

Comparing the Data Sets

Online Appendix to Accompany

"Rival Strategies of Validation: Tools for Evaluating Measures of Democracy"

Jason Seawright and David Collier

Comparative Political Studies 47, No. 1 (2014), 112-39

This article has examined four traditions of measurement validation that have been applied to cross-national data sets on democracy. Whereas the main body of this article is organized around the four traditions, this appendix is structured around the resulting insights into the six data sets that have been examined. A substantial literature has assessed democracy measures from diverse perspectives (Munck & Verkuilen, 2002; Munck, 2009), and the discussion below considers only issues of measurement validation. The assessment is also partial, because some indicators have not been evaluated from the standpoint of all four traditions of validation. Nevertheless, the observations below may suggest productive avenues for future work.

Polity Data

The Polity data set (Polity IV) has been evaluated from the perspective of all four traditions. These data are a time series of annual democracy scores starting in 1800 and currently ending in 2011. Polity provides a 20-point summary scale, spanning full authoritarianism (-10) to full democracy (10). For earlier periods, these scores are based on historical records and summaries of regime-related events. Current scores are necessarily based on news reports.

Structural Equations. Treier and Jackman (2008) use an item-response theory model to conclude that “there is considerable error in the latent levels of democracy underlying the Polity scores.”

Levels of Measurement. Munck and Verkuilen (2002) argue that the Polity scale uses an inappropriate aggregation procedure, and thus produces scores that do not fully meet the criteria for ordinal measurement. Treier and Jackman (2008) similarly criticize the Polity aggregation rule.

Pragmatic. Casper and Tufis (2003) show that, relative to competing indicators, the Polity measure is more weakly connected with economic growth and primary education, but more strongly related with lack of economic openness. For the sub-period from 1951 until 1973, the Polity measure appears more strongly related with secondary education than the other indicators. In the last period, between 1975 and 1992, the Polity indicator has a negative relationship with economic growth and with primary education that falls in the middle of the range of estimates. However, it once again shows a more statistically significant negative relationship between economic openness and democracy, compared with competing indicators. Mainwaring, Brinks, and Pérez-Liñán (2003) report that their indicator of democracy in Latin America has a correlation of 0.86 with the Polity scores for the region. Munck and Verkuilen (2002) note that the Polity indicator has demonstrated high levels of intercoder reliability.

Case-Based. Bowman, Lehouq, and Mahoney (2005) criticize many Polity scoring choices for Central American countries. Polity gives Costa Rica a perfect democracy score for every year between 1900 and 1999, despite sixteen coup attempts between 1900 and 1955. The Polity data also code Nicaragua during the 1920s and 1930s as more democratic than other indicators. Berg-Schlosser (2004) argues that the Polity scores neglect “certain broader aspects of social and political reality such as the extent and kind of actual participation or the observance of civil liberties and human rights,” and he also finds a “coding bias favoring an American type of democracy with a strict separation of powers” (p. 253). Examining African countries more closely, Berg-Schlosser argues that the Polity scores for 2000 miss seven probable democracies due to inattention to the actual functioning of institutions, as opposed to formal rules (p. 261).

Freedom House Data

The organization Freedom House publishes an annual report, *Freedom in the World*, that ranks the degree of “freedom”—they basically mean “democracy”—in 192 countries for the period from 1973 to the present. The rankings are given on a 13-point scale, in which democracy is associated with full freedom and authoritarianism with a total lack of freedom. According to the methodology statement on the Freedom House website,¹ the rankings draw on “foreign and domestic news reports, academic analyses, nongovernmental organizations, think tanks, individual professional contacts, and visits to the region.” Each year, these data are coded according to a checklist of 10 questions on political rights and 15 on civil liberties, which are aggregated to produce the final ranking. Each coding decision is reviewed by multiple members of the project team.

Structural Equations. Shen and Williamson (2005), using structural-equation models, find that the Freedom House indicators of political rights, civil liberties, and press freedom have little measurement error; estimates of the variance of measurement error range from 0.06 to 0.08 out of a standardized total variance of 1. By contrast, Bollen and Paxton (2000), based on similar models, find that about 30 percent of the variance of the Freedom House measure of broadcast freedom consists of error (systematic or random). Likewise, about 20 percent of the variance of the Freedom House measures of print freedom and of civil liberties consists of error. At the same time, less than 10 percent of the variance of the political rights measure comes from error. Results from Bollen’s (1993) earlier analysis are essentially identical.

Levels of Measurement. Munck and Verkuilen (2002) argue that the aggregation rule used by Freedom House to convert subscales into an overall democracy score is inappropriate and fails to meet the assumptions of an ordinal scale.

Pragmatic. Drawing on comparisons among regression models, Casper and Tufis (2003) show that, for the period from 1975 until 1992, the Freedom House indicator has a weaker negative relationship with economic growth and with primary education than competing indicators. However, it has a notably stronger positive relationship with presidentialism than other available measures. Mainwaring, Brinks, and Pérez-Liñán (2001) report that their indicator of democracy in Latin America has a correlation of 0.82 with the Freedom House scores for the region.

Polyarchy Data

The Polyarchy data set, created by Tatu Vanhanen, covers all independent countries for the period from 1810 until 2000. The Polyarchy index uses the smaller parties' vote share in legislative and presidential elections as a measure of competition, while the percentage of the population which actually votes in the election is used as a measure of participation. For non-elected regimes, participation and competition are both scored as zero. An overall democracy indicator is constructed by multiplying a country-year's competition and participation scores and dividing the result by 100.¹

Levels of Measurement. Munck and Verkuilen (2002) criticize the multiplicative aggregation rule used for converting subscales into the overall Polyarchy measure of democracy. Vanhanen, these authors argue, offers no theoretical argument for why multiplication should be the correct aggregation rule, or why different subscales should have the same weight in the product. Thus, the resulting scores may not appropriately preserve ordering and relative difference among cases on the subscales.

Pragmatic. Casper and Tufis (2003) find that the Polyarchy indicator has a more pronounced negative relationship with growth rates and with education than other measures

during the period from 1951 until 1992. It has a somewhat less substantial positive relationship with secondary education than other indicators during the period from 1951 to 1973. Finally, during the period from 1975 until 1992, the Polyarchy indicator has relationships with independent variables that are roughly comparable to those of the Polity indicator. Bogaards (2007) criticizes the Vanhanen scale for its direct incorporation of incumbents' margin of victory as an indicator of democratic competition. He shows that electoral results in Africa have a relatively weak relationship with subjective measures of democracy that do not incorporate electoral results, specifically the Polity and Freedom House scales. Munck and Verkuilen (2002) observe pragmatically that the Polyarchy data set has such clear decision rules that it can be perfectly replicated by any interested researcher. Bollen (1980) shows that turnout—a key component of the Polyarchy indicator—is weakly or negatively correlated with several other indicators of democracy, thereby calling into question its appropriateness as part of the overall measure.

Case-Based. Bowman, Lehoucq, and Mahoney (2005) criticize some aspects of the Polyarchy indicator for Central America, while praising others. For example, the indicator codes the percentage of Costa Rican citizens voting in elections prior to 1914 as near zero—even though indirect presidential elections with substantial turnout were held three times during that period. By contrast, the Polyarchy indicator for Nicaragua during the early part of the twentieth century corresponds with the results of a close analysis of country history.

Gasiorowski Data

Gasiorowski's (1996) Political Regime Change Dataset provides democracy scores for the 97 largest developing countries. The time period for each country extends from the year of independence or of the formation of what Gasiorowski classifies as a "modern state" (1996) until

1992. Each country-year is coded as democratic regime, semi-democratic regime, authoritarian regime, and transitional regime. Data used to classify country-years were drawn from Diamond et al.'s (1989) case studies of democracy in the developing world, *Keesing's Record of World Events*, and similar sources (Gasiorowski, 1996).

Pragmatic. Munck and Verkuilen (2002) argue that the Gasiorowski scale suffers from important pragmatic weaknesses in that no attempt at replication by other coders was ever conducted, and may not even be feasible. Thus, dependence on potential idiosyncrasies of the individual coder is impossible to assess.

Case-Based. Bowman, Lehoucq, and Mahoney (2005) note that in the Gasiorowski index, Nicaragua is coded as completely authoritarian throughout the 1920s and 1930s. This corresponds with the available evidence about Nicaraguan elections before 1928, which evidently were purely ceremonial. However, these authors argue that the Gasiorowski index understates the level of democracy during the period between 1928 and 1932, when the United States supervised national elections that were apparently relatively free and fair, in comparison with the previous period.

Przeworski et al. Data

Przeworski et al. (2000) provide data on regime type for 141 countries in the period from 1950 through 1990. A country is scored as democratic if the chief executive is elected, the legislature is elected, there is more than one political party, and there has been alternation in power. Any country that does not meet these criteria is scored as authoritarian. Data for these coding decisions are drawn in part from Arthur S. Banks's (1993) *Cross-National Time-Series Data Archives*, and also from unspecified historical sources.

Pragmatic. Elkins (2000) finds the dichotomous measure of democracy preferred by Przeworski et al. less useful in specific applications than a 10-category graded version of their indicator. In a test of a hypothesis drawn from the democratic peace literature, the graded version proves statistically significant while the dichotomous version does not. In an analysis of the effects of democracy on regime longevity, the graded measure allows the discovery that democracy has a nonlinear relationship with longevity. Simulated results also suggest that the dichotomous measure has a greater proportion of measurement error than does the graded measure. Mainwaring, Brinks, and Pérez-Liñán (2001) report that their indicator of democracy in Latin America has a correlation of 0.83 with the Przeworski et al. scores for the region. Bogaards (2007) criticizes the Przeworski et al. indicator for its reliance on electoral turnover as a necessary condition for democracy. He shows that electoral turnover in Africa has a relatively weak relationship with subjective measures of democracy that do not incorporate electoral results, specifically the Polity and Freedom House scales.

Coppedge and Reinicke Data

The Coppedge and Reinicke (1990) indicator of democracy provides regime rankings for 1985 and 2000. For 137 out of 170 countries analyzed, a successful Guttman scale was formed, based on five dimensions: suffrage, free and fair elections, freedom of organization, freedom of expression, and availability of alternative sources of information. An additional 26 cases were treated as “approximately equivalent variants” of the Guttman scale, and with slight adjustments were included in the index. The remaining 7 cases were treated as “anomalies” and were excluded (1990, pp. 56-7). For the 1985 rankings, specific data sources are not specified; 2000 data rely on the State Department's *Country Reports on Human Rights Practices*.¹

Levels of Measurement. Munck and Verkuilen (2002) praise the use of Guttman scaling as the aggregation rule for producing the Coppedge and Reinicke data. This technique helps ensure that the assumptions of the ordinal level of measurement are fully met.

Pragmatic. Przeworski et al. (2000) report that in a probit analysis, the Coppedge and Reinicke data correctly predict 92 percent of the Przeworski et al. scores. Therefore, the two measures are empirically similar. Inkeles (1990) finds that the correlation between the Coppedge and Reinicke measure and the Freedom House civil liberties measure is 0.94. Once again, this shows a high degree of empirical similarity between the two indicators. Munck and Verkuilen (2002) argue, from a pragmatic perspective, that the Coppedge and Reinicke indicator has important strengths. In particular, this scale was developed through a process that tested intercoder reliability and found that scale construction did not depend on idiosyncratic individual interpretations.

Supplemental Bibliography for Appendix¹

Bogaards, M. (2007). Measuring Democracy through Election Outcomes: A Critique with African Data. *Comparative Political Studies*, 40, 1211-1237.

Gasiorowski, M. J. (1996). An Overview of the Political Regime Change Dataset. *Comparative Political Studies*, 29(4), 469-483.

¹ All the sources cited in the appendix are listed in the bibliography for the published article, except for the following two.

¹ Available: <http://www.state.gov/j/drl/rls/hrrpt/>. Viewed March 31, 2013.