>
<td>Components of attributes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ex ante</th>
<th>Ex post</th>
<th>Repeatability</th>
</tr>
</thead>
<tbody>
<tr>
<td>uncertainty</td>
<td>irreversibility</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 displays the ACLP definition of democracy structured logically according to its level of abstraction. All attributes and their components are easily distinguishable and there are no redundant elements. Munck and Verkuilen criticise other democracy
indices for including too many attributes, components and subcomponents with little thought to the relationship among them (Freedom House), and yet others for confusing which components relate to which attribute (Bollen 1980, and Arat 1991). The simplicity of minimalist definitions like the ACLP one, are in contrast praised for their logical organization of attributes (Muck and Verkuilen 2002: 14).

3.2 Disputed measurement

“There is no measurement without error”

- Przeworski et al. (2000: 35)

While Alvarez et al. lean on Dahl’s basic definition of democracy, they diverge from Dahl on important aspects of his assumption of democracy as a theoretical ideal type of which regimes approach to a varying degree. Dahl in his conceptualization takes it as a conceptual given that there is “an underlying, hypothetical continuum that extends from the greatest to the least opportunities for oppositions” (Dahl 1971: 231). Alvarez et al. in this respect draws on Sartori’s rationale that political systems are “bounded wholes”, and thus characterized by constitutive mechanisms and features that are either present (albeit imperfectly) or absent (albeit imperfectly) (Sartori 1987: 184).

Prominent scholars engaged in comparative research on democratic regimes are in sharp disagreement over the choice between a dichotomous or graded approach to the distinction between a democracy and a non-democratic regime. As pointed out by Cheibub and Gandhi, among those that study regimes empirically, this is probably the one debate that has generated the highest level of controversy, with most views being
inclined towards preferring polychotomous approaches (2004: 10). Collier and Adcock stress the importance of the choice by pointing to both a substantive concern and a methodological concern:

“The choice is substantively important because it affects the findings of empirical research. It is methodologically important because it raises basic issues concerning appropriate standards for justifying choices about concepts” (Collier and Adcock 1999: 537).

To the authors of ACLP and the updated version, there is simply no choice to be made since they understand the concept categories as in essence mutually exclusive. The authors thus argue that insisting on that there are borderline cases that require a graded approach is, in their view, “ludicrous”, because in a carefully applied classification of regimes the ambiguous status of a case is due to bad rules or insufficient information (Przeworski et al. 1996: 52). If a regime is not clearly distinguished it is therefore not a flaw pertaining to the dichotomy, rather to the scoring procedures (ibid).

As pointed out by Collier and Adcock, however, the justification provided by the authors does not fully explain why a system that holds competitive elections for the executive, maintain rotation in this office, and have more than one party - but lacks elections for the legislative, would not be justified as partially democratic (Collier and Adcock 1999: 549). One would perhaps expect a further justification for why each of the component parts according to Alvarez et al. should be understood dichotomously, and not as part of a gradable whole. The authors do however not rule out grading of democracies on the whole, however, they advocate the two-step procedure introduced by Sartori which entail that first one need to distinguish the democracies from the
dictatorships (dichotomously), then one can perform comparisons of the regimes falling into the democracy category (Sartori 1987: 182-83).

While it is recognized that democracies may differ as to how democratic they are, according to Cheibub and Gandhi, it makes little sense to speak of positive levels of democracy in dictatorships, which frequently will be the result when deploying continuous measures. For example, according to Bollen’s measure, the level of democracy in Albania in 1950 and 1955, under the communist regime of Hoxa was about 24 out of 100, while Chile under Pinochet between 1974 and 1980 scored –7 on a scale from –10 to 10 on the Polity measure etc. (Cheibub and Gandhi 2004: 11). Treating democracy as an attribute that can be measured over the spectrum of cases will thus produce what Cheibub and Gandhi describe as “absurd assertions” (ibid). Pursuing this logic, one could conversely imagine the scores of Sweden or Switzerland on a dictatorship scale.

Arguments for the graded approach have often been made in circumstances where there exists no clear gap between democracy and non-democracy, either across national units or over time (Hadenius and Teorell 2005). In other cases where such a gap is observed and where the empirical observations cluster, a dichotomy may provide an adequate summary of the empirical contrasts among the regimes studied. In such cases the dichotomous approach would arguably be good enough (Collier and Adcock 1999: 554). According to Cheibub and Gandhi, however, this argument would in principal deem a dichotomy appropriate in most cases since the distribution of most polychotomous measures, such as Freedom House and Polity are actually bimodal, with high concentration of cases in their low and high ends (Cheibub and
Gandhi 2004: 13). A high number of categories thus add little information in the case of a highly bimodal distribution. In fact, Cheibub and Gandhi argue that polychotomous indices in general do not convey much more information than dichotomous ones due to their often vague and subjective conceptualizations and arbitrary construction of measures. Many of the standards-based scales of democracy are for example not entirely explicit if a move from 1 to 2 is the same as a move from 3 to 4; or how qualitative information yields the score that is assigned (Landman and Häusermann 2002: 10). Resulting, researchers are often required to collapse regimes into “democratic” and non-democratic’ in order to make meaningful interpretations (Cheibub and Gandhi 2004: 12).

However, if the choice of approach can be justified on the basis of the empirical distribution, it can also be argued that a dichotomous approach will be relevant only as long as the distribution is not too skewed. That is, if there is a larger structural change, such as a democratization wave, leaving relatively few regimes in the dictatorship category, it would arguably influence the relevance of the dichotomy. Diamond e.g. makes the argument with reference to Latin America that the shift to democracy has made it “more fruitful to view democracy as a spectrum, with a range of variation in degree and form” (Diamond 1996, in Collier and Adcock 1999: 555).

Przeworski et al also makes the argument about the empirical distribution based on measurement error. While the authors acknowledge that a dichotomy will produce potentially larger errors than a polychotomous measure, the dichotomy will produce fewer errors than the graded alternative. Furthermore, Przeworski at al. assert that the distribution of democracy versus nondemocracy tends to be “u-shaped”, meaning that the advantage of a dichotomy in terms of reducing error is even greater than it is for a
uniform distribution (Przeworski et al. 1996: 52). The authors hold that it is easier to sort cases the fewer the categories, and that a dichotomous measure will be more reliable than a graded measure even if the true nature of democracy would be continuous (ibid). Such an argument is challenged by Elkins who carried out simulations intended to test the reliability of dichotomous versus graded approaches and found that the reliability of the dichotomy is not as superior as Przeworski et al. claims. Elkins questions their logic based on diverging conceptions of error, or rather on the ACLP authors’ formulation of error, which he deems “incomplete” (Elkins 2000: 299). In calculating expected error with (i) the probability of an error of a certain magnitude, (ii) the magnitude of the error and (iii) the number of such errors, Elkins claims that since the authors do not take total variance into account, this is in essence a measure of error variance. In order to determine the reliability of a measure, it is essential according to Elkins to compare the error variance to the total variance since a polychotomous measure will almost always produce more total variance as well as error variance (ibid).

4 Data sources

The political regime indicators form the core of the dataset, and also surely the most often discussed ones. However, the ACLP include a broad range of additional variables that also merit attention.

There are 23 economic indices (e.g. annual economic growth, rate of inflation, and central government expenditures), 21 social variables (e.g. birth and death rates, the number of riots in a given year, ethno-linguistic fractionalization, and the level of education of the labour force) and four religious ones (indicators of percentage of Catholics, Moslems and Protestants). There are also other political variables and a
number of path-dependence indicators, which try to capture different aspects of the institutional history of the regimes, e.g. type of civil military relations, the mode of executive and legislative election, and the percentage other democracies in the region/world.

Most of the economic variables were gathered from widely used sources including the World Bank, the IMF, the Penn World Tables, while most social, political and path dependence variables were compiled from sources such as Banks *Handbook of Political Indicators*, Singer and Small’s *Correlates of War* dataset and Keesings *Contemporary Archives*. By compiling data from these sources, the ACLP is not the origin of any potential bias or measurement error pertaining to the majority of its variables. However, in examining the overall quality of the dataset these indicators should not be ignored. In fact, Bowman, Lehoucq and Mahoney (2005) argue that the use of inaccurate, partial or misleading secondary sources threatens the validity of existing over-time indices of democracy (2005: 940).²

Problems of comparability and data quality affect most kinds of data, though some statistics are more susceptible than others. Problems often affect the quality of socio-economic statistics in areas such as income inequality, social indicators, education, demography, poverty and inequality, and social welfare spending. According to de la Fuente and Doménech (2006), poor quality of education data may distort analyses to such a degree that education variables frequently turn out to be insignificant or to have the “wrong” sign in growth regressions (Fuente and Doménech in Rydland et al. 2008: 6). In addition, although sources such as the World Bank and the Penn World

² The authors specifically study the Gasiorowski index (1996), the Polity IV and the Vanhanen index focusing on five Central American countries between 1990 and 1999.
Tables mainly contain data from official registers, administrative records and national accounts, users should keep in mind that the data quality vary greatly across countries and that especially developing countries face a number of problems in providing reliable statistics, since under-investment in national statistical systems often results in data of poor quality, and that this would have implications for the true comparability of the data (Rydland et al. 2008: 294-296). Perhaps as a testimony of this problem, about 23 percent of the observations (mostly from developing countries) lack information in the ACLP dataset (Sousa 2003: 28). Sources such as the World Bank and the Penn World Tables, although not perfect in their reliability, probably constitute the best alternative available and as such do not make the ACLP any less reliable than other similar sources. However, there are other data in the ACLP that can be questioned from a validity-concerned point of view. As pointed out by Sousa, the widely used Ethno-Linguistic Fractionalization index (ELF60) collected from Easterly and Levine (1997) has been criticised for not having incorporated important theoretical advances, a fact which could serve to compromise the validity of the measure (ibid).

5 Other methodological concerns

5.1 Validity and reliability

The most common claim for the validity of democratic performance is based on high statistical correlations with other such measures. Przeworski et al. also point to the high correlations with measures such as the Coppenedge-Reinecke scale, the Bollen scale, the Gurr scales and the Gastil scale. While most democracy indices vary considerably concerning conceptualizations, measurement and aggregation, they still tend to correlate highly in aggregates. One reason for this is that they, as mentioned
above, to a great extent rely on the same sources and in some cases also on the same precoded data (Munck and Verkuilen 2002: 29). Thus, high correlations may mean that biases are also reproduced in most indices, and do not altogether disband concerns of validity.

There is also a range of other reasons why the strong correlations should be interpreted with caution. Hadenius and Teorell find that the apparently strong aggregate level correlations tend to weaken significantly when computed at different levels of the democracy scale, or within a critical zone around the cut points of dichotomous indices (Hadenius and Teorell 2005: 2). Foweraker and Krznaric point to the dubious statistical basis of these high correlations since they generally assume that the data are normally distributed, while most of the datasets in fact are not so distributed (Foweraker and Krznaric 2000: 769). Their other objection rests on most validity tests using interval level data techniques like Pearson’s product-moments correlations on data that are mainly ordinal (ibid). Finally, Foweraker and Krznaric find evidence that comparisons of performance measures sometimes are manipulated in order to boost correlation results, e.g. by adjusting time frames to suit particular purposes (ibid). Even without manipulation, they find that the global tests may obscure regional variations, since the degree of congruence among performance measures for the established Western democracies tends to inflate the overall correlation coefficients (ibid: 769-770).

Reliability of measures is sometimes assessed using tests of inter-coder reliability. These normally set out to demonstrate the coherence in coding produced by multiple coders. The value of such tests is however not undisputed as they merely account for
agreement/disagreement about how cases are to be coded, possibly due to similar biases and not for high levels of validity (Munck and Verkuilen 2002: 18). Consequently, since neither the correlation tests or tests of inter-coder reliability provide a test of the validity of the indices (Munck and Verkuilen 2002: 29), these measures should not necessarily be taken to mean that the choice of index is insignificant even though many indexes appear to produce more or less the same results.

5.2 Replicability

While the demand for replicability may be more closely tied to reliability and validity concerns, and less to a need to reproduce the process of generating data itself, it is nevertheless vital that the opportunity to scrutinize and question data is maintained (Munck and Verkuilen 2002: 18-19). Such a process requires the publication of coding rules, the coding process, and the disintegrated data generated on all indicators (ibid: 19).

As we have seen, the coding rules of the REG variable are clearly stated. However, in the original version, the other regime variables,\(^3\) are not equally well documented. There is little information of the coding of these variables, which affects the potential for their replication. The authors do not provide the sources that were used in constructing these variables, the coding rules, nor do they publicize who coded these variables and whether tests of inter-coder reliability were carried out (Sousa 2003: 28). However, in the updated version by Cheibub and Gandhi, the coding of the six-

\(^3\) These variables include the one classifying type democracies (INST), one that distinguish dictatorships with at least one political party (MOBILIZE), one that distinguish dictatorships according to the number of formal powers (DIVIDED) and one that classifies dictatorships according to whether rule is exercised by laws (AUT) (Przeworski et al. 2000: 31-32).
fold classification is far better documented, although there is no reference to tests of inter-coder reliability.

Nonetheless, on a general basis and compared with other indices, Munck and Verkuilen judges Alvarez et al. to stand out as “having set a high standard what regards reporting their coding and coding process” and thus assuring replicability of their dataset (Munck and Verkuilen 2002: 21).

### 5.3 Aggregation

Existing democracy indices tend to display a rather low level of sophistication concerning their aggregation procedures. Quite a few simply assume that it is acceptable to move up to the highest level of aggregation, i.e. to a one-dimensional index, without much justification or going by specific and consistent aggregation rules (ibid: 27). In an evaluation of nine democracy indices, Munck and Verkuilen found that Alvarez et al. were among the few index constructors to offer appropriate or at least acceptable rationale for their aggregation procedures. They deem a regime democratic only as long as there are positive scores on all three attributes individually (contestation, elected executives, and elected legislature) (Przeworski et al. 1996: 51). The authors do not, however, report to having tested out the implications of any alternative aggregation rules (Munck and Verkuilen 2002: 27).

### 5.4 Missing data

As mentioned above, Sousa asserts that about 23 percent of the observations are lacking information in the dataset. Since the missing data for the most part pertain to developing countries or less developed countries, this could potentially constitute a
problem when creating sub-samples of the data. Consequently, Sousa suggests that both the pattern of missing data and the reasons for them missing be checked by scholars intending to use the dataset. Such measures would be recommendable in order to make sure that the missing information can be safely ignored or that compensating measures can be used without running the risk of biases harmful to hypothesis testing (Sousa 2003: 28).

6 Conclusions

Like most other democracy indices, ACLP focus attention on selected aspects of the democracy concept, while neglecting others. It should be bore in mind then that the definitions employed by the authors do influence the results of causal assessment. Similarly, the pertaining crude measure used to measure limited facets of the phenomenon, also have a bearing upon empirical analysis. In the original version of the dataset, highly heterogeneous regimes are thus lumped together in two broad categories. Resulting from the 2004 update, six categories of regimes are distinguished. This does nonetheless, not remedy the fact that the ACLP’s REG variable is insensitive to the characteristics of partial transition processes (Sousa 2003: 27). Recognizing that there always will be trade-offs between different types of indicators (Landman and Häusermann 2002: 6-7), the ACLP index appears nonetheless appears to be a robust measure of regimes. One of the main advantages of the dataset is simply put that you know what you get. Definitions and coding rules are clear and replicable, meaning that is a resource that provides users with an opportunity to accurately judge its potentials and whether suits the particular research context. Such clarity and transparency can unfortunately not be taken for granted in the universe of democracy indices. Another advantage is of course the substantial
empirical scope of the dataset, which definitely constitutes a valuable contribution to the field of comparative politics.

The ACLP represents a simple, well-documented, and innovative conceptual approach. It furthermore has the advantages of making it easy to find instances of the concept and allowing for the study of numerous empirical questions. Nonetheless, its minimalist definition and dichotomous measure are especially controversial to many scholars who hold a firm conviction that democracy is a highly compound construct that should be measured accordingly in order to tap into its many facets. However, this author embraces the view advocated by Collier and Adcock (1999) that for concepts such as democracy it may be more realistic and useful to aim for an accumulation of knowledge grounded in mutual comprehension among scholars, rather than an absolute standardization (Collier and Adcock 1999: 562).
References


