

Correlation Versus Interchangeability: The Limited Robustness of Empirical Findings on Democracy Using Highly Correlated Data Sets

Gretchen Casper and Claudiu Tufis

*Department of Political Science, Pennsylvania State University,
107 Burrowes Building, University Park, PA 16802-6202*

e-mail: gcasper@psu.edu

e-mail: cdt130@psu.edu

This article shows that highly correlated measures can produce different results. We identify a democratization model from the literature and test it in more than 120 countries from 1951 to 1992. Then, we check whether the results are robust regarding measures of democracy, time periods, and levels of development. The findings show that measures do matter: Whereas some of the findings are robust, most of them are not. This explains, in part, why the debates on democracy have continued rather than been resolved. More important, it underscores the need for more careful use of measures and further testing to increase confidence in the findings. Scholars in comparative politics are drawn increasingly to large-*N* statistical analyses, often using data sets collected by others. As in any field, we show how they must be careful in choosing the most appropriate measures for their studies, without assuming that any correlated measure will do.

1 Introduction

Democracy, like representation or power, is a basic concept in political science that is inherently difficult to measure. Although different scholars have accepted different trade-offs between consistency and operationalization when constructing their measures, resulting in a range of different measures of democracy, their measures correlate highly. It would be a reasonable assumption for a researcher to treat the measures interchangeably, selecting one that best fits the time period, number of countries, or particular variables that he or she was interested in. However, as we show here, despite high correlations, the use of these different measures can produce different results.

This article proceeds as follows. First, we describe three measures of democracy that are commonly used to study democratization, noting their high levels of correlation. Second, we

Authors' note: We acknowledge generous advice from Frank Baumgartner, Scott Bennett, David Brown, Suzanna DeBoef, Scott Gates, Barbara Geddes, and Quan Li. A more complete version of this article, containing further discussion of the democratization literature, data sources, and complete footnotes and references, is available on the *Political Analysis* Web site.

identify a model of democratization from the literature and test it in more than 120 countries from 1951 to 1992. Third, we check the robustness of the findings across measures of democracy, time periods, and levels of development. The conclusion points out that different measures lead to different results: Coefficients change directions, size, and statistical significance. This has strong implications for democratization scholars, but probably is a common problem for researchers in other areas as well. Scholars should beware: Correlation does not imply interchangeability.

2 Measuring Democracy

Dahl (1971, p. 4) argues that democracy entails competition and participation: Candidates for public office compete in elections and citizens participate in the process by selecting the winners. Scholars have revised this minimal definition by including additional dimensions such as political rights and civil liberties, socioeconomic equality, uncertainty, and the absence of military influence. Not surprisingly, the number of measures of democracy mirrors this diversity of definitions. These measures include Polity (Marshall and Jaggers 2000), Polyarchy (Vanhanen 2000), and Freedom House (2001), as well as measures constructed by several other scholars. It is common for the authors to show that their measures correlate with the methods of others to establish the reliability of their measures. However, although these measures may correlate highly, they vary in a number of ways, including the dimensions used to measure democracy, the time period covered, and the number of countries included.

This article compares three measures of democracy: Polity IV, Polyarchy 1.2, and Freedom House. First, they are three of the most widely used measures in democratization research. Second, they are the most similar in that all three consciously start from Dahl's definition of democracy. Finally, these three measures are highly correlated with each other (range: .85–.92). By choosing these highly correlated measures all drawing from a similar conceptual measure, we set up a most difficult case for our analysis. If we find that the results from testing a model of democratization are not robust using these three measures, then the problem would be even worse if other, more diverse, measures were incorporated into our research. Below, we discuss how these three measures operationalized democracy.

Polity was originally constructed to test the durability of states. It includes democracy and autocracy indicators for more than 160 countries from 1800 to 1999 by coding five institutional dimensions. Both of the democracy and autocracy indicators were constructed as 11-point scales. Vanhanen (2000) constructed Polyarchy to explain historical patterns of democratization. Polyarchy codes 187 countries from 1810 to 1998 on competition and participation and then weights them equally to create a democracy index. Freedom House set out to study freedom by measuring countries' levels of democracy. The annual surveys coded 192 countries from 1973 to the present across two dimensions—political rights and civil liberties—on a 7-point scale, whereby countries coded 1 were *most free* and those coded 7 were *least free*.

Democracy is a difficult concept to measure, yet the high correlations suggest that these three measures generally code countries in very similar ways. It is not surprising, then, that scholars would treat the three measures as virtually interchangeable, choosing to use whichever one had more complete coverage of the time period they were interested in, more favorable geographic coverage, or some other reason. This practice, though seemingly reasonable, is more hazardous than scholars realize. The next section tests a model of democratization to assess the robustness of the findings by using these three highly correlated measures of democracy.

3 Model of Democratization

Scholars have constructed a wide range of models to explain why some countries are democratic and others are not. The results from these tests have enhanced rather than ended debates in democratization, as findings have been inconsistent. Democratization theorists see these as conceptual debates. However, it may be that there is a methodological explanation for why these debates have continued. As we shall see, use of these highly correlated measures leads to different results. Below, we introduce the model of democratization used in this study. (Complete documentation of the measures is available on the *Political Analysis* Web site.)

We used the three measures of democracy discussed earlier. As is commonly done with Polity data, we subtracted a country's autocracy score from its democracy score and added 10 points to create an annual measure that ranged between 0 (*not democratic*) and 20 (*most democratic*). For Polyarchy, we used Vanhanen's democracy index. With Freedom House, we followed standard practice: we transposed the two Freedom House scales (for political rights and civil liberties) and added them together to create a 13-point scale from 2 (*least free*) to 14 (*most free*).

To explain levels of democracy, we included socioeconomic and institutional explanatory variables commonly used in democratization studies. The six socioeconomic variables are income, growth, trade dependence, inflation, primary education, and secondary education. Income is measured as logged real GDP per capita. Growth is the growth rate of real per capita income. Trade dependence is measured as openness, which is the sum of exports and imports divided by real GDP per capita. Inflation is measured by the Consumer Price Index. Primary education and secondary education are measured as the percentage of the population enrolled in primary or secondary school. The two institutional variables are types of executive systems and party fractionalization. A country's executive system was coded annually as one of three possible types: presidential, parliamentary, or other. To measure party fractionalization, we adopted Rae's index, which calculates the probability that any two legislators selected at random are members of the same political party. The next section presents the results from the democratization model and the robustness checks.

4 Results

We tested the democratization model in more than 120 countries, from 1951 to 1992. Because the data set pools across countries and years, ordinary least squares results may report incorrect standard errors as a result of potential problems with heteroskedasticity, contemporaneous correlation, and serial correlation. Because of these issues, we applied regression analysis with panel-corrected standard errors (PCSEs) (Beck and Katz 1995, p. 636). We then checked the robustness of the findings, paying particular attention to the measures for democracy, time periods, and levels of development.

First, we tested the democratization model from 1951 to 1992, using two measures of democracy: Polity and Polyarchy. (We did not include Freedom House in the initial test because it starts in 1973.) We lagged the explanatory variables and tested the model by using the same country-years from both Polity and Polyarchy. The results are shown in the first two columns of Table 1. Four of the explanatory variables are significant regardless of the measure of democracy used—income, parliamentary systems, and party fractionalization are positively related to democracy, whereas growth is negatively associated with democracy.

Looking at the other variables, we can see that although primary education is significantly associated with democracy when using Polyarchy, its significance drops out when using

Table 1 Democratization model: measures of democracy and time

	1951–1992		1951–1973		1975–1992	
	<i>Polity</i>	<i>Polyarchy</i>	<i>Polity</i>	<i>Polyarchy</i>	<i>Polity</i>	<i>Freedom House</i>
GDP pc, logged	3.614*** (.207)	6.580*** (.349)	2.806*** (.356)	5.782*** (.619)	3.372*** (.280)	2.236*** (.132)
Real GDP pc growth	–0.018** (.006)	–0.038*** (.009)	–0.013 (.011)	–0.023 (.013)	–0.023** (.009)	–0.010* (.004)
Openness	–0.009* (.004)	–0.012 (.007)	–0.016 (.009)	–0.021 (.016)	–0.009* (.004)	–0.004 (.002)
Inflation	–0.009 (.008)	0.009 (.009)	0.030 (.030)	0.089 (.050)	–0.002 (.010)	–0.003 (.003)
Primary education	–0.043 (.035)	–0.212*** (.056)	0.069 (.061)	–0.142 (.076)	–0.102* (.047)	–0.038* (.018)
Secondary education	0.036 (.071)	0.121 (.124)	0.368*** (.102)	0.213 (.183)	0.055 (.099)	0.030 (.042)
Presidential	0.245 (.248)	–0.080 (.382)	0.238 (.377)	–0.061 (.569)	0.508 (.416)	0.607*** (.191)
Parliamentary	2.061*** (.410)	3.446*** (.548)	3.390*** (.561)	5.043*** (.648)	2.924*** (.592)	0.768*** (.230)
Party fractionalization	2.788*** (.514)	7.294*** (.741)	3.931*** (.894)	8.050*** (1.176)	3.598*** (.869)	1.864*** (.465)
Constant	–17.812*** (1.463)	–40.396*** (2.280)	–14.589*** (2.670)	–36.292*** (4.214)	–16.291*** (1.704)	–9.713*** (.882)
<i>N</i>	2812	2812	1245	1245	1402	1402
<i>R</i> ²	.35	.38	.44	.41	.39	.58
Wald test	864.48	1814.86	425.13	832.87	822.01	1345.52
Probability > χ^2	0.000	0.000	0.000	0.000	0.000	0.000

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. All tests are two-tailed. All explanatory variables are lagged 1 year.

Polity. Trade openness, in turn, is significant only when using Polity as a measure of democracy. The unstable results of these two explanatory variables may be influenced by Polity's operationalization of democracy, which focuses on institutional aspects. Inflation and presidential systems, although not significant, switch the directions of their associations, depending on whether we use Polity or Polyarchy. The results from the democratization model show that using the same model and the same country-years but different measures of democracy generate different results.

Next, we checked the robustness of the findings for the democratization model across time and level of development. The first test considers whether the relationships between certain variables and levels of democracy change across time, and if so, whether these changes are consistent for all three measures of democracy. The extension of the right to vote over time has generated higher levels of political participation, whereas the effect of inflation has changed over time. To test the model across different time periods, we divided the data into two sets: 1951–1973 and 1975–1992. We chose 1974 as the dividing point for theoretical and practical purposes. Because the Third Wave of democratization began in 1974 (Huntington 1991), resulting in a sizeable increase in the number of democracies in the world, we can group the Third Wave period as a distinct set. In addition, this division results in two data sets of relatively equal size. For the first time period, 1951–1973, we used the same country-years from Polity and Polyarchy. For the second time period, 1975–1992, we included Freedom House and used the same country-years from all three measures.

As we can see from Table 1, three of the nine explanatory variables have stable results across all three measures and both time periods: income, parliamentary systems, and party fractionalization are positively and significantly associated with democracy. The remaining variables have inconsistent results. Also, the choice of measures of democracy has an important influence on the size of variance explained. For example, for 1975–1992, the model best explains levels of democracy when it uses Freedom House (58% of total variance), followed by Polyarchy (42%) and Polity (39%). The comparison suggests that the results one obtains are a function of the measures used for democracy.

Growth is negatively associated with democracy during 1975–1992 for all three measures, but with different levels of significance. Openness is significantly and negatively related to democracy during 1975–1992 only when using Polity. Primary education is not significant prior to 1974, but is significantly and negatively related to democracy for 1975–1992 with all three measures. Secondary education, however, has a significant positive association for 1951–1973, but only when using Polity. Presidential systems show a significant positive association in 1975–1992, but only with Freedom House.

One could explain the different coefficients in the two subsamples by arguing that the effects of some variables might change over time. Such an interpretation, however, holds only if the variables are consistent inside each subsample, regardless of the measure of democracy used. As shown earlier, this is not the case for openness, secondary education, and presidential systems. Thus, the findings in these two subsamples add more evidence supporting our conclusion that the choice of measures affects the results of the analysis.

Finally, regarding level of development, some scholars argue that the effects of the variables are different in developed versus developing countries. To address this issue, they limit their cases to developing countries. To see whether the effects are different for countries with different levels of development, we divided the sample into two—OECD countries and non-OECD countries—and reran the model separately for each subset. Again, we used the same country-years for all three measures.

As Table 2 shows, party fractionalization is positively and significantly associated with democracy for both OECD and non-OECD members. However, results for the other eight

Table 2 Democratization model: measures of democracy and level of development

	OECD (1951–1992)			Non-OECD (1951–1992)		
	Polity	Polyarchy	Freedom House	Polity	Polyarchy	Freedom House
GDP pc, logged	3.340* (1.579)	4.952 (2.634)	2.534*** (.481)	2.055*** (.294)	2.891*** (.548)	1.414*** (.137)
Real GDP pc growth	–0.044 (.026)	–0.091 (.048)	–0.019 (.010)	–0.015* (.007)	–0.034** (.011)	–0.007 (.004)
Openness	0.015 (.009)	0.043* (.018)	0.009** (.003)	–0.003 (.004)	0.002 (.006)	0.000 (.002)
Inflation	0.003 (.010)	0.019 (.024)	0.003 (.004)	–0.006 (.011)	–0.004 (.011)	–0.010* (.004)
Primary education	0.161*** (.048)	–0.272* (.126)	0.108*** (.021)	0.014 (.046)	–0.050 (.071)	0.025 (.020)
Secondary education	0.234*** (.068)	–0.691*** (.188)	0.111*** (.034)	0.045 (.106)	0.241 (.158)	0.025 (.046)
Presidential	1.717 (.942)	–1.107 (1.874)	1.027** (.339)	0.546 (.380)	0.724 (.488)	0.617*** (.184)
Parliamentary	0.898 (.943)	2.863 (1.608)	0.522 (.296)	2.120** (.680)	1.729* (.799)	0.446 (.292)
Party fractionalization	5.756* (2.297)	16.976*** (3.253)	2.932*** (.693)	3.567*** (.821)	7.465*** (1.090)	1.626*** (.420)
Constant	–20.233 (14.382)	–24.553 (24.272)	–14.934*** (4.362)	–9.031*** (1.837)	–18.150*** (3.080)	–4.691*** (.890)
<i>N</i>	431	431	431	1296	1296	1296
<i>R</i> ²	.70	.52	.88	.24	.22	.40
Wald test	40.80	143.40	100.87	137.23	188.76	233.70
Probability > χ^2	0.000	0.000	0.000	0.000	0.000	0.000

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. All tests are two-tailed. All explanatory variables are lagged 1 year.

explanatory variables are unstable. The model explains a greater amount of variance for OECD countries compared to non-OECD members (although there are differences in the sizes of the coefficients and the levels of significance), suggesting that we know more about democracy in developed than in developing countries. Again, the choice of measures of democracy significantly affects the size of variance explained: In the OECD subsample, the explained variance ranges between 88% (using Freedom House) and 52% (using Polyarchy); in the non-OECD subsample the explained variance ranges between 40% (Freedom House) and 22% (Polyarchy).

Income is significant in both subsamples for all three measures, with the exception of OECD countries when using Polyarchy. Growth is significantly but negatively associated with democracy for non-OECD countries when using Polity or Polyarchy. Openness has a positive significant association only for OECD countries and only when democracy is measured by Polyarchy or Freedom House. Inflation has a significant negative relationship for non-OECD countries when using Freedom House, but no effect otherwise. Both primary and secondary education have a positive significant relationship for OECD members when using Polity or Freedom House, yet a significant negative effect for OECD members with Polyarchy. Presidential systems are positively and significantly associated with democracy for both OECD and non-OECD countries, but only when using Freedom House. Parliamentary systems have a positive significant association for non-OECD countries only with Polity and Polyarchy.

The results in Table 2 strengthen our conclusion. Within each subsample, eight of the nine explanatory variables either have unstable significance or change the sign of their association (in the case of the education variables) depending on the measure of democracy one uses. Because the results are unstable within groups of countries with similar levels of development, the importance of the measure of democracy becomes evident.

To summarize this section, the results for only one variable—party fractionalization—are consistent across the three measures of democracy and two robustness checks. The results for other variables are inconsistent: coefficients became larger or smaller, significance levels increased or disappeared, and signs changed direction. The same explanatory variable is either significant or not and has a positive or a negative effect, depending only on the measure used in the analysis. We argue that the confusion in the democratization literature is explained largely by the instability of the results.

5 Conclusion

Over the last forty years, political scientists have studied why some countries are democratic while others are not, and why some countries are more democratic than others. To conduct this research, they have used different measures of democracy. The scholars who constructed these measures and the researchers who used them have consistently pointed out that they are highly correlated with each other. The implication is that they are interchangeable, leaving researchers free from worrying about alternative conceptions or sensitivity checks. Thus, researchers have based their choice of a particular measure on the time period covered, their preferred statistical applications, or the number of valid cases for the variables with which they are most concerned. Furthermore, scholars tend to assume that because the measures are highly correlated with each other, their findings do not depend on the particular measure used, but rather are robust across many or all of them. In this article, we show that this assumption is often erroneous.

We tested a model based on the democratization literature, using three different measures of democracy, and then checked for robustness. Of the nine variables tested here,

three (income, parliamentary system, and party fractionalization) are consistent across the democracy measures and time periods, but only one (party fractionalization) is consistent across the measures and all robustness checks. The results explain in part why the debates in the field have continued rather than been resolved.

It is not uncommon for researchers to engage in venue-shopping. This is true not only for studies of democratization, but also for other topics of study in political science. Given the inconsistent results, however, from using different measures that are highly correlated, scholars must select measures carefully and explain their selection based on theoretical reasons rather than expediency or taste. Furthermore, they should also explain why they did not choose other measures that are considered standards in their field. Finally, researchers should explain why their results might vary depending on the measure used. Given the availability of data and the relative ease of adding them to a data set, one solution is for researchers to test their models by using different measures and keep the model that satisfies some requirement for robustness. Their analyses and articles will be longer, but they will be taking statistical testing seriously in democratization, as well as in other fields of political science.

References

- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data." *American Political Science Review* 89:634–647.
- Dahl, Robert A. 1971. *Polyarchy: Participation and Opposition*. New Haven, CT: Yale University Press.
- Freedom House. 2001. "Freedom House Country Ratings." (Available from <http://www.freedomhouse.org/ratings/index.htm>.)
- Huntington, Samuel P. 1991. *The Third Wave*. Norman, OK: University of Oklahoma Press.
- Marshall, Monty G., and Keith Jaggers. 2000. *Polity IV Dataset and Users' Manual: Political Regime Characteristics and Transitions, 1800–1999*. (Available from <http://www.bsos.umd.edu/cidcm/polity>.)
- Vanhanen, Tatu. 2000. "The Polyarchy Dataset: Vanhanen's Index of Democracy." (Available from <http://www.svt.ntnu.no/iss/data/vanhanen>.)