

The Democracy Cluster Classification Index

Mihaiela Ristei Gugiu

*Department of Political Science, The Ohio State University, 2189 Derby Hall,
154 N Oval Mall, Columbus, OH 43210
e-mail: gugiu.1@osu.edu (corresponding author)*

Miguel Centellas

*Croft Visiting Assistant Professor of Political Science, 335 Deupree Hall,
University of Mississippi, University, MS 38677
e-mail: mcentell@olemiss.edu*

Edited by R. Michael Alvarez

Utilizing hierarchical cluster analysis, a new measure of democracy, the DCC index, is proposed and constructed from five popular indices of democracy (Freedom House, Polity IV, Vanhanen's index of democratization, Cheibub et al.'s index of democracy and dictatorship, and the Cingranelli-Richards index of electoral self-determination). The DCC was used to classify the regime types for twenty-four countries in the Americas and thirty-nine countries in Europe over a thirty-year period. The results indicated that democracy is a latent class variable. Sensitivity and specificity analyses were conducted for the five existing democracy indices as well as the newly proposed Unified Democracy Scores index and a predicted DCC score. This analysis revealed significant problems with existing measures. Overall, the predicted DCC index attained the highest level of accuracy although one other index achieved high levels of accuracy in identifying nondemocracies.

1 Introduction

Three decades since the “third wave” (Huntington 1991) of democratization swept Latin America and two decades since communism fell in Central and Eastern Europe (CEE), concerns remain over the quality of democracy in many countries. Although there have not been any outright democratic reversals in most countries, questions remain over the consolidation of individual democratic regimes and long-term prospects of democracy. The return of populism in Latin America in the 1990s in places like Peru and Venezuela raised doubts about the quality of democracy and suggested the emergence of new forms of “illiberal” regimes (Smith and Zeigler 2008). In CEE, countries like Bulgaria, Romania, and Slovakia were consistently classified as “illiberal democracies” for most of the 1990s (Vachudova 2005).

The ability of social movements to drive out sitting presidents in places like Bolivia or to prevent the return to power of Victor Yanukovich in Ukraine raised concerns about the ability of these regimes to effectively channel social demands. Although no Latin American democracy has reverted to outright authoritarianism, significant differences in quality of democracy exist in Venezuela and Costa Rica. Similarly, important differences in democratic quality between the Czech Republic and Ukraine are noticeable. Yet, despite a proliferation of democracy measures, such differences remain difficult to empirically measure, limiting the ability to pursue research questions in which quality of regime is an important variable.

Another challenging problem is determining when a country has transitioned from a democratic to a nondemocratic regime type. Determining when a democratic regime erodes to the point where

Authors' note: We thank P. Cristian Gugiu for his valuable analytical suggestions, Jose Antonio Cheibub for so generously making the data set “Dictatorship and Democracy Revisited” available to us, and the editors and the anonymous reviewers for their helpful comments which improved our article. Data to replicate the statistical analyses in this article are available at <http://hdl.handle.net/1902.1/20393>. Supplementary materials for this article are available on the *Political Analysis* Web site.

it should be qualified as either a “diminished subtype” (Collier and Levitsky 2009) or a “competitive-authoritarian” regime (Levitsky and Way 2002) is an important conceptual and empirical challenge. The same is true of determining when such a regime crosses the threshold to where it should be classified as a democracy. Determining the timing of Mexico’s transition to democracy is such a problem. Mexico is often considered to have formally transitioned to democracy in 2000, after the long-dominant Institutional Revolutionary Party (PRI—*Partido Revolucionario Institucional*) lost an election after nearly seventy years in power. Yet elections in Mexico were increasingly competitive throughout the 1980s and 1990s, prompting debate about whether Mexico in the 1990s was truly a nondemocratic regime or simply (like Japan or India) a case of a dominant-party democratic regime (Greene 2007; Wong and Friedman 2008; Bogaards and Boucek 2010). However, existing measures of democracy often provide different and sometimes contradictory ratings, thereby further complicating scholars’ decision regarding which measure of democracy is the most reliable and valid.

An attempt to address this problem was recently made by Pemstein, Meserve, and Melton (2010), who suggested that by employing a Bayesian latent variable approach one could combine various existing indices into a “Unified Democracy Scores” (UDS) index. Although in a few instances a UDS score was produced from ten democracy indices, on average, only six were available for the period (1946–2000) examined by the authors. Moreover, the coverage of the indices varied widely from five to 191 countries over this span of time. On average, five or fewer indices were employed to generate the posterior means for a given country year. This minor observation aside, the UDS potentially improves upon prior measures of democracy by minimizing the role of measurement error. Nevertheless, the strength of the observed results is a function of the robustness of the method employed in relation to violations of its assumptions, as well as the reliability and validity of the underlying data.

Pemstein et al. (2010) identified a number of assumptions, including that the latent variable (democracy) is unidimensional and continuous, errors are independent and normally distributed, and the variance parameters for each latent measure are independent and conform to an inverse-gamma prior density. Although the authors reported that the model fit the data, the tests performed did not investigate whether the ten indices indeed measured a unidimensional construct or whether regimes should be classified along a democracy continuum. Yet, given the overlap between the confidence intervals (CIs) included in their figure for 2000, it seems the UDS cannot easily discern between countries that are not on opposite ends of the democracy spectrum. Furthermore, it was not apparent whether the statistical model employed properly accounted for the use of time-series and nested data, thereby violating the independence assumption. That is, error terms for point estimates closer in time for a given country are likely to be more correlated than point estimates farther in time. Similarly, error terms for a given country are more likely to be correlated than the error terms for different countries. Since it is not clear whether the method they employed is robust to such violations, further research is warranted for generating a composite measure of democracy.

The primary purpose of this study was to classify the regime types for twenty-four countries in the Americas and thirty-nine countries in Europe over a thirty-year period (1980–2010) using existing democracy measures. This study expands on previous research by utilizing hierarchical cluster analysis (HCA) to develop an index—the “democracy cluster classification” (DCC)—that can distinguish among regime types. By grouping countries based on their similarities within and across groups, one is able to examine the continuity of the democracy construct. If the latent construct representing democracy is indeed continuous, then one expects the HCA to produce a dendrogram with small Euclidean distances between the branches (i.e., a large number of clusters each composed of a few elements). If one observes only a few clusters with relatively large Euclidean distances, then the differences within a cluster are minor relative to the differences between clusters, thereby suggesting that the construct democracy is a latent class variable. Furthermore, if the explained variance is very high, then the error that entered the model is inconsequential—provided the variables used in the HCA are representative of the latent construct. This does not imply that some variation does not exist within each cluster, although the variance between clusters is greater than the variance within each cluster.

A secondary purpose of the study was to critically appraise existing measures by assessing their strengths and weaknesses. Furthermore, the overall classification rate of these measures was compared against the DCC index to determine whether single measures performed as well as the composite variable. Finally, the classification rate of each index was examined by cluster so as to determine whether some indices were better suited for identifying democracies, whereas others were better for identifying dictatorships.

2 Challenges of “Measuring” Democracy

Discussions about the quality of democracy take many different forms. Typically, the term describes the extent to which a democratic regime achieves a normative standard for democracy. Such discussions often focus on various “outputs” of democracy, such as the rule of law (Maravall and Przeworski 2003; O’Donnell 2004), policies that reduce inequality (Rueschemeyer 2004; Bermeo 2009), or efforts to protect or expand civil liberties (Beetham 2004). Although these issues are important—and perhaps difficult to separate from our basic understanding of what democracy is and why it matters—they move us further away from conceptual discussions of democracy as a type of political regime.

Schmitter and Karl (1991) warned of the confusion generated by imputing additional values into democracy. We may very well wish for democracies to produce a number of normative outputs, but there is no inherent reason that every democracy *must* produce them. Schmitter and Karl push for a “procedural” or “minimal” definition of democracy. Their definition is self-consciously derived from Schumpeter’s formulation of democracy as an “institutional arrangement for arriving at political decisions by means of competitive struggle for the people’s vote” (1975, 269). While acknowledging, like Dahl (1971), that a number of “procedural minimal” conditions must exist for democracy to function, these are essentially reduced to background or prerequisite conditions, not specific indicators of democracy itself. To the seven “institutional guarantees” proposed by Dahl (1982), Schmitter and Karl (1991) added that elected officials must have autonomy from other actors (e.g., the military) and the polity must be self-governing (i.e., enjoy full sovereignty). However, none of these “guarantees” by themselves define democracy.

Agreeing on conceptual definitions of—and operational indicators for—democracy is critical for comparative scholarship that uses regime type as a key variable of analysis. Similarly, studies that seek to explain differences in the quality of democracy across cases—or to use such differences as possible causes for other variations across cases—also require valid and reliable measures. Munck and Verkuilen (2002) highlight this problem in their review of existing democracy measures. The three strongest indices they identified were those produced by Alvarez et al. (2011), Coppedge and Reinicke (1991), and Hadenius (1992); the three weakest indices were those produced by Freedom House (FH), Gasiorowski (1996), and Vanhanen (2000). Nonetheless, all measures were beset by a host of problems of conceptualization, measurement, aggregation, and validity. Some of the most common issues identified were the use of a minimalist definition of democracy, insufficient sensitivity to key issues, assumptions of continuity and unidimensionality, lack of clear coding rules, ad hoc application of the aggregation rules, and interpretation of an index’s reliability as its validity (Munck and Verkuilen 2002). Furthermore, the established high correlations among all these different measures of democracy (Adcock and Collier 2001) should be cautiously considered since most measures were constructed using the same sources and, sometimes, the same coders, which would explain the high correlation scores (Munck and Verkuilen 2002).

Despite well-known weaknesses, the two indicators most commonly used to distinguish between democratic and nondemocratic regimes—and to distinguish differences in the degree of quality of democratic regimes—are the FH and Polity indices. Although different in purpose, both indices are similarly constructed: country experts code individual countries along a set of prescribed, standardized criteria. Where they differ significantly is in their operationalization of component indicators: Polity primarily codes on a limited set of regime “characteristics,” whereas FH codes along a broad set of regime “outcomes” as reflected in the levels of freedom enjoyed by a regime’s citizens. In recent years, there has been a preferential trend toward the Polity index, which has the benefit of being able to project back to 1800—although the reliability of scores before 1900 is highly

suspect (Munck and Verkuilen 2002). One significant disadvantage of the Polity index is that it does not cover a number of countries, particularly those with small populations; in contrast, FH index is more comprehensive, though it only extends back to 1973. It should also be noted that Vanhanen's index of democratization (2000), which appeals to some researchers because it relies exclusively on measures of participation and competition derived from (objective) electoral data, employs a rather arbitrary aggregation rule (Munck and Verkuilen 2002) that generates scores often inconsistent with those of other democracy indices (Pemstein, Mesere, and Melton 2010).

In an earlier examination of democracy measures, Collier and Adcock (1999) proposed a "pragmatic approach" to the conceptualization and operationalization of democracy that would account for "the goals and context of research" (561). They claim democracy scholarship would not benefit from a "standardized" measure of democracy. Indeed, the lack of consensus over the meaning of this protean word has resulted in the proliferation of various measures of democracy. However, the choice of which measure of democracy to use can have significant implications for a study's conclusions. Some scholars found that when democracy is used as an independent variable, ignoring measurement error significantly increases the likelihood of inferential error (Treier and Jackman 2008)—consistent with studies that cautioned about the "sizable" measurement error in democracy measures (Bollen 1990, 8; Bollen and Paxton 2000).

While cautioning against the potential proliferation of many subtypes of democracy, Collier and Adcock (1999) indicated that carefully employed gradated measures of democracy can capture cases that lack one or more attributes associated with democracy. However, scholars are often concerned with the evolution of democracies across time and/or countries. Hence, a major limitation of gradated measures of democracy is that they presume that nominal (or at best ordinal) scales can be treated as interval. That is, although gradation suggests a rank-ordering of regime types according to some specified attributes, gradated measures are not suitable for comparative or longitudinal case studies because they are nominal measures. For example, if one scholar classifies a country as a "male democracy" at one point in time and a different scholar classifies it as a "controlled democracy" a few years later, what could one conclude regarding the evolution of democracy in that country? Could one say the absence of women's suffrage was more or less important for democracy than serious limitations on contestation? Moreover, how could one meaningfully compare two countries classified as subtypes of democracy with different "missing attributes" and place them on the democracy continuum? Such a comparison necessitates the use of an interval or ratio scale that quantifies the distance along the democracy continuum associated with each dimension (or lack) of this construct. To the best of our knowledge, such a determination is not possible without making unsubstantiated assumptions about the rank-order and impact of dimensions on the democracy continuum. Therefore, although relative comparisons between countries classified in different regime types is possible, relative comparisons within a regime type cannot be made until a method for rank-ordering and quantifying differences between values (e.g., lack of women's suffrage versus serious limitations on contestation) is established and validated. Although it may be useful to discuss a specific case by using a gradated measure of democracy, classifying regime types requires measures that allow for meaningful comparisons.

A recent collaborative project proposed a new approach to measuring democracy which emphasized the use of a disaggregated set of indicators, as opposed to most of the extant measures which do not provide disaggregated data (Coppedge and Gerring 2011). However, completing such an ambitious project rests upon the availability of scarce resources (human, financial, institutional).¹ An alternate approach may be to move away from measuring democracy in favor of a more limited concept, such as democratic "competitiveness" (Centellas 2011). However, the usefulness of such an approach would be limited to cases already identified as democracies. In the meantime, the search continues for a democracy index that produces a reliable and valid measure of the concept in a timely fashion and yet is pertinent to different researchers.

¹ Coppedge and Gerring (2011) list thirty-two mid-level indicators, which must be further disaggregated into low-level indicators, whose numbers can vary between 1 to 10 per mid-level indicator. If, for example, on average, only three low-level indicators were used to measure each mid-level indicator, ninety-six indicators must be measured for each country and each year in order to produce comparable cross-country and cross-time measures of democracy.

3 Methodology

3.1 Data

The study focused on twenty-four countries in Latin America and thirty-nine countries in Europe, including Russia, from 1980 to 2010. These are regions with a large number of countries that underwent a process of democratic transition during the third wave. With the exception of Cuba (excluded from the study), every country in Latin America has held at least five competitive elections since 1980. However, the context under which these elections took place varies significantly. The study included a number of cases of “first elections,” which took place under authoritarian regimes (e.g., Chile 1989; Uruguay 1984), during periods of civil war (e.g., Guatemala and El Salvador throughout the 1980s), or during periods of democratic erosion (e.g., Peru in the 1990s, Venezuela in the 2000s), as well as countries considered democratic during the entire period (e.g., Costa Rica, Jamaica) and one country that gained independence during the period of observation (Belize in 1981). Additionally, countries rarely included in studies of democracy in the region (Haiti, Belize, Suriname, Guyana, Trinidad and Tobago, and Jamaica), but which increase diversity (e.g., cultural, institutional) within this sample, were included.

Similarly, half of Europe underwent significant changes since 1990. Whereas the Western democracies are a mark of stability on the continent, the CEE countries make it particularly diverse and significant for democracy scholarship. In addition to the Baltic states (Estonia, Latvia, and Lithuania), this study also included Russia and three other former Soviet republics: Ukraine, Belarus, and Moldova. Although Ukraine and Moldova have expressed interest in joining the European Union, they are very different from other CEE countries. The European sample also includes most of the former Yugoslav republics, with the exception of Serbia and Montenegro and the recently formed Republic of Kosovo. Hence, this degree of regional variation is unique and allows for a strenuous test of existing democracy measures.

For the HCA, five commonly used democracy measures were included: FH (2011); Cheibub, Gandhi, and Vreeland’s index of democracy and dictatorship (2009) (henceforth, DD); the Polity IV index (Marshall, Gurr, and Jaggers 2011); Vanhanen’s index of democratization (2011); and the Cingranelli-Richards index of electoral self-determination (2011) (henceforth CIRI), formerly known as the index of political participation. Although the first four measures are more familiar to scholars of democracy, CIRI was included because it captures an important aspect of democracy: political and electoral rights (Bollen 1990). Additionally, the HCA required indicators that provided scores for all years and all countries included in the study. Thus, it was not possible to include democracy measures that have been developed only recently (e.g., the Economist Intelligence Unit Democracy Index), only measured a few countries like BLM (Bowman, Lehoucq, and Mahoney 2005), or estimated democracy scores for only one or a very few years (e.g., Hadenius 1992). The recently proposed UDS index (Pemstein, Mesere, and Melton 2010) included the BLM and Political Regime Change (PRC) (Gasiorowski 1996; Reich 2002) in addition to the FH, DD, Polity, and Vanhanen indices. The two additional indices were omitted from the present study because BLM covers only five countries until 2000, whereas PRC does not provide scores after 1998. Finally, UDS was not included to avoid issues of multicollinearity due to its dependence on four of the aforementioned indices.

3.2 Statistical Procedure

Agglomerative HCA was employed to determine how well the five selected measures classified the selected countries with regard to the quality of democracy. HCA groups the units of analysis (e.g., people, cases, country-years) based on similarities and dissimilarities found in the pattern of responses across the set of variables that describe the objects to be clustered (Tan, Steinbach, and Kumar 2006). HCA is conceptually, but not statistically, related to factor analysis. Whereas factor analysis seeks to reduce the number of items (variables) to a few factors wherein the items corresponding to a factor are more highly correlated (either positively or negatively) with each other than with the items corresponding to other factors, HCA seeks to reduce the number of individual units of analysis to a smaller number of clusters whose members are more similar to

each other than they are to the members of other clusters. Generally, clusters are formed based on the Euclidean distance between the units of analysis, which can be computed by the Pythagorean formula (Johnson and Wichern 1998). That is, the distance between the p th and q th units of analysis is given by $d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$, where $\mathbf{p} = (p_1, p_2, \dots, p_n)$ and $\mathbf{q} = (q_1, q_2, \dots, q_n)$. Computing distances for each of the n observations yields a $p \times q$ matrix with $n(n-1)/2$ distances. The general HCA procedure entails five simple steps: (1) compute the distances between all possible combinations of units of analysis (initially there are n units of analysis); (2) find the smallest value for distance; (3) merge the clusters associated with that distance; (4) replace the row and column associated with the distance obtained in Step 1 with the distance between the merged cluster and the remaining units of analysis (or other clusters); and (5) repeat Steps 1–4 a total of $n-1$ times until all units are merged into a single cluster. Note that no assumptions are made about the distribution of the variables, the number of groups, or the group structure.

In agglomerative HCA, clusters are allowed to have subclusters that lead to a set of nested clusters organized as a tree (Tan, Steinbach, and Kumar 2006). This bottom-up procedure starts with the individual objects and, at each stage of the analysis, clusters together the most similar objects—a process that continues until all subclusters are eventually merged into a single cluster (Johnson and Wichern 1998; Friedman, Hastie, and Tibshirani 2009). If the groups produced are meaningful, then the clusters capture the natural structure of the data. The greater the similarity of objects within a group and dissimilarity between groups, the more robust and interpretable the cluster structure will be. Although several agglomerative hierarchical methods exist, Ward's minimum variance method was used because it produces mutually exclusive clusters from objects that are similar with respect to specified characteristics. At each level of clustering, Ward's method minimizes the within-cluster sum of squared error produced by merging two clusters from the previous iteration (Ward 1963). That is, in Step 1, Ward's method computes the distances for every possible combination of singletons (i.e., clusters composed of a single unit of analysis), singletons and clusters (composed of two or more units of analysis), and clusters obtainable from the previous iteration, where distance is equal to the sum of squares between the two clusters added up over all the variables. In Step 2, Ward's method identifies the two singletons and/or clusters that when merged produce the smallest within-group variance.

Since data for the HCA came from five different sources measured with different scales, the variables were standardized to balance for differences in variability. Although the traditional approach is to standardize variables by transforming them to have a mean of zero and unity variance, this is not the best approach for HCA. Comparing seven standardization approaches and their impact on the cluster analyses results, Milligan and Cooper (1988) found that the common z-score standardization approach was the least effective. The authors examined the standardized data sets in four different types of error environment conditions: error free data, error perturbed distances, presence of outliers, and presence of random noise dimensions. Furthermore, they tested the impact of various standardization processes using four different hierarchical clustering methods: single link, complete link, group average, and Ward's method. The standardization process that used the range of the variables as the divisor and Ward's method was found to best reproduce the natural clustering of the original data. This standardization approach employed the following mathematical algorithm: $Z = [X - \text{Min}(X)] / [\text{Max}(X) - \text{Min}(X)]$, where Z denotes standardization, X denotes the original variable, and Max and Min denote the maximum and minimum values, respectively (Milligan and Cooper 1988, 185). This study employed the aforementioned algorithm and Ward's method of hierarchical clustering. The HCA was first conducted on the Latin American data set and validated on the European data set.² All analyses were conducted using SAS 9.3.

² Note: standard practice in statistical model developments requires the validation of statistical models on an independent data set, so as to minimize the probability of capitalizing on chance.

3.3 Multiple Data Imputation

Inspection of the data revealed a small percentage of missing values for each measure, which represented 4.5% of the Latin American data set (743 country-years by five variables) and 2.3% of the entire European data set (950 country-years by five variables). Although these missing scores appeared reasonable, we imputed missing values to avoid the bias produced by listwise deletion (i.e., HCA requires complete cases for each variable), which would have resulted in 18.6% reduction of the Latin American data set and 10.1% reduction of the European data set. Compared to mean substitution, which tends to reduce the overall variance and produces biased results (e.g., underestimated error variances and inflated test statistics) (Horton and Kleinman 2007), multiple imputation introduces variability (across imputed samples), thereby generating unbiased results when the results are aggregated. Numerous imputation methods exist depending upon the level of measurement of the variables and existence of a monotonic or arbitrary missing pattern. The Markov Chain Monte Carlo (MCMC) method, which operates under the assumption of multivariate normality, was employed due to the arbitrary missing pattern and because the variable that contained the largest number of missing values (Polity) is reasonably continuous (SAS Institute 2012a).³ Following standard recommendations, five imputations were used to estimate the missing values via SAS 9.3 Proc MI. Although standard practice calls for separate analyses of each imputed data set prior to merging the results, examination of the five data sets revealed the variability between variable means to be extremely small (<0.1 point difference per variable across the five imputed data sets). Hence, the five imputed data sets were merged into a single data set (via averaging) and all subsequent analyses were conducted on this data set.⁴

4 Results

Although Ward's method of hierarchical clustering was found to be one of the best-performing methods, especially when data were standardized using the aforementioned algorithm (Milligan and Cooper 1988), it is sensitive to multivariate outliers (Milligan 1980). To control for the distortion of results by outliers, 10% of cases were trimmed from the initial analysis, in line with the recommendation made by the SAS Institute (2010). Furthermore, to minimize the impact of capitalizing on chance and to enhance the interpretability of a cluster, the minimum size permitted for a cluster was 10. After the first HCA was conducted, three criterion measures were used to determine the number of clusters: the cubic clustering criterion (CCC), pseudo F statistic, and pseudo t^2 statistic (Milligan and Cooper 1985). According to both the CCC and pseudo F statistics, the optimal number of clusters is equal to the maximum hierarchy level in each of the corresponding plots. In contrast, according to the pseudo t^2 statistic, also known as $Je(2)/Je(1)$ ratio criterion (Duda and Hart 1973), the optimal number of clusters is equal to the minimum hierarchy level in the plot (Milligan and Cooper 1985). As is standard practice, the combination of the results provided by these three criterion measures was used to make the final decision regarding the number of clusters in the data set along with the visual inspection of the dendrogram and the R^2 value.

The three criterion measures produced by the HCA on the Latin American data set indicated a potential five-cluster structure. However, a visual inspection of the dendrogram⁵ (see supplementary material) showed that a three-cluster solution was appropriate since it accounted for 92.1% ($R^2=0.921$) of the total variance. Moreover, since the goal was to produce a parsimonious model with high explanatory power, the three-cluster solution best met these criteria. Once the

³ As a means of detecting whether the multivariate normality assumption adversely impacted cluster assignment, a confirmatory analysis was conducted that omitted all the cases with missing values. Comparison of the assignment produced by this analysis to the one on the imputed data set revealed changes in less than 2% of the data. Hence, the MCMC imputation method did not adversely impact the results reported herein.

⁴ The between-imputation variances were extremely small and, therefore, their inclusion in the analysis was deemed unnecessary. For example, the largest between-imputation variance was 0.00017 for the Polity IV measure in the Latin American data set. The between variances were smaller for the European data set.

⁵ Given the large sample sizes, it was not possible to list the countries on the vertical axes of the dendrograms. Thus, the cluster assignments are included in appendices, also available on the *Political Analysis* Web site.

number of clusters was identified, a second HCA was conducted on the same data set to include the seventy-five cases previously omitted. Although the presence of outliers affected the proportion of variance accounted for by the three-cluster solution, the R^2 remained very high (0.890), suggesting that minimal amounts of error were present in the classification (particularly in the outlier cases). Note: some caution should be taken in interpreting the cluster assignment for the cases identified as outliers⁶ by the initial HCA.

The results produced by the HCA on the Latin American data set were validated on the European data set. As the dendrogram (see supplementary material) clearly illustrates, the three-cluster solution was confirmed with almost 100% of the variance ($R^2=0.996$) accounted for. As previously, a second HCA was conducted on the European data set to include the 10% of cases trimmed from the initial analysis. Although the presence of outliers slightly affected the proportion of variance accounted for by the three-cluster solution, the R^2 remained very high (0.928), thus indicating that negligible amounts of error were present in the classification.⁷ As a final check of validity of the three-cluster solution, an HCA was performed on all the data. This analysis also supported the model with a three-cluster solution explaining 96.3% of all the variance.

It is important to note that these results call into question the assumption that democracy is a *continuous* latent construct. Instead, the HCA results suggest that democracy may be a *discrete* latent (class) construct. This inference is supported by the fact that large Euclidean distances were found between a small number of clusters. Had the observed data supported a continuous latent model, then one would have found a dendrogram with a large number of branches separated by small Euclidean distances. This does not mean that there are no differences within clusters, but that the variance between clusters is considerably greater.

4.1 Interpretation of the Cluster Structure

Table 1 summarizes the means and CIs for each cluster and the overall mean for Latin American, European, and merged data sets. To facilitate interpretation of the means, recall that FH ranges from 1 to 7, with lower scores denoting free countries and higher scores denoting countries that are not free; Vanhanen's index ranges from 0 to 100, with higher scores denoting democratic regimes; DD index is a dichotomous measure where 1 denotes democracy and 0 denotes dictatorship; CIRI ranges from 0 to 2, with 0 denoting no political rights, 1 denoting some political rights, and 2 denoting full political rights; and Polity ranges from -10 to 10, with low scores denoting autocracy and high scores denoting democracy.

Examination of the means and CIs for Clusters 1 to 3 relative to the overall mean for the merged Latin American and European data set (Table 1) and the benchmarks provided for each measure of democracy easily enable interpretation of the clusters (Table 2). Comparison of the CI for the FH index for Cluster 1 to the corresponding CI for the overall mean revealed that the interval for Cluster 1 was lower than the overall interval. In contrast, the CI for the FH for Cluster 3 was higher than the overall CI. Extending this approach to all the indices and clusters yielded the results for Table 2, left panel. Moreover, since the FH, CIRI, and Polity indices provide guidance that enables one to classify their scores as either dictatorships, transitional regimes or anocracies, or democracies, one may use these benchmarks to interpret the means from the previous tables. Finally, the DD index can be interpreted along the dictatorship-democracy dichotomy.

However, since the benchmarks provided are based on whole numbers whereas our means are integers, an additional step was needed to facilitate benchmark interpretations. Namely, equal-sized intervals were identified for each whole number associated with an index score. This can be accomplished by subtracting the lowest scale value from the highest scale value and dividing the difference by the total number of the scale points. Using this increment, one can identify the equal-size intervals associated with each whole-number score. For example, in the case of FH,

⁶ A case was identified as an outlier if major disagreements existed among the scores of the five democracy indices.

⁷ The cluster assignment for each European country year is included in Appendix II and is available on the *Political Analysis* Web site along with the other supplementary materials.

Table 1 HCA cluster sample sizes, means, and corresponding 95% CIs^a

	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>	<i>Overall</i>
Latin America	<i>n</i> = 427	<i>n</i> = 127	<i>n</i> = 114	<i>n</i> = 668
FH	2.29 (2.21, 2.37)	2.65 (2.52, 2.78)	4.63 (4.40, 4.87)	2.76 (2.66, 2.85)
Vanhanen	18.69 (18.01, 19.38)	17.23 (16.22, 18.23)	5.73 (4.38, 7.08)	16.20 (15.56, 16.84)
DD	1 (1, 1)	1 (1, 1)	0 (0, 0)	0.83 (0.80, 0.86)
CIRI	2 (2, 2)	1 (1, 1)	0.86 (0.74, 0.98)	1.62 (1.57, 1.66)
Polity	8.12 (8.00, 8.25)	7.63 (7.38, 7.87)	-3.83 (-4.70, -2.97)	5.99 (5.61, 6.37)
Europe	<i>n</i> = 657	<i>n</i> = 111	<i>n</i> = 97	<i>n</i> = 855
FH	1.33 (1.29, 1.37)	2.49 (2.32, 2.66)	6.41 (6.28, 6.54)	1.99 (1.88, 2.10)
Vanhanen	33.80 (33.27, 34.32)	27.36 (26.05, 28.67)	1.92 (1.02, 2.82)	29.78 (29.00, 30.56)
DD	1 (1, 1)	1 (1, 1)	0 (0, 0)	0.90 (0.88, 0.92)
CIRI	2 (2, 2)	1 (1, 1)	0.04 (-0.00, 0.08)	1.67 (1.63, 1.72)
Polity	9.58 (9.51, 9.65)	8.23 (7.96, 8.51)	-7.33(-7.57, -7.09)	7.72 (7.37, 8.06)
Overall	<i>n</i> = 1084	<i>n</i> = 238	<i>n</i> = 199	<i>n</i> = 1521
FH	1.71 (1.66, 1.76)	2.58 (2.47, 2.68)	5.39 (5.20, 5.58)	2.33 (2.25, 2.40)
Vanhanen	27.85 (27.24, 28.45)	21.95 (20.92, 22.99)	4.11 (3.21, 5.00)	23.82 (23.20, 24.44)
DD	1 (1, 1)	1 (1, 1)	0 (0, 0)	0.87 (0.85, 0.89)
CIRI	2 (2, 2)	1 (1, 1)	0.51 (0.42, 0.60)	1.65 (1.62, 1.68)
Polity	9.01 (8.93, 9.09)	7.91 (7.73, 8.10)	-5.33(-5.88, -4.77)	6.96 (6.70, 7.22)

^aThese figures exclude the countries denoted as outliers.

Table 2 Interpretation of the overall cluster structures

Index	<i>Relative interpretation</i>			<i>Benchmark interpretation</i>		
	<i>Cluster 1</i> (<i>n</i> = 1084)	<i>Cluster 2</i> (<i>n</i> = 238)	<i>Cluster 3</i> (<i>n</i> = 199)	<i>Cluster 1</i> (<i>n</i> = 1084)	<i>Cluster 2</i> (<i>n</i> = 238)	<i>Cluster 3</i> (<i>n</i> = 199)
FH	Low	Low/average	High	Free	Free/partly free	Not free
Vanhanen ^a	Average	Average	Low	—	—	—
DD	High	High	Low	Democracy	Democracy	Dictatorship
CIRI	High	Average	Low	Full rights	Some rights	No rights
Polity	High	Average/high	Low	Democracy	Democracy	Dictatorship

^aVanhanen's index of democratization does not provide cutoffs for classifying scores into low-, middle-, and high-democracy groups.

the increment value is equal to 0.857 $[(7 - 1)/7]$. Adding this value to the scale points yields the following association: a score of 1 on FH has a corresponding interval of (1, 1.857), a score of 2 has a corresponding interval of (1.858, 2.714), and so forth. Since FH interprets scores of 1 and 2 as free countries, mean values less than 2.714 denote free countries. Applying this method to all indices yielded the following transformations: for FH, means less than 2.714 denote free countries, means greater than 5.286 denote not free countries, and means between 2.715 and 5.286 denote partly free countries; for DD, means less than 0.499 denote dictatorships, whereas means greater than or equal to 0.50 denote democracies; for CIRI, means less than 0.666 denote no political rights, means greater than 1.334 denote full political rights, and means between 0.667 and 1.333 denote some political rights; and for Polity, means less than -5.238 denote autocracies, means greater than 5.239 denote democracies, and means between -5.239 and 5.239 denote anocracies.

Table 2 reveals that Cluster 1 is composed of cases identified as democracies, Cluster 3 is composed of cases identified as dictatorships, and Cluster 2 is composed of cases identified as being in regime transition (i.e., anocracies). Examination of cases included in each cluster (see supplementary materials) also indicates that the outliers identified for both Latin America and Europe were included almost exclusively in either Cluster 2 or Cluster 3 in the second HCA conducted for each data set. Whereas the individual measures agreed in their classification of democracies, the scores assigned to cases classified as anocracies or dictatorships were often very

different. For example, Mexico in 1997 was given a score of 3.5 (partly free) by FH, a 2 (full political rights) by CIRI, a 6 (democracy) by Polity, but a score of 0 (dictatorship) by DD and a score of 19 by Vanhanen's index. In 2003, Russia was classified as "not free" (score of 6.5) by FH, a democracy (score of 6) by Polity, a dictatorship (score of 0) by DD, a country with some political rights (score of 1) by CIRI, and received a score of 22.5 from Vanhanen's index. These results indicate that using just one individual measure of democracy can potentially lead to problems for scholars interested in correctly classifying countries by their regime type.

Examination of Table 1 reveals an interesting finding; namely, the existence of mean cluster differences and their corresponding CIs between Latin America and Europe. Although the three cluster structure is produced in both instances, the difference between democracy (Cluster 1) and anocracy (Cluster 2) is more evident in Europe than in Latin America. This finding is not utterly surprising since half of the European countries have a long tradition of democracy, whereas most of the countries in Latin America underwent democratization and democratic consolidation during the period under study. Additionally, most former communist countries had more consistent paths toward democratic consolidation compared to their Latin American counterparts. Nonetheless, this difference may be important to scholars conducting comparative studies on democratization that employ cases from both regions. Moreover, it is very likely that such a difference may also be present in other regions in the world.

4.2 Generalizing Beyond the Current Study

Though informative, the DCC index only classifies twenty-four countries in the Americas and thirty-nine countries in Europe over a three-decade period. Nevertheless, others may wish to classify these countries for different time periods or to classify other countries in regions not covered by the present study. This can be done by utilizing a set of linear discriminant equations (one for each cluster) to compute a coefficient from four of the democracy indices.⁸ The discriminant equation with the highest coefficient indicates the cluster to which the case should be classified (Tabachnick and Fidell 2001). The discriminant equations employ the results produced by the discriminant analysis conducted on the merged Latin American and European data set ($n = 1693$), which correctly classified 95% of cases:

$$C_1 = -44.64 + 1.28(\text{Polity}) + 0.55(\text{Vanhaven}) + 10.11(\text{FreedomHouse}) + 22.58(\text{CIRI})$$

$$C_2 = -31.28 + 1.93(\text{Polity}) + 0.50(\text{Vanhaven}) + 10.48(\text{FreedomHouse}) + 10.11(\text{CIRI})$$

$$C_3 = -35.60 + 0.19(\text{Polity}) + 0.58(\text{Vanhaven}) + 10.95(\text{FreedomHouse}) + 15.75(\text{CIRI}).$$

For example, if one were interested in classifying Serbia in 2008, a country not included in this study, one could use the scores assigned by each individual index and the discriminant equations to classify the country regime. For example, in 2008, Polity assigned Serbia a score of 8, Vanhanen a score of 24.3, FH a score of 2.5, and CIRI a score of 2. Inputting these numbers, one obtains the following scores: $C_1 = 49.4$, $C_2 = 42.7$, and $C_3 = 38.95$. Following the recommended rule, Serbia in 2008 should be classified in Cluster 1 (democracy). In fact, in 2008, the European Union adopted the European Partnership with Serbia, recognizing the significant progress toward democracy since the removal from power of Slobodan Milosevic in 2000 (European Commission 2008). Hence, in the absence of a classification provided by the present study, one can still determine a country's regime type using these equations. A note of caution is warranted for countries outside the Americas or Europe. Since the equations were generated based on the regimes prevalent in these two regions, further research is needed to determine whether the predicted DCC scores generated by these equations generalize beyond the countries included in this study.

⁸ The DD index was excluded from this analysis because it failed the tolerance test for the discriminant function analysis due to the fact that a dichotomous variable cannot be used to predict a trichotomous variable. For a more detailed explanation of the discriminant analysis, interested readers are directed to Tabachnick and Fidell (2001).

4.3 *The Sensitivity and Specificity of Existing Measures of Democracy*

According to recent research (Treier and Jackman 2008; Pemstein, Mesere, and Melton 2010), existing measures of democracy are more susceptible to measurement error than a composite of these indices. To investigate whether the DCC index, reported in supplementary appendices, represents a significant improvement over single democracy indices, we performed a series of contingency table analyses comparing each democracy index to the DCC index. It is important to note that since a contingency table is a nonparametric method that relies on frequency counts rather than variance estimates, its results are not susceptible to violations of distributional assumptions or the independence of errors associated with the use of nested and time-series data. To perform the contingency table analysis, the democracy indices (except DD index) were recoded as follows: FH and Polity were recoded using the equal-sized intervals presented earlier (see Section 4.1). For example, cases with a FH score at or below 2.714 were coded 1 (democracy), cases with a score between 2.715 and 5.286 were coded 2 (anocracy), and cases with a score above 5.286 were coded 3 (dictatorship). Since the DD index is a dichotomous measure, no recoding was needed.

Vanhanen's index of democratization and Pemstein et al.'s (2010) UDS index do not provide any information on how to classify regimes except to indicate that low scores represent dictatorships, whereas high scores represent democracies. Hence, these indices were recoded so as to match the proportion of cases found in each of the three clusters produced by the DCC measure; namely, 64.74% of cases were classified as democracies, 17.09% as anocracies, and 18.07% as dictatorships. Cases that received a score at or below 11 on Vanhanen's index were assigned to Cluster 3 (dictatorship), cases with a score of 17.8 or higher were assigned to Cluster 1 (democracy), and the remaining cases were assigned to Cluster 2 (anocracy). Thus, 64.7% of all cases were deemed democracies, 17.2% anocracies, and the remaining 18.1% dictatorships. Since the UDS data set did not include scores after 2008 and no missing value imputation was performed on this data set, the proportion of cases that fell in each of the three clusters was slightly smaller ($n = 1577$). Namely, 65.1% of cases were classified as democracies, 15.9% as anocracies, and the remaining 19% as dictatorships. Employing these cutoffs to recode Vanhanen's index and the UDS, the overlap between each of these indices and the DCC was artificially maximized. That is, we selected these cutscores to maximize the degree of overlap between these measures and the DCC. It is highly probable that if cutscores were available for these measures, the consistency of their classification in comparison to the DCC would be substantially lower.

Although numerous statistics can be computed from contingency tables, sensitivity and specificity statistics are the most appropriate for investigating the performance of an index. Originating in medical research, sensitivity and specificity are used to evaluate the accuracy of diagnostic tests (Weinstein, Obuchowski, and Lieber 2005). Sensitivity refers to the ability of a test to correctly identify the presence of disease, whereas specificity refers to the ability of a test to correctly identify the absence of disease. Hence, a test with 100% sensitivity correctly identifies all the patients who have the disease (i.e., true positives), whereas a test with 100% specificity correctly identifies all the patients who do not have the disease (i.e., true negatives) (Altman and Bland 1994; Lalkhen and McCluskey 2008). Although a test can have perfect sensitivity and specificity, perfect accuracy rarely occurs in practice. Therefore, both statistics need to be interpreted together. If a test has 100% sensitivity but low specificity, it will lead to a high number of false positives. Conversely, if a test has 100% specificity but low sensitivity, the results will indicate a high number of false negatives. Hence, considering both sensitivity and specificity provides the most accurate measure of a test's diagnostic ability.

For the purpose of the present study, sensitivity and specificity were employed to estimate the accuracy of each index of democracy to correctly classify cases by the three regime types identified by DCC.⁹ Hence, they were employed as instruments for assessing the validity (Bjelland et al. 2002)

⁹ Since the *actual* classification of a case must be inferred (because it is unknown) and the HCA results demonstrated that the three-cluster solution accounted for almost all the variance of five democracy indices, the analyses reported in this section treated the DCC index as the "gold standard" for identifying cluster membership and compared all other indices to its results. By this comparison, however, we do not wish to imply that the DCC contains no measurement error. Its dependence on sample data most assuredly results in some error. However, the use of a composite score and

of the democracy measures. Ideally, a test should attain high levels for both statistics, else it will inflate Type II error rate (low sensitivity) or Type I error rate (low specificity). Herein, sensitivity was computed from the ratio of the number of cases correctly identified as belonging to a particular cluster divided by the total number of cases belonging to that cluster. In contrast, specificity was computed from the ratio of the number of cases correctly identified as not belonging to a particular cluster divided by the total number of cases not belonging to that cluster. For example, based on the frequency counts (see supplementary material) for FH by DCC indices, sensitivity is equal to $930/(930 + 166) = 0.849$ (i.e., FH index only correctly identifies 84.9% of the 1096 cases identified as democracies by the DCC) and specificity is equal to $426/(426 + 171) = 0.975$ (i.e., FH index correctly identifies 97.5% of the cases identified as nondemocracies by the DCC). The sensitivity and specificity for Cluster 2 (anocracy) were 44.3% and 78.5%, respectively, whereas the sensitivity and specificity for Cluster 3 (dictatorship) were 50.7% and 99.5%, respectively. The overall sensitivity and specificity for an index can be computed by the weighted average of their corresponding statistics for the three clusters. Hence, the overall sensitivity of the FH index is $84.9(1096/1693) + 44.3(291/1693) + 50.7(306/1693) = 71.7\%$. Similarly, the overall specificity for FH index can be shown to equal 94.6%. As a benchmark, one can compare these statistics to the sensitivity (33.3%) and specificity (66.7%) one would attain if cases were assigned to clusters at random. In the case of the FH index, its sensitivity is more than twice the rate one would observe by chance alone, whereas its specificity is almost one-third higher than the rate one would observe by chance alone.

The sensitivity and specificity statistics reported in Table 3 provide a means by which the accuracy of the democracy indices can be compared. Examination of these statistics revealed that Vanhanen and Polity were the least accurate indices overall and were among the least accurate measures within each cluster. In contrast, the predicted DCC¹⁰ was the most accurate index overall, followed by CIRI. Although the predicted DCC index attained the highest level of accuracy across the three clusters, it was not always the most accurate measure within a cluster. With respect to Clusters 1, 2, and 3, the most accurate indices were the predicted DCC, CIRI, and DD, respectively. Although the DD index is a dichotomous measure and cannot identify anocracies, it improves the classification ability of the DCC, particularly pertaining to its ability to correctly identify dictatorships. The UDS index was slightly more accurate than the predicted DCC in classifying dictatorships, which may have been a function of how the UDS was recoded. It is probable that if different cutoffs were employed, the UDS would have underperformed the predicted DCC. Consequently, researchers who want to focus exclusively on democracies or

Table 3 Sensitivity and specificity for indices of democracy

<i>Index</i>	<i>Democracy</i>		<i>Anocracy</i>		<i>Dictatorship</i>		<i>Overall</i>	
	<i>Sensitivity</i>	<i>Specificity</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Sensitivity</i>	<i>Specificity</i>
FH	84.9	97.5	44.3	78.5	50.7	99.5	71.7	94.6
Vanhanen ^a	78.7	61.5	24.7	84.6	68.6	92.7	67.5	71.1
DD	100.0	51.3	0	100.0	100.0	100.0	82.8	68.5
CIRI	100.0	92.5	90.0	90.5	41.8	97.9	87.8	93.1
Polity	97.7	51.9	15.1	90.2	49.7	99.9	74.8	67.2
UDS	86.6	75.5	35.5	87.8	87.3	97.0	78.6	81.5
Predicted DCC	99.9	94.0	93.5	98.0	79.1	98.6	95.0	95.5

^aVanhanen's index of democratization and UDS do not provide cutoffs for classifying scores into democracy, anocracy, and dictatorship. Hence, the cutoffs used were the percentage of cases classified into these three categories by the DCC, thus maximizing the overlap between the two indices and the DCC.

the fact that a three-cluster solution nearly explained 100% of the data suggests that the measurement error contained in the DCC is relatively small compared to extant measures.

¹⁰ The predicted DCC is the score produced by the discriminant analysis.

anocracies should use the predicted DCC score to identify the cases that fall within these groups. Alternatively, researchers who want to focus exclusively on dictatorships should identify cases based on a DD score of 0 (dictatorship). However, these guidelines should only be followed when the country or the time period fall outside the cases covered in this study because, absent knowledge of the true state of a country for a given year, the best measure of its latent class is provided by the DCC index, not the predicted DCC index.

5 Discussion

The results of the HCA suggest that “democracy” may be best thought of as a latent class variable rather than as a continuous variable. This conclusion is supported by the fact that a three-cluster HCA solution fit the observed data better than one with many branches and small Euclidean distances separating each cluster. Consistent with conventional thought, the three-cluster structure was clearly representatives of democracies, anocracies, and dictatorships. Although the three-cluster solution identified in Latin America was confirmed in the European case (as well as in the combined data sets), further research is needed to determine whether this result generalizes to other regions. The results of this study partially explain why the Pemstein et al. (2010) study found it difficult to distinguish between countries when the UDS point estimates fall in the same tertile. Interestingly, although the UDS relied upon four of the five indices included in the DCC, it did not assign cases to the three clusters with great accuracy. This suggests that the diagnostic ability of the UDS index may have been compromised by violations of the underlying assumptions of the analysis employed or the weighting scheme used to combine indices.

This study demonstrated that the predicted DCC score is the most accurate overall estimate of the latent class one can obtain without replicating the HCA for new data, particularly if researchers are interested in democracies or transitioning countries. Nevertheless, the analysis also showed that the predicted DCC is not always the most accurate measure if one wishes to focus on dictatorships. In those cases, researchers may be better off using the DD index. It is important to note that these recommendations should only be followed when the country or year of interest fall outside the cases included in the present study. For cases included in the study, the DCC remains the most accurate measure of regime type.

A serious limitation of efforts to use existing measures and indices to operationalize “democracy” or assess its quality should be underscored. Most existing measures, including four of the five indices we used, rely primarily on subjective interpretation of regimes. Although these are subject to rigorous coder reliability tests, they nevertheless impose significant constraints on cross-national research that must be acknowledged. Typically, coders are only asked to code a single country, rather than a basket of countries, which, in turn, increases the likelihood that coders will use different operational standards from one another. This seems to be particularly a problem in transitional cases, as evidenced by the high number of outliers identified in Latin America, which were classified by the DCC as either anocracies or dictatorships. During transition years, individual democracy measures are more likely to assign conflicting scores than in periods of democratic consolidation or dictatorship. Hence, particular caution should be used in employing any one individual measure to classify countries in regime transition.

For example, if one were interested in studying Romania in 1990, the country was classified as a democracy by the DD index (i.e., score of 1), a dictatorship by the FH index (i.e., score of 5.5), and an anocracy by the Polity, Vanhanen, and CIRI (i.e., scores of 5, 14.7, and 1, respectively). However, the DCC classified Romania in 1990 as an anocracy. Scholars familiar with Romania know it was what Linz and Stepan (1996) called a “totalitarianism-cum-sultanism” regime under the leadership of Nicolae Ceaușescu and that his removal from power in December 1989 was not accompanied by the downfall of the former communist nomenclature. Hence, the leader in 1990 was a former high-ranking member of the Romanian Communist Party, Ion Iliescu, and his political party (National Salvation Front) was characterized by the predominance of ex-party cadres and Securitate (Secret Service Police) (Hollis 1999). The control and corruption exhibited by these leaders made the transition to democracy slower and more difficult compared to other CEE neighbors (Ristei 2010; Gugiu 2012). This is an illustrative case where various democracy indices assign

contradictory scores, leaving scholars to choose—based on their knowledge of the country or others' recommendations—which score may be best to use in their research.

6 Limitations of Current Study and Future Directions

The DCC employed five measures of democracy that provided data for the entire time period and all countries included in this study. Arguably, the composite variable would have been even stronger if one had access to the disaggregated level indicators used to create the five indices employed in this study or to variables with a more nuanced level of measurement. Unfortunately, such information is not publicly available, despite our efforts to gather it. Moreover, even if more nuanced variables were developed, it is not clear how one would rank-order the multitude of dimensions that underlie democracy, not to mention quantify the distance between these dimensions on the regime-type continuum, assuming the latent variable is indeed continuous, without resorting to assumptions that are either untenable or cannot be empirically validated. Therefore, we contend that by building off the strengths of existing measures of regime type, the DCC and the predicted DCC are the most accurate measures of regime type presently available. That said, further research is needed to investigate the validity and reliability of this new index.

One line of future research may be to expand upon the measure of the “quality of democracy” proposed by Altman and Pérez-Liñán (2002), which incorporates three dimensions of democracy: civil rights, effective participation, and effective competition. Although an assessment of their measure is outside the scope of this article, our preliminary analysis indicated potential problems with this measure as well. Specifically, the inclusion of their proposed indicator of effective competition in the HCA analysis had no impact on the cluster structure and so was excluded from the list of variables that comprised the DCC index. Additionally, the validity and replicability of their measure has yet to be investigated. We believe the DCC index will aid researchers in moving forward with existing measures, provided researchers treat regime type as a discrete variable.

References

- Adcock, Robert, and David Collier. 2001. Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review* 95(3):529–46.
- Altman, David, and Anibal Perez-Linan. 2002. Assessing the quality of democracy: Freedom, competitiveness and participation in eighteen Latin American countries. *Democratization* 9(2):85–100.
- Altman, Douglas G., and J. Martin Bland. 1994. Diagnostic tests 1: Sensitivity and specificity. *British Medical Journal* 308:1552.
- Alvarez, Michael, Jose Antonio Cheibub, Fernando Limongi, and Adam Przeworski. 2011. Classifying political regimes. *Studies in Comparative Political Development* 31(2):1–37.
- Arat, Zehra F. 1991. *Democracy and human rights in developing countries*. Boulder, CO: Lynne Rienner.
- Beetham, David. 2004. Freedom as the foundation. *Journal of Democracy* 15(4):61–75.
- Bermeo, Nancy. 2009. Does electoral democracy boost economic equality? *Journal of Democracy* 20(4):21–35.
- Bjelland, Ingvar, Alv A. Dahl, Tone Tangen Haug, and Dag Neckelmann. 2002. The validity of the hospital anxiety and depression scale: An updated literature review. *Journal of Psychosomatic Research* 52(2):69–77.
- Bogaards, Matthijs, and Françoise Boucek. 2010. *Dominant political parties and democracy: Concepts, measures, cases and comparisons*. New York: Routledge.
- Bollen, Kenneth A. 2001. *Cross-national indicators of liberal democracy, 1950–1990*. 2nd ed. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.
- . 1990. Political democracy: Conceptual and measurement traps. *Studies in Comparative International Development* 25(1):7–24.
- Bollen, Kenneth A., and Pamela Paxton. 2000. Subjective measures of liberal democracy. *Comparative Political Studies* 33(1):58–86.
- Bowman, Kirk, Fabrice Lehoucq, and James Mahoney. 2005. Measuring political democracy: Case expertise, data adequacy, and Central America. *Comparative Political Studies* 38(8):939–70.
- Calinski, T., and J. Harabasz. 1974. A dendrite method for cluster analysis. *Communications in Statistics* 3(1):1–27.
- Carbone, Giovanni. 2009. The consequences of democratization. *Journal of Democracy* 20(2):123–37.
- Centellas, Miguel. 2011. Revisiting assessing the quality of democracy: Measuring democratic competitiveness and participation in Latin America since the “Left Turn.” Paper presented at the annual meeting of the Midwest Political Science Association. Chicago, IL.

- Cheibub, Jose Antonio, Jennifer Gandhi, and James Raymond Vreeland. 2009. Democracy and dictatorship revisited. http://netfiles.uiuc.edu/cheibub/www/DD_page.html (accessed January 27, 2012).
- . 2010. Democracy and dictatorship revisited. *Public Choice* 143(1–2):65–101.
- Cingranelli, David L., and David L. Richards. 2011. The Cingranelli-Richards Human Rights Dataset. *CIRI Human Rights Data Project*. <http://ciri.binghamton.edu/> (accessed January 27, 2012).
- Collier, David, and Robert Adcock. 1999. Democracy and dichotomies: A pragmatic approach to choices about concepts. *Annual Review of Political Science* 2:537–65.
- Collier, David, and Steven Levitsky. 2009. Democracy: Conceptual hierarchies in comparative research. In *Concepts and methods in the social sciences: The tradition of Giovanni Sartori*, eds. David Collier and John Gerrings, 269–88. London: Routledge.
- Coppedge, Michael. 2001. Political Darwinism in Latin America's lost decade. In *Political parties and democracy*, eds. Larry Diamond and Richard Gunther, 173–205. Baltimore, MD: Johns Hopkins University Press.
- Coppedge, Michael, and Wolfgang H. Reinicke. 1991. Measuring polyarchy. In *On measuring parties and democracy*, ed. Alex Inkeles, 47–68. New Brunswick, Canada: Transaction Press.
- Coppedge, Michael, and John Gerring. 2011. Conceptualizing and measuring democracy: A new approach. *Perspectives on Politics* 9(2):247–67.
- Dahl, Robert. 1982. *Dilemmas of pluralist democracy: Autonomy v. Control*. New Haven, CT: Yale University Press.
- . 1971. *Polyarchy: Participation and opposition*. New Haven, CT: Yale University Press.
- Duda, Richard O., and Peter E. Hart. 1973. *Pattern classification and scene analysis*. New York: Wiley.
- European Commission. 2008. Enlargement. February 18, 2008. *European Commission*, http://ec.europa.eu/enlargement/countries/detailed-country-information/serbia/index_en.htm (accessed November 13, 2012).
- Freedom House. 2011. Freedom in the World Country Ratings 1972–2011. www.freedomhouse.org/sites/default/files/inline_images/FIWAAllScoresCountries1973-2011.xls (accessed December 14, 2011).
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2009. *The elements of statistical learning: Data mining, inference, and prediction*. 2nd ed. New York: Springer.
- Gasiorowski, Mark J. 1996. An overview of the political regime change data set. *Comparative Political Studies* 29(4):469–483.
- Greene, Kenneth. 2007. *Why dominant parties lose: Mexico's democratization in comparative perspective*. Cambridge, UK: Cambridge University Press.
- Gugiu, Mihaiela Ristei. 2012. EU enlargement and anticorruption: Lessons learned from the Romania case. *Journal of European Integration* 34(5):429–46.
- Gugiu, Mihaiela Ristei, and Miguel Centellas. 2013. *Replication data for: The Democracy Cluster Classification Index*. <http://hdl.handle.net/1902.1/20393>, IQSS Dataverse Network [Distributor] V1 [Version] (accessed February 28, 2013).
- Hadenius, Axel. 1992. *Democracy and development*. Cambridge, UK: Cambridge University Press.
- Hadenius, Axel, and Jan Teorell. 2007. Pathways from authoritarianism. *Journal of Democracy* 18(1):143–56.
- Hollis, Wendy. 1999. *Democratic consolidation in Eastern Europe: The influence of the communist legacy in Hungary, the Czech Republic, and Romania*. New York: Columbia University Press.
- Horton, N. J., and K. P. Kleinman. 2007. Much ado about nothing: A comparison of missing data methods and software to fit incomplete regression models. *American Statistician* 61(1):79–90.
- Huntington, Samuel. 1991. *The third wave: Democracy in the late twentieth century*. Norman: University of Oklahoma Press.
- Johnson, Richard A., and Dean W. Wichern. 1998. *Applied multivariate statistical analysis*. 4th ed. Upper Saddle River, NJ: Prentice Hall.
- Karatnycky, Adrian. 2005. Ukraine's orange revolution. *Foreign Affairs* 84(2):35–52.
- Lalkhen, Abdul Ghaliq, and Anthony McCluskey. 2008. Clinical tests: Sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care & Pain* 8(6):221–23.
- Levitsky, Steven, and Lucan A. Way. 2002. The rise of competitive authoritarianism. *Journal of Democracy* 13(2):51–65.
- Linz, Juan J., and Alfred Stepan. 1996. *Problems of democratic consolidation: Southern Europe, South America, and post-communist Europe*. Baltimore, MD: Johns Hopkins University Press.
- Lipset, Seymour M. 2000. The indispensability of political parties. *Journal of Democracy* 11(1):48–55.
- Maravall, Jose Maria, and Adam Przeworski. 2003. *Democracy and the rule of law*. Cambridge, UK: Cambridge University Press.
- Marshall, Monty G., Ted Rebot Gurr, and Keith Jagers. 2011. INSCR Data Page. *Polity IV Project*. <http://www.systemicpeace.org/inscr/p4manualv2009.pdf> (accessed January 3, 2012).
- Milligan, Glenn W. 1980. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* 45(3):325–42.
- Milligan, Glenn W., and Martha C. Cooper. 1988. A study of standardization of variables in cluster analysis. *Journal of Classification* 5(2):181–204.
- . 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50(2):159–79.
- Mulaik, Stanley A. 2010. *Foundations of factor analysis*. 2nd ed. New York: Taylor and Francis Group.
- Munck, Gerardo L., and Jay Verkuilen. 2002. Conceptualizing and measuring democracy: Evaluating alternative indices. *Comparative Political Studies* 35(1):5–34.
- O'Donnell, Guillermo. 2004. Why the rule of law matters. *Journal of Democracy* 15(4):32–46.

- Pemstein, Daniel, Stephen A. Mesere, and James Melton. 2010. Democratic compromise: A latent variable analysis of ten measures of regime type. *Political Analysis* 18(4):426–49.
- Pridham, Geoffrey. 2011. Ukraine, the European Union and the democracy question. *Romanian Journal of European Affairs* 11(4):18–33.
- Przeworski, Adam, Michael Alvarez, Jose Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and development: Political regimes and economic well-being in the world, 1950–1990*. Cambridge, UK: Cambridge University Press.
- Reich, Gary. 2002. Categorizing political regimes: New data for old problems. *Democratization* 9(4):1–24.
- Ristei, Mihaela. 2010. The politics of corruption: Political will and the rule of law in post-communist Romania. *Journal of Communist Studies and Transition Politics* 26(3):341–62.
- Rueschemeyer, Dietrich. 2004. Addressing inequality. *Journal of Democracy* 15(4):76–90.
- SAS Institute. 2010. *SAS user's guide*. 2nd ed. Cary, NC: SAS Institute.
- . 2012a. Multiple imputation for missing data. *SAS*. <http://support.sas.com/rnd/app/da/new/dami.html> (accessed January 28, 2012).
- . 2012b. The cluster procedure. *SAS*. http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_cluster_sect002.htm (accessed February 1, 2012).
- Schmitter, Philippe, and Terry Lynn Karl. 1991. What democracy is... and is not. *Journal of Democracy* 2(3):75–88.
- Schumpeter, Joseph A. 1975. *Capitalism, socialism, and democracy*. New York: Harper & Row.
- Smith, Peter H., and Melissa R. Zeigler. 2008. Liberal and illiberal democracy in Latin America. *Latin American Politics & Society* 50(1):31–57.
- Sorensen, Georg. 2008. *Democracy and democratization: Process and prospect in a changing world*. 3rd ed. Boulder, CO: Westview Press.
- Tabachnick, Barbara E., and Linda S. Fidell. 2001. *Using multivariate statistics*. 4th ed. Boston, MA: Allyn & Bacon.
- Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. 2006. *Introduction to data mining*. Boston: Addison-Wesley.
- Treier, Shawn, and Simon Jackman. 2008. Democracy as a latent variable. *American Journal of Political Science* 52(1):201–17.
- Vachudova, Milada Ana. 2005. *Europe undivided: Democracy, leverage, and integration after communism*. Oxford, UK: Oxford University Press.
- Vanhanen, Tatu. 2000. A new dataset for measuring democracy, 1810–1998. *Journal of Peace Research* 37(2):251–65.
- . 2003. *Democratization: A comparative analysis of 170 countries*. New York: Routledge.
- . 2011. FSD1289 measures of democracy 1810–2010. *Finish Social Science Data Archive*. <http://www.fsd.uta.fi/english/data/catalogue/FSD1289/meF1289e.html> (accessed January 9, 2012).
- Ward, Joe H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58(301):236–44.
- Weinstein, Susan, Nancy A. Obuchowski, and Michael L. Lieber. 2005. Fundamentals of clinical research for radiologists. *American Journal of Roentgenology* 184:14–19.
- Wong, Joseph, and Edward Friedman. 2008. *Political transitions in dominant party systems: Learning to lose*. New York: Routledge.