Model-Based Cluster Analysis of Democracies

Jaewon Jang

University of South Carolina

Department of Political Science

David B. Hitchcock

University of South Carolina

Department of Statistics *

September 9, 2011

Abstract

We apply model-based cluster analysis to data concerning types of democracies, creating an instrument for typologies. Noting several advantages of model-based clustering over traditional clustering methods, we fit a normal mixture model for types of democracy in the context of the majoritarian-consensus contrast using Lijphart's (1999) data on ten variables for 36 democracies. The model for the full period (1945-1996) finds four types of democracies: two types representing a majoritarian-consensus contrast, and two mixed ones lying between the extremes. The four-cluster solution shows that most of the countries have high cluster membership probabilities, and the solution is found to be quite stable with respect to possible measurement error in the variables included in the model. For the recent-period (1971-1996) data, most countries

remain in the same clusters as for the full-period data.

Keywords: Model-based clustering, Multivariate analysis, Political applications, Typology, Types of democracy.

Running Title: Model-Based Clustering of Democracies

1 Introduction

Cluster analysis has found wide application across disciplines. Within the social sciences, cluster analysis has appeared frequently in sociology but not in political science and economics (Ahlquist and Breunig, 2009). This article applies model-based cluster analysis to political data as an instrument for building typologies. Typological theorizing has a distinguished tradition in social sciences (George and Bennett, 2005; Elman, 2005). Typology, defined by Lehnert (2007, p. 62) as "a theoretically or empirically derived concept which systematically orders complex phenomena according to a limited number of attributes," has served as "a conceptual tool to simplify and order complex social phenomena" (Lehnert, 2007, p. 62). It can be utilized to "clarify similarities and differences among cases to facilitate comparisons" (George and Bennett, 2005, p. 233).

In political science, some scholarly works develop typologies for the purpose of descriptive inferences. For example, Andeweg and Thomassen (2005) propose a typology of modes of representation, and they capture aspects of the relationship between voters and representatives. Ghunther and Diamond (2003) develop a party typology of the contemporary political parties. The works of Collier and Adcock (1999) and Alvarez et al. (1996) develop typologies concerning types of democracies. In contrast, some scholars use typologies as a tool for causal inferences, to challenge the questions

2

such as what affects political orders and their performance. For example, Lijphart (1999) advances a new typology (Majoritarian-Consensus models) to investigate how the performance of a democracy depends on its institutional characteristics.

The conventional approach in typologies is to arrange the variables' values in rows and columns to construct an associated property space, and each cell of the space captures a possible combination of values of the variables (Lazarsfeld and Barton, 1951, p. 169). For example, the simplest property space has four cells in a 2×2 table consisting of two variables, each of which has two values, and each cell constitutes a type in that typology. However, including too many variables and/or values for variables might result in an intractably large number of types or too many empty cells in the property space, especially for small-sample studies. To address these problems, Lehnert (2007, p. 70–72) provides some practical guidelines and Elman (2005, p. 300–308) proposes "compression (or reduction) of the number of cells in a property space." This article finds that cluster analysis can deal with this problem without reducing property space or data dimension.

Since cluster analysis identifies groups of observations that are cohesive and separated from other groups (Hartigan, 1975; Jain and Dubes, 1988; Jain et al., 1999; Ghosh, 2003; Zhong and Ghosh, 2003; Fraley and Raftery, 2002), it can be utilized to create types (or clusters) of political systems such that objects in one cluster are very similar while objects in different clusters are quite distinct. However, its application to political science has been as yet limited. Wolfson et al. (2004) employ Ward's (1963) method to find a relationship among economic and political variables. Webb (2008) and Chae (2010) introduce hierarchical clustering to find groups in political attitude. Numerous works employ cluster analysis to investigate similarities and differences among welfare regimes (Franzoni, 2008; Lee and Ku, 2007; Gough and Sharkh, 2010;

Rudra, 2007). However, most of them draw on traditional approaches such as hierarchical and K-means clustering methods. Only recently, Ahlquist and Breunig (2009) and Spirling and Quinn (2010) provide rigorous discussion of model-based cluster analysis for investigating types of welfare regimes and legislative voting behavior, respectively. Recognizing several advantages of model-based clustering over traditional clustering methods, we employ model-based clustering using variables characterizing types of democracies in subsequent sections.

2 Model-based Cluster Analysis

There has been an explosion in the theory and application of cluster analysis. Traditional cluster analysis frequently used in practice has been founded on sensible (yet heuristic) algorithms. In recent years, it has become common to formulate cluster analysis methodology using probability models (see Bock, 1996; Fraley and Raftery, 2002). There are three major classes of clustering methods: From oldest to newest, they are hierarchical, partitioning, and model-based methods. Hierarchical clustering methods proceed by either a series of successive mergers or a series of successive divisions to optimize some criterion at each stage of the algorithm. Some criteria used in popular hierarchical algorithms include the sum of within-group sums of squares (Ward, 1963) and the shortest distance between groups. Another common class of methods is iterative partitioning methods, in which data objects are partitioned into K (often specified *a priori*) groups by shifting objects across groups until the value of a particular criterion fails to improve. However, as Fraley and Raftery (2002, p. 611) note, these two traditional methods lack "systematic guidance ... for solving basic practical questions in cluster analysis, such as how many clusters there are, which clustering method should be used, and how outliers should be handled [and also pre-

4

clude] the possibility of formal inference." To alleviate this, model-based clustering methods have been developed.

Since the observed data set may be considered a sample from some population of interest, a number of statisticians have suggested methods involving statistical models (Everitt, 1993). Compared to the traditional approach, model-based methods offer better interpretability since the resulting submodel for each cluster directly characterizes that cluster (Zhong and Ghosh, 2003). Model-based clustering is based on the idea that data are generated by a mixture of underlying probability distributions in which each component represents a different group or cluster. Thus we need to model each of the subpopulations separately and the overall population as a mixture of these subpopulations, using finite mixture models (Raftery and Dean, 2006).

Fraley and Raftery (2002) note that finite mixture models have long been suggested for the purpose of clustering data (Edwards and Cavalli-Sforza, 1965; Day, 1969; Scott and Symons, 1971; Duda and Hart, 1973; Binder 1978). More recent research (McLachlan and Basford, 1988; Banfield and Raftery, 1993; Cheeseman and Stutz, 1995) has shown that such mixture models offer a unified statistical approach to practical clustering questions such as the number of clusters, choice of clustering method, model choice, and variable selection.

2.1 Finite Normal Mixture Models

We follow the standard notation of Fraley and Raftery (2002) in this section. Suppose we observe independent d-dimensional multivariate observations $\mathbf{y}_1$, ..., $\mathbf{y}_n$. A mixture model with G components has likelihood

$$L_{MIX}(\theta_1, \ldots, \theta_G | \mathbf{y}) = \prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k f_k(\mathbf{y}_i | \theta_k), \tag{1}$$

where f_k is the density of the kth mixture component (indexed by parameters θ_k), and τ_k represents the probability of an observation belonging to the kth component, where $\tau_k \geq 0$ and $\sum_{k=1}^{G} \tau_k = 1$ (Fraley and Raftery, 2002). Gaussian models are widely used for multidimensional data (McLachlan and Basford, 1988; Banfield and Raftery, 1993), although for specialized data structures, other models such as multivariate Bernoulli, multinomial, and von Mises-Fisher models may be used. In this clustering context we let f_k be the multivariate normal density ϕ_k, having mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$,

$$\phi_k(\mathbf{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \equiv \frac{\exp\left\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_k)\right\}}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_k)}}. \tag{2}$$

Clusters of data following such a mixture density tend to be centered at the multidimensional means $\boldsymbol{\mu}_k$, with shape, volume, and orientation of the clusters being determined by the parameters of the covariance matrices $\boldsymbol{\Sigma}_k$, which may also induce cross-cluster conditions (Fraley and Raftery, 2002). The number of parameters requiring estimation ranges from $d(d+1)/2$ for constant $\boldsymbol{\Sigma}_k$ to $G(d(d+1)/2)$ for unrestricted $\boldsymbol{\Sigma}_k$.

Banfield and Raftery (1993) used the eigenvalue decomposition

$$\boldsymbol{\Sigma}_k = \lambda_k \mathbf{D}_k \mathbf{A}_k \mathbf{D}_k^T, \tag{3}$$

where $\mathbf{D}_k$ is the orthogonal matrix of eigenvectors, $\mathbf{A}_k$ is a diagonal matrix whose elements are proportional to the eigenvalues such that $\mathbf{A}_k = \text{diag}\{\alpha_{1k}, \alpha_{2k}, \ldots, \alpha_{dk}\}$ and $1 = \alpha_{1k} \geq \alpha_{2k} \geq \cdots \geq \alpha_{dk} > 0$, and λ_k is the first eigenvalue of $\boldsymbol{\Sigma}_k$. Banfield and Raftery (1993) suggested considering λ_k, $\mathbf{A}_k$, and $\mathbf{D}_k$ as independent parameters (possibly varying across clusters). In addition, $\mathbf{D}_k$, $\mathbf{A}_k$, and λ_k determine the ori-

entation, shape, and volume of the kth component of the mixture respectively. The commonly used structures of $\mathbf{\Sigma}_k$ are summarized in Fraley and Raftery (2010).

The E-M algorithm (Dempster et al. 1977, McLachlan and Krishnan, 2008) is a general approach to maximum likelihood estimation for mixture models. Celeux and Govaert (1995) provide detail both on the E and M steps for the case of multivariate normal mixture models parameterized via the eigenvalue decomposition in (3).

In the mixture modeling approach, choosing the clustering method and deciding upon the number of clusters become simply a problem of model selection, since the various combinations of methods and numbers of components represent a set of possible models (Fraley and Raftery, 2002). Model selection in cluster analysis is typically handled through Bayesian model selection approaches (Kass and Raftery, 1995).

A common Bayesian approach is to choose the *a posteriori* most likely model, based on the Bayes factor: the posterior odds for one model against the other, assuming neither is favored *a priori*. Since computing Bayes factors is not easy, the Bayes Information Criterion (BIC) of Schwarz (1978) can be used as an approximation. Since the mixture model fit can only improve as more terms are added to the model, the BIC involves a penalizing term for the complexity of the model, so that it may be maximized for more parsimonious parameterizations and smaller numbers of groups. Model selection for the cluster analyses in this paper will be based primarily on the BIC. A conventional rule of thumb for calibrating BIC differences is that differences of less than 2 correspond to weak evidence, differences between 2 and 6 to positive evidence, differences between 6 and 10 to strong evidence, and differences greater than 10 to very strong evidence (Jeffreys, 1961; Kass and Raftery, 1995; Raftery, 1999).

3 Data

Lijphart (1999) develops a new concept for types of democracies via a majoritarian-consensus contrast, representing two extremes along a continuum. The majoritarian model describes democracies where a bare majority–often a mere plurality–governs, while the consensus model aims at broad participation in government and broad agreement on the policies through sharing of power and division of power. Lijphart's dataset includes ten variables capturing "the most important democratic institutions and rules" deduced from the majoritarian and consensus principles (Lijphart, 1999, p. 2). Then Lijphart groups these ten variables into two clearly separate dimensions of the majoritarian-consensus contrast: the joint-power and divided-power dimensions. The crucial distinction between two dimensions is whether power is "dispersed to political actors operating together *within* the same political institutions" (joint-power dimension) or "dispersed to *separate* political institutions" (divided-power dimension) (Lijphart, 1999, p. 185). Operational definitions of five variables for each dimension are below. Note that each definition states a key feature of majoritarian democracies followed by that of consensus democracies. Higher values for the "Cabinets," "Executive-legislature relations," "Electoral systems," and "Interest groups" stand for majoritarian characteristics while higher values of the other six variables indicate consensual characteristics. In addition, the first five variables belong to the joint-power dimension while the last five variables belong to the divided-power dimension.

1. **Party systems**: Contrasts a two-party system vs. multiparty system, which is measured through the "effective number of legislative parties" of Laakso and Taagepera (1979). The index is defined as $N = 1/\sum s_i^2$, where s_i is the seat share of the ith party.

2. **Cabinets**: Contrasts concentration of executive power in a single-party vs. ex-

ecutive power sharing through multiparty coalitions. The degree of concentration of executive power is measured by the mean of proportions of minimal winning cabinet life and one-party cabinet life (Lijphart, 1999, p. 110–111).

3. **Executive-legislature relations**: Contrasts dominant executive branch vs. balanced executive-legislative relationship. Lijphart (1999) proposes cabinet durability as a proxy, since a long cabinet duration indicates its dominance vis-à-vis the legislature while a short indicates its relative weakness. This variable is measured by each country's average "executive dominance index" based on cabinet durations in years (Lijphart, 1999, p. 132–133).

4. **Electoral systems**: Contrasts majority and plurality methods vs. proportional representation. Lijphart (1999, p. 157–162) measures this using Gallagher's (1991) least squares measure, $\sqrt{0.5(s_i - v_i)^2}$, where s_i and v_i are the seat and vote shares of the ith party respectively.

5. **Interest groups**: Contrasts pluralist interest group systems with free-for-all competition vs. corporatist interest group systems aimed at compromise and concertation. This variable is measured by Siaroff's (1999) complex aggregate index reflecting eight basic aspects of the pluralism-corporatism contrast (Lijphart, 1999, p. 175–180).

6. **Division of power**: Contrasts unitary and centralized government vs. federal and decentralized government. Lijphart (1999, p. 186–191) constructs a fivefold classification and assigns values ranging from 1 to 5, based on two criteria: formal federal constitution and centralization/decentralization.

7. **Parliaments and congresses**: Contrasts concentration of legislative power in unicameral legislature vs. division of legislative power between two equally strong but differently constituted houses. This makes distinctions between bicameralism and unicameralism, between symmetric and asymmetric bicameralism, and between

congruent and incongruent bicameralism, Lijphart (1999, p. 212) constructs a classification of the cameral structures and assigns values ranging from 1 to 4.

8. **Constitutions**: Contrasts flexible constitutions that can be amended by simple majorities vs. rigid constitutions that require extraordinary majorities for amendment. Lijphart (1999) introduces the flexible vs. rigid constitution contrast which is measured via amendment procedure. Various constitutional provisions are reduced to four basic types (Lijphart, 1999, p. 218–223) and values ranging from 1 to 4 are assigned to each type.

9. **Constitutional review**: Contrasts systems in which legislatures judge the constitutionality of their own legislation vs. systems in which supreme or constitutional courts have the power of judicial review. Based on the distinction between the presence or absence of judicial review of the constitutionality of laws and on the degree of judicial activism, a fourfold classification scheme is employed for the variable (Lijphart, 1999, p. 225–228).

10. **Central banks**: Contrasts central banks that are dependent on the executive vs. independent central banks. The central bank independence is measured by averaging three indicators for (1) legal central bank independence, (2) turnover rate of the central bank governor, and (3) political and economic autonomy of the central bank (Lijphart, 1999, p. 235–240).

The data set includes 36 democracies, which are countries with population over a quarter million and rated as "free" for at least nineteen years in the Freedom House surveys (produced since 1972 and based on political rights and civil liberties), as of 1995. All values for the ten variables are averaged over a long period, close to 50 years for the 20 older democracies and a minimum of 19 years for the newest democracies such as India, Papua New Guinea, and Spain. To explore similarities and differences

in majoritarian-consensus characteristics of the 36 democracies between the full period and a recent period, two periods (1945-1996 and 1971-1996) are separately analyzed.

4   Normal Mixture Model for Types of Democracy

The `Mclust` function within the `mclust` package (Fraley and Raftery, 2010) in `R` (R Development Core Team, 2010) is designed to implement the finite mixture model-based cluster analysis described in Section 2. We use `Mclust` to implement the model-based clustering on Lijphart's ten variables. Formal variable selection (using the `clustvarsel` function in R) and informal diagnostic plots were used to determine whether any of the ten variables played no role in the clustering structure. While there was some evidence that certain variables (primarily "Central banks") were less important, the evidence favoring variable reduction was not overwhelming, and so we conducted the cluster analysis with all ten variables. (The substantive conclusions were similar with cluster analyses based on eight or nine of the variables.)

Following Milligan and Cooper's (1988) finding that standardization by dividing each variable by its range gives consistently superior recovery of the underlying cluster structure, all the variables are standardized by dividing by each variable's range. While applying principal components analysis to reduce the dimension of the data before clustering is frequently practiced, Ahlquist and Breunig (2009) argue that "traditional data reduction techniques and cluster analysis do not easily go together" and Chang (1983) shows that it is not justified in general. Our analysis includes all ten variables in the model, and scores on the first two principal components are employed to display the clustering results graphically.

As Fraley and Raftery (2007) point out, maximum likelihood estimation (MLE) for normal mixtures using the EM algorithm may fail as the result of singularities or

degeneracies in the covariance estimate, typically for models in which the covariance is allowed to vary between components. To avoid this, they propose replacing the MLE by a maximum a posteriori (MAP) estimate from a Bayesian analysis. A prior distribution on the parameters eliminates this failure, while having little impact on the stable results obtainable without a prior. The function `priorControl` is provided in the `mclust` package to specify the prior and its hyperparameters. For the multivariate data in this analysis, we employ a normal prior on the mean vector, and an inverse Wishart prior on the covariance matrix, with hyperparameter values set to the default values suggested by Fraley and Raftery (2007). Note that with this prior regularization, a modified BIC (with the likelihood evaluated at the MAP) is used for model selection; see Fraley and Raftery (2007) for details.

4.1 Model-based Clustering Results for the Full Period

With the prior regularization, models with four and five spherical clusters with varying volume and equal shape (labeled VII) are identified as the two best (see Figure 1) with BIC values of -106.7 and -117.3 for the four-cluster and five-cluster models respectively. Following Jeffreys (1961) and Kass and Raftery (1995), differences in BIC greater than 10 correspond to very strong evidence. Thus we conclude that the model with four spherical clusters with varying volume and equal shape is the best one. The solution is projected using the first two principal component scores in Figure 2.

The complete assignment of countries to clusters for the full period is given in the left side of Table 1. Graphically, Figure 2 depicts the four clusters with their member countries. Cluster 3 includes most Scandinavian counties along with Switzerland (which can be regarded as an ideal type of consensus democracy), while Cluster 4
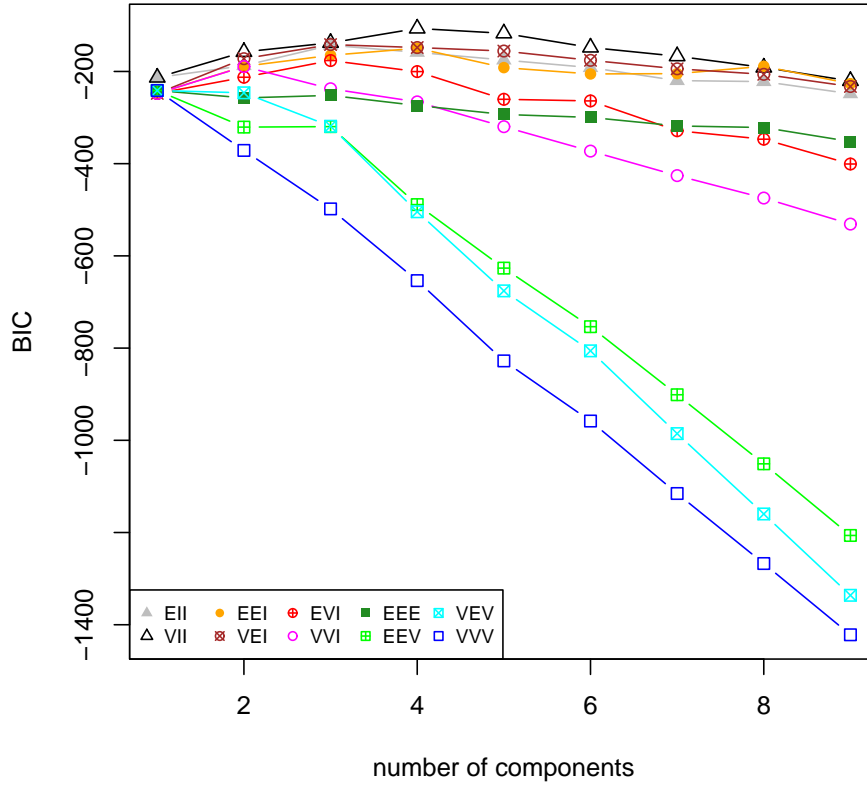
Figure 1: BICs for candidate mixture models for clustering the full-period (1945-1996) data set. Following Fraley and Raftery (2010), the candidate component structures are labeled using: EII=spherical, equal volume; EEI=diagonal, equal volume and shape; EVI=diagonal, equal volume, varying shape; EEE=ellipsoidal, equal volume, shape, and orientation; VEV=ellipsoidal, equal shape; VII=spherical, unequal volume; VEI=diagonal, varying volume, equal shape; VVI=diagonal, varying volume and shape; EEV=ellipsoidal, equal volume and equal shape; VVV=ellipsoidal, varying volume, shape, and orientation.
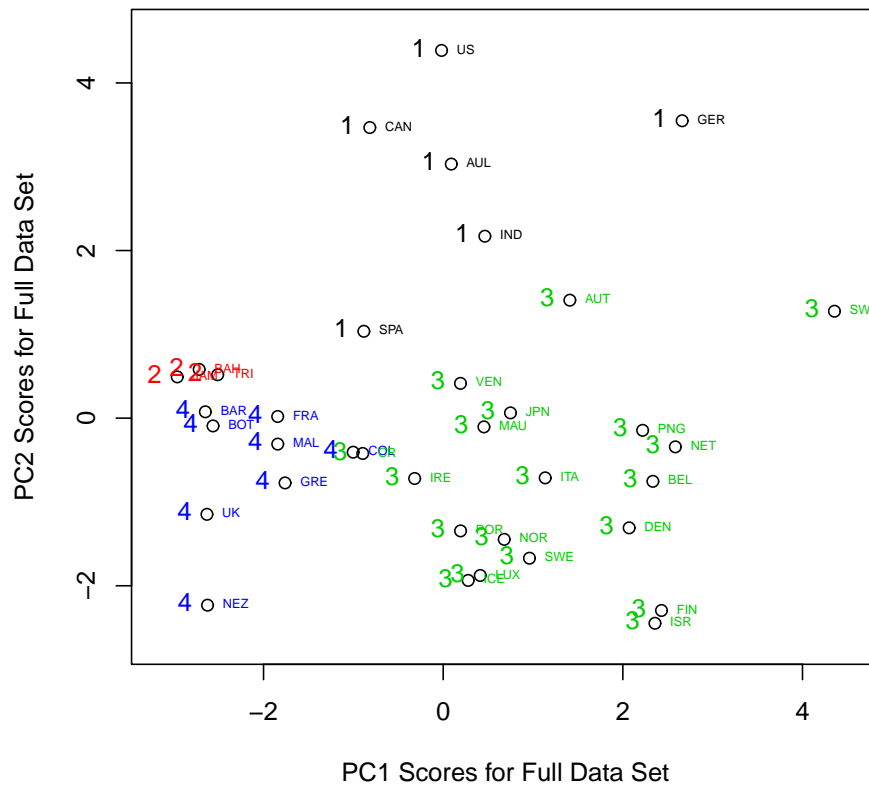
Figure 2: Country membership by cluster (4-cluster solution), 1945-1996 data.

includes UK and New Zealand, which are regarded as prototypes of majoritarian democracy. Cluster 2 (which comprises three Caribbean counties) and Cluster 1 are neither fully majoritarian nor fully consensus. Table 2 allows a clearer sense of characteristics of each cluster through cluster means for the ten variables. Clusters 3 and 4 highlight the majoritarian-consensus contrast very well, especially for the five variables in the joint-power dimension. However, note that for Clusters 3 and 4, all five variable means in the divided-power dimension are lower than the corresponding grand means, and the "Parliaments and congresses," "Constitutional review" and "Central banks" variables have very similar values across Clusters 3 and 4. This implies that the variables in the joint-power dimension are more influential in categorizing countries with respect to the typical majoritarian-consensus contrast between Clusters 3 and 4. Means of the joint-power dimension variables for Cluster 1 lie between those of Clusters 3 and 4, but closer to those of Cluster 4. Uniformly high means of divided-power dimension variables characterize Cluster 1; all of these countries have values close to the highest possible scores for each variable. Countries in Cluster 1 show mixed characteristics: majoritarian for the joint-power dimension and consensual for the divided-power dimension. In Cluster 2, all the variables in the joint-power dimension are "more majoritarian" than those in Cluster 4 (to which prototype countries for majoritarian democracies belong), while variable means in the divided-power dimension are close to those of Cluster 3 except for the "division of power" and "Parliaments and congresses" variables. In general, countries in Cluster 2 are mixed in the "opposite direction" as those in Cluster 1.

In addition, we show star plots summarizing the ten variables' values for all 36 countries in Figure 3. We can see Cluster 1 characterized by relatively large values of the "Division of power," "Parliaments and congresses," "Constitutions," and
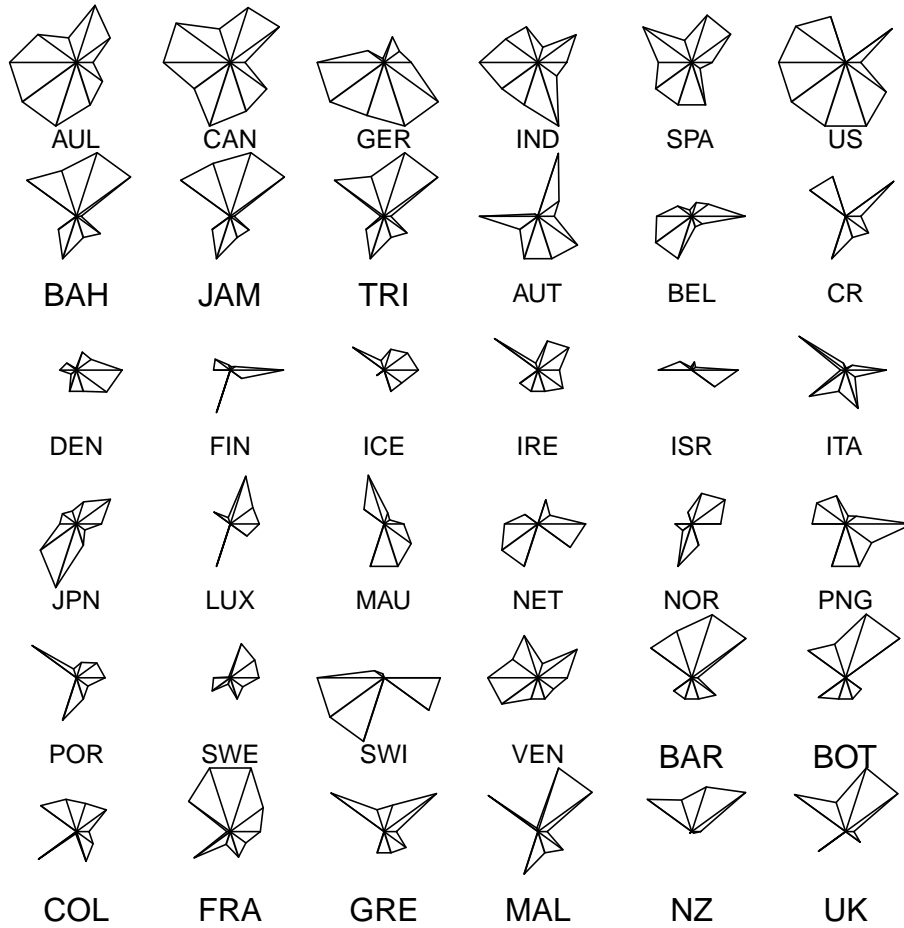
Figure 3: Star plots indicating the variable values for the 36 countries. Countries are listed in order of the first through the fourth clusters, with countries in Clusters 2 and 4 having large-font labels in the star plot. Magnitudes of variables 1 through 10 (see Section 3 for variable names in order) are shown, starting with variable 1 at the 3 o'clock position and proceeding counterclockwise around the star.

"Constitutional review" variables (on the lower portion of the stars). Countries in Cluster 2 (Bahamas, Jamaica, Trinidad and Tobago) have nearly identical stars (also confirmed by the near-zero standard deviations for Cluster 2 in Table 2), with large values for the "Cabinets," "Executive-legislature relations," "Electoral systems," and "Interest groups" variables. Cluster 3 is the most populous and wide-ranging cluster, but countries appear fairly similar, with relatively high values of the "Party systems" variable along with modest values for other variables. Switzerland is potentially an outlying member of this cluster in some ways. The fourth cluster is characterized by lower values of the "Party systems" variable and larger values of the "Cabinets," "Executive-legislature relations" and "Interest groups" variables.

Table 1: Country cluster membership for two periods: 1945-1996 data vs. 1971-1996 data.

Cluster	Full Period, 1945-1996		Recent Period, 1971-1996	
	Member Countries	#	Member Countries	#
1	Australia, Canada, Germany, India, Spain, United States	6	Australia, Austria, Canada, Germany, India, Spain, United States, Venezuela	8
2	Bahamas, Jamaica, Trinidad and Tobago	3	Bahamas, Jamaica, Trinidad and Tobago	3
3	Austria, Belgium, Costa Rica, Denmark, Finland, Iceland, Ireland, Israel, Italy, Japan, Luxembourg, Mauritius, Netherlands, Norway, Papua New Guinea, Portugal, Sweden, Switzerland, Venezuela	19	Belgium, Costa Rica, Denmark, Finland, Iceland, Ireland, Israel, Italy, Japan, Luxembourg, Mauritius, Netherlands, Norway, Papua New Guinea, Portugal, Sweden	16
4	Barbados, Botswana, Colombia, France, Greece, Malta, New Zealand, United Kingdom	8	Barbados, Botswana, Colombia, France, Greece, Malta, New Zealand, United Kingdom	8
5			Switzerland	1

Overall, Figure 2 indicates that the first principal component addresses the ma-

Table 2: Cluster means and standard deviations for ten variables, 1945-1996 data.

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Grand Mean
Party systems	2.688	1.707	3.851	2.127	3.095
	(0.536)	(0.103)	(1.030)	(0.597)	(1.173)
Cabinets	69.570	99.700	38.830	89.090	60.190
	(21.277)	(0.520)	(22.442)	(18.185)	(31.636)
Executive-legislature relations	3.370	5.520	2.366	4.701	3.315
	(1.659)	(0.000)	(1.169)	(1.182)	(1.650)
Electoral systems	9.865	15.630	5.312	11.474	8.300
	(4.487)	(2.050)	(4.691)	(5.433)	(5.693)
Interest groups	2.743	3.300	1.604	2.990	2.243
	(0.813)	(0.000)	(0.838)	(0.370)	(0.986)
Division of power	4.583	1.067	2.258	1.025	2.272
	(0.801)	(0.116)	(1.262)	(0.071)	(1.520)
Parliaments and congresses	3.500	2.000	1.905	2.025	2.206
	(0.548)	(0.000)	(0.970)	(0.888)	(1.016)
Constitutions	3.583	3.000	2.595	1.712	2.597
	(0.492)	(0.000)	(0.875)	(0.688)	(0.939)
Constitutional review	3.550	2.000	1.911	1.825	2.172
	(0.505)	(0.000)	(0.697)	(0.529)	(0.857)
Central banks	0.465	0.380	0.364	0.333	0.375
	(0.158)	(0.027)	(0.110)	(0.070)	(0.112)

Note: Standard deviations are in parentheses

joritarian/consensus contrast well, especially with respect to joint-power-dimension variables. More majoritarian countries like those in Cluster 4 are found on the left side of Figure 2, while more consensus countries like those in Cluster 3 have larger scores on the first principal component. The second principal component relates more to the countries' values on the divided-power variables, which separate the countries in Cluster 1 from the rest.

A major advantage of the model-based clustering technique is that it produces an estimate of the conditional probability that observation i belongs to group k given the current parameter estimates. For this clustering solution, 33 of 36 countries have conditional probabilities greater than 0.995 of belonging to each of their respective

groups as identified by the four-cluster model, while Colombia, Ireland, and Spain have probabilities of 0.846, 0.708, and 0.926 of belonging to each of their respective clusters.
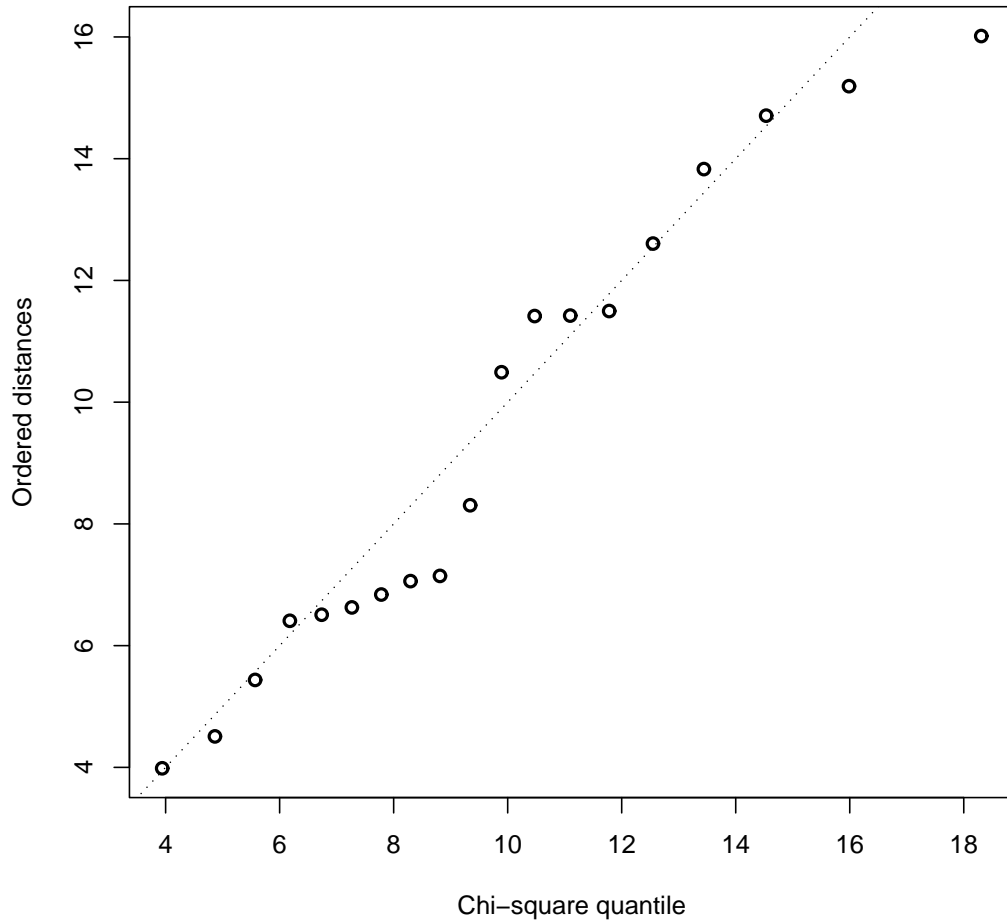

Figure 4: Chi-square plot for countries from Cluster 3. A relatively straight plot indicates the multivariate normality assumption is reasonable.

We note that the mixture-model setup we employ assumes each component follows a multivariate normal distribution. Admittedly, this normality assumption makes more sense for some of the variables, such as those representing complex indices, than for others such as variables 6-9, which are ordinal classifications. Conceptually,
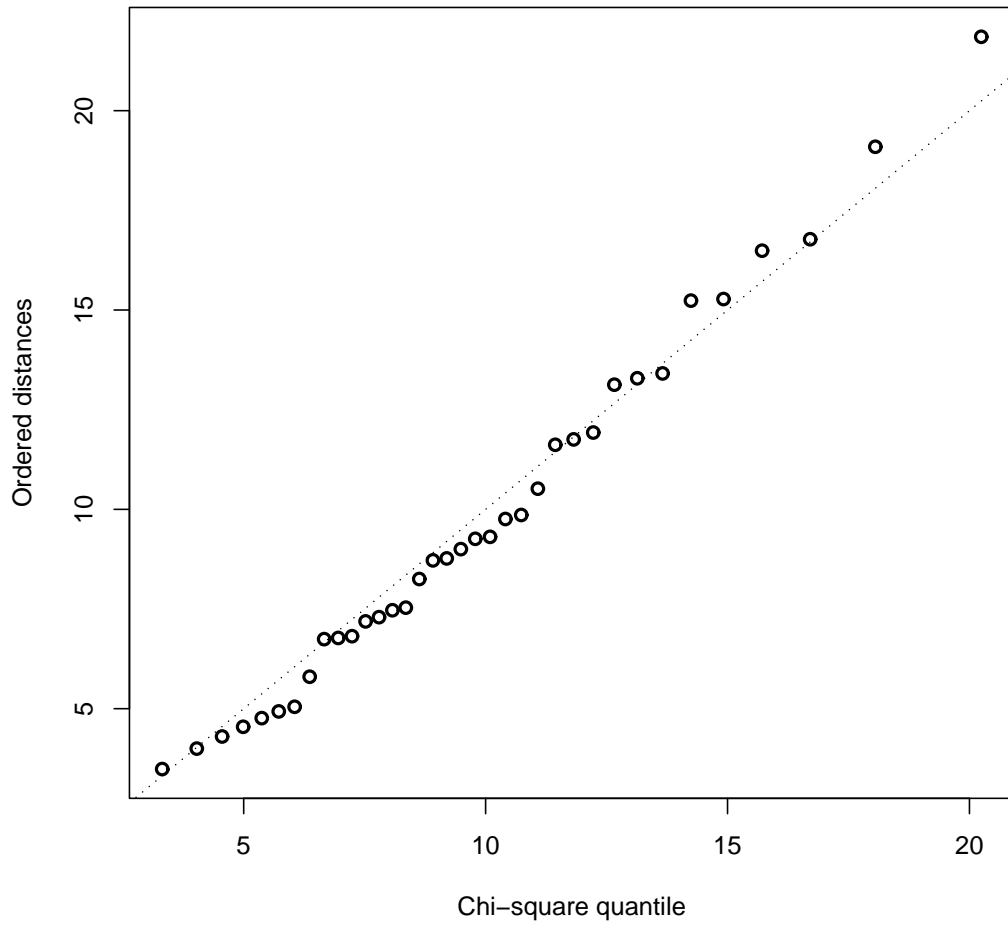
Figure 5: Chi-square plot for countries from all four clusters together. A relatively straight plot indicates the multivariate normality assumption is reasonable.

all ten variables are averaged over a substantial series of years and thus may have approximately normal distributions (whether this is practically true is another matter, since the yearly ratings are not independent). To judge how well the multivariate normal distribution fits the data, we used chi-square plots (Everitt, 2005) on Cluster 3 and on the entire data set. (Clusters 1, 2, and 4 have too few objects – or have objects with too little variability on certain variables – for chi-square plots to be appropriate for these individual clusters.) The two chi-square plots in Figures 4 and 5 show relatively straight plots, indicating a fidelity to the assumption of multivariate normality.

4.2   Comparison with Other Clustering Methods

As discussed in Section 2, in addition to model-based clustering, other popular clustering algorithms include hierarchical methods and partitioning methods; see Everitt (1993) for a detailed review. For comparative purposes, we now use average linkage (Sokal and Michener, 1958)–a popular hierarchical method–and K-means (MacQueen, 1967)–a popular partitioning method–to cluster the countries.

The average linkage approach produces the dendrogram shown in Figure 6. Again four clusters are apparent, but the most notable difference between this solution and the model-based solution is that the average linkage solution immediately separates Switzerland into its own cluster. While this choice is potentially reasonable and may make sense for recent data patterns (see Section 6 for more discussion), it also points to a weakness of hierarchical methods: Note that once Switzerland is isolated in the first step of the algorithm, it cannot be incorporated in a cluster with other countries in a later step.

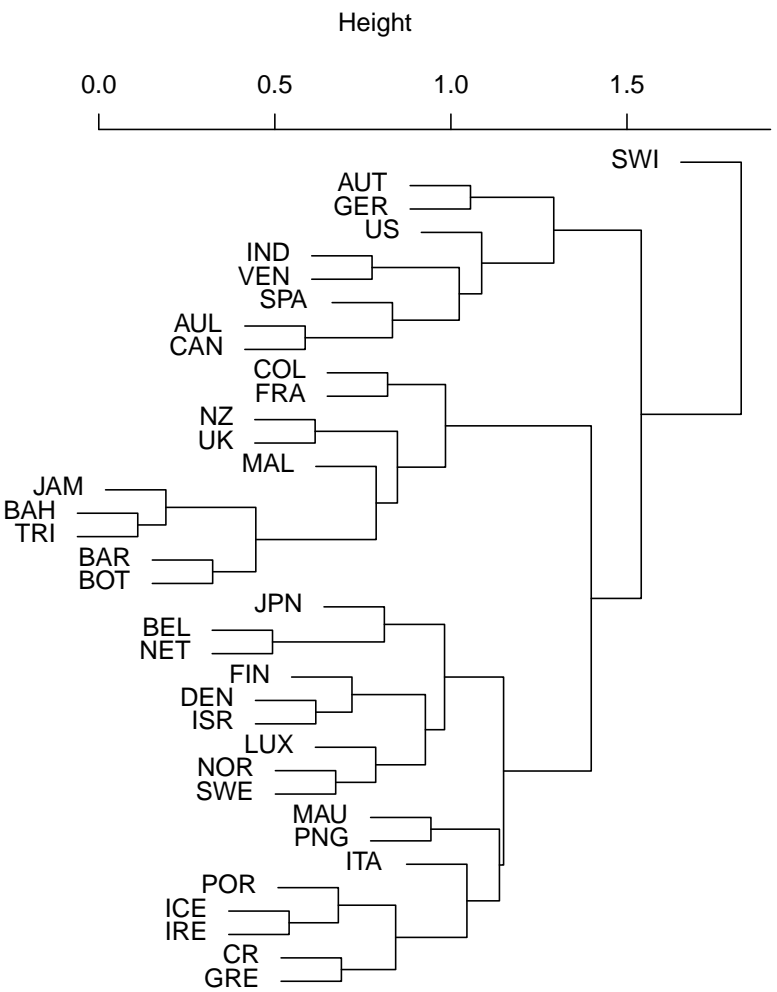In addition, a brief summary of the four-cluster K-means solution is:

Figure 6: Dendrogram showing the average linkage hierarchical solution to the full (1945-1996) data set.

22

```
[1]: AUL AUT CAN GER IND SPA US VEN.

[2]: CR DEN FIN ICE IRE ISR ITA LUX MAU NOR PNG POR SWE.

[3]: BEL JPN NET SWI.

[4]: BAH BAR BOT COL FRA GRE JAM MAL NZ TRI UK.
```

Comparing this to the model-based solution shown in the left side of Table 1, we see that the K-means algorithm places Caribbean countries (Bahamas, Jamaica, Trinidad and Tobago) with the majoritarian democracies of the model-based solution's fourth cluster. We note that this is problematic in that these countries strongly differ (especially with regard to the joint-power-dimension variables) from the countries in Cluster 4 in the model-based solution. In addition, the large third cluster of the model-based solution is split into two clusters, with a smaller subset (Belgium, Japan, Netherlands, Switzerland) broken off. Overall, we prefer the model-based approach because of its reliance on formal likelihood theory, its inherent ability to produce cluster membership probabilities, and its seamless, unified approach toward choosing the correct number of clusters and proper partitioning.

5 Sensitivity Analysis

Although the four-cluster model is chosen as the best one based on the BIC and the majority of the observations are well classified, we now consider the uncertainty stemming from measurement error in the variables in the data set. While we believe the data values given by Lijphart (1999) to be generally trustworthy, it may be realistic to view the values for each country to be "close to the truth" rather than the certain, exact truth. Therefore, sensitivity analysis is performed to evaluate the stability of the chosen clustering solution with respect to slight perturbations to the observed

data.

Hubert and Arabie (1985) created the adjusted Rand index (ARI hereafter) to measure the correspondence between two clustering solutions, based on how pairs of objects are classified, by

$$ARI_{HA} = \frac{\binom{N}{2}(a+d) - [(a+b)(a+c) + (c+d)(b+d)]}{\binom{N}{2}^2 - [(a+b)(a+c) + (c+d)(b+d)]},$$

where $\binom{N}{2}$ represents the total count of item pairs resulting in four different categories of pairs. Here, a is a count of pairs of items placed in the same group in both clustering solutions, while d counts pairs of items placed in different groups in both clustering solutions. Also, b and c count pairs of items placed in the same group in one clustering solution and in different groups in the other. This index is a widely accepted measure of the concordance between two proposed clustering partitions. Steinley (2004) proposes a set of heuristics for determining, in terms of ARI, the degree of concordance between two proposed clustering partitions: Values greater than 0.90 and 0.80, respectively, can be considered excellent and good concordance, while 0.65 can be viewed as the threshold between moderate and poor concordance. The ARI is employed to compare the fixed clustering solution in the left side of Table 1 with clustering solutions on perturbed data constructed by adding normal noise to the original data set. To create each perturbed data set, normal random noise, with mean zero and standard deviation equaling k times the per-cluster standard deviation of each respective variable, was added to each data value. Values of k were chosen to be $k = 0.01$, 0.05, 0.1, 0.2, 0.3, and 0.5, yielding very small to moderate perturbations of the data. Table 3 lists the resulting ARIs averaged over 1000 perturbed data sets. To measure for the stability of each individual cluster, recovery rates (defined as the rate of items having the same cluster label on the perturbed data as the initial cluster

Table 3: Sensitivity of clustering solution to normal perturbations.

k	Average Adjusted Rand Index	Recovery Rate			
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
0.01	0.998	1.000	1.000	0.999	0.998
0.05	0.938	0.993	1.000	0.947	0.931
0.10	0.831	0.978	0.998	0.694	0.710
0.20	0.704	0.959	0.941	0.405	0.473
0.30	0.661	0.950	0.872	0.316	0.429
0.50	0.603	0.918	0.732	0.247	0.419

label) for each cluster for each chosen value of k are also obtained.

As expected, the larger the added noise, the smaller the average ARI. Based on Steinley's (2004) criterion, the average adjusted Rand Indices for $k = 0.01$, 0.05, and 0.1 show very high stability in the sensitivity analysis. The smaller average ARI values for $k = 0.30$ and 0.50 mainly stem from the low stabilities in Clusters 3 and 4. Clusters 1 and 2 remain relatively stable when the added normal noise gets larger, while Clusters 3 and 4 become unstable more quickly because of their relatively larger within-cluster standard deviations depicted in Table 2.

6 Types of Democracy for the Recent Period

As discussed in the previous section, all values for the ten variables analyzed above are averaged over a long period (nearly fifty years for the twenty older democracies and at least nineteen years for the three newest democracies). To explore whether recent changes in political systems have led to changes in the cluster structure for the majoritarian-consensus characteristics of the thirty-six democracies, we perform model-based clustering on a recent subset of the full-period data. Here, we use a subset provided by Lijphart (1999) that involves the period from 1971-1996. Lijphart (1999) does not provide analogous data for the 1945-1970 period, but some informal

conclusions about the clustering structure for 1945-1970 could be gleaned by contrasting our full-period and recent-period results. We also note that analogous data for years more recent than 1996 are not yet available and cannot be reconstructed in a way consistent with Lijphart's data: The first four variables are periodically updated or easily calculated based on clearcut formulas for each variable, but the other variables are coded by Lijphart based on other experts' estimates, which are not regularly updated.

For the recent-period (1971-1996) data, a five-cluster model with spherical clusters having varying volume is identified as the best model based on its superior BIC (-116.3 compared to the next-best BIC value of -127.4). The clustering solution is plotted on the first two principal component scores in Figure 7. In addition, Table 1 compares the best models for the full period (left side) and for the recent period (right side). The two solutions differ in membership for three countries: Austria, Switzerland, and Venezuela. The model for the recent period isolates Switzerland as a separate cluster, and switches Austria and Venezuela from Cluster 3 to Cluster 1. This makes sense based on the data: For the recent period, consensual characteristics of Switzerland have become stronger; the effective number of parliamentary parties and central bank independence index increased from 5.24 to 5.57 and from 0.60 to 0.63, respectively. In contrast, Austria and Venezuela show changes in the opposite direction. The proportion of minimal winning coalition and one-party cabinets, and executive dominance index for Austria have increased from 41.4% to 65.1% and 5.47 to 5.52 respectively. For Venezuela, the effective number of parliamentary parties decreased from 3.38 to 3.07 and the proportion of minimal winning coalition and one-party cabinets increased from 73.4% to 82.4%. Notice that other countries do have changes in their political systems, but the changes are not enough to shift from

26

one cluster to another, and for both the recent and the longer periods the consensus democracies are more numerous than majoritarian democracies.
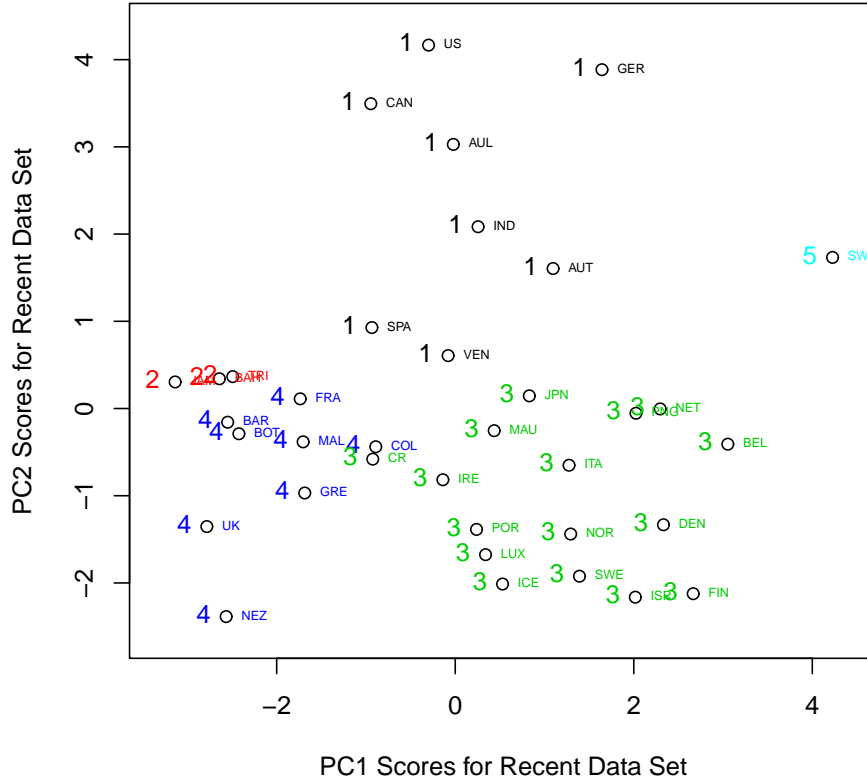

Figure 7: Country membership by cluster (G=5), 1971-1996 data.

7   Discussion

Lijphart (1999) systematically distinguishes majoritarian and consensus democracies according to the ways in which they have institutionalized the decision-making process in two dimensions: joint-power and divided-power dimensions. His framework is broadly applicable to democratic political systems in analyzing the relationships between different parts of an institutional arrangement as well as investigating the

effect of types of democracies on their performance.

Nonetheless, Lijphart's treatment of the typology of democracy has some limitations for general use. While he discusses types of democracy in two-dimensional space, it may be preferable to develop majoritarian/consensus scores using all 10 variables simultaneously, rather than splitting the variables into two sets of five. In his separate analyses of the effect of types of democracy on macro-economic management, the control of violence, and the quality of democracy, he uses bivariate regressions with separate scores for types of democracy in each dimension as covariates. Admittedly, unidimensional majoritarian-consensus contrast scores for each dimension are very informative in analyzing the effect of types of democracy on economic performance as well as quality of democracy. There exist limitations in their interpretability, however. When interpreting models with the majoritarian-consensus score as one of predictor variables, it is hard to interpret the parameter estimate in terms of the original variables' units. Furthermore, in the process of combining five variables to gain the majoritarian-consensus score for each dimension, characteristics of each variable disappear, masking each variable's contribution to democracy type.

In addition, Lijphart's typology is not classificatory but continuous, and so we can only say some countries are more or less majoritarian or consensual than other countries, rather than being able to classify countries to a specific type. Therefore his typology is not applicable to qualitative research in that it does not recognize specific country types lying somewhere between two extremes. These limitations motivated employing model-based cluster analysis as an alternative for types of democracy in particular and for typology in general.

In sum, we apply the appealing features of normal mixture model-based clustering to the substantive problem of grouping types of democracy, and we find a reason-

able cluster structure for the majoritarian-consensus characteristics of the thirty-six democracies. We are able to classify countries to specific types which will be conducive to examining relationships among types of democracy with other variables for future research. In the future, as the third wave of democratization (Huntington, 1991) finds that approximately thirty democratic regimes replaced authoritarian regimes in Europe, Asia, and Latin America, the scope of the research will likely be broadened once data for these newer democracies (along with more recent data for the 36 democracies studied here) become available.

References

Ahlquist, J. S. and Breunig, C. (2009). Country clustering in comparative political economy. *MPIfG Discussion Paper 09/5, Max-Planck-Institut für Gesellschaftsforschung.*

Alvarez, M., Cheibub, J. A., Limongi, F. and Przeworski, A. (1996). Classifying political regimes. *Studies in Comparative International Development* **31**, 3-36.

Andeweg, R. B. and Thomassen, J. (2005). Modes of political representation: Toward a new typology. *Legislative Studies Quarterly* **30**, 507-528.

Banfield, J. D. and Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics* **49**, 803-821.

Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika* **65**, 31-38.

Bock, H. H. (1996). Probabilistic models in cluster analysis. *Computational Statistics and Data Analysis* **23**, 5-28.

Celeux, G. and Govaert, G. (1995). Gaussian parsimonious clustering models. *Pattern Recognition* **28**, 781-793.

Chae, H. (2010). South Korean attitudes toward the ROK-U.S. alliance: Group analysis. *PS: Political Science and Politics* **43**, 493-501.

Chang, W. C. (1983). On using principal components before separating a mixture of two multivariate normal distributions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **32**, 267-275.

Cheeseman, P. and Stutz, J. (1995). Bayesian classification (autoclass): Theory and results. In *Advances in Knowledge Discovery and Data Mining* (Edited by U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy). AAAI Press, Menlo Park, CA.

Collier, D. and Adcock, R. (1999). Democracy and dichotomies: A pragmatic approach to choices about concepts. *Annual Review of Political Science* **2**, 537-565.

Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika* **56** 463-474.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**, 1-38.

Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*, Wiley, New York.

Edwards, A. W. F. and Cavalli-Sforza, L. L. (1965). A method for cluster analysis. *Biometrics* **21**, 362-375.

Elman, C. (2005). Explanatory typologies in qualitative studies of international politics. *International Organization* **59**, 293-326.

Everitt, B. (1993). *Cluster Analysis*, John Wiley & Sons, Inc, New York.

Everitt, B. (2005). *An R and S-PLUS Companion to Multivariate Analysis*, Springer-Verlag, London.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97**, 611-631.

Fraley, C. and Raftery, A. E. (2007). Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classification* **24**, 155-181.

Fraley, C. and Raftery, A. E. (2010). MCLUST version 3 for R: Normal mixture modeling and model-based clustering. *Technical Report No. 504., Department of Statistics, University of Washington.*

Franzoni, J. M. (2008). Welfare regimes in Latin America: Capturing constellations of markets, families, and policies. *Latin American Politics and Society* **50**, 67-100.

Gallagher, M. (1991). Proportionality, disproportionality and electoral systems. *Electoral Studies* **10**, 33-51.

George, A. L. and Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*, MIT Press, Cambridge, MA.

Ghosh, J. (2003). Scalable clustering. In *Handbook of Data Mining* (Edited by N. Ye), 341-364. Lawrence Erlbaum Associates, Inc..

Ghunther, R. and Diamond, L. (2003). Species of political parties: A new typology. *Party Politics* **2**, 167-199.

Gough, I. and Sharkh, M. A. (2010). Global welfare regimes: A cluster analysis. *Global Social Policy* **10**, 27-58.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification* **2**, 193-218.

30

Huntington, S. P. (1991). *The Third Wave : Democratization in the Late Twentieth Century*, University of Oklahoma Press, Norman, OK.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data*, Prentice Hall, Englewood Cliffs, NJ.

Jain, A. K., Murty, M. N. and Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys* **31**, 264-323.

Jeffreys, H. (1961). *Theory of Probability*, Oxford University Press, New York, NY.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.

Kass, R. E. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* **90**, 928-934.

Laakso, M. and Taagepera, R. (1979). "Effective" number of parties: A measure with application to West Europe. *Comparative Political Studies* **12**, 3-27.

Lazarsfeld, P. F. and Barton, A. H. (1951). Qualitative measurement in the social sciences: Classification, typologies, and indices. In *The Policy Sciences* (Edited by D. Lerner and H. D. Lasswell), 155-192. Stanford University Press, Stanford, CA.

Lee, Y. J. and Ku, Y. W. (2007). East Asian welfare regimes: Testing the hypothesis of the developmental welfare state. *Social Policy and Administration* **41**, 197-212.

Lehnert, M. (2007). Typologies in social inquiry. In *Research Design in Political Science: How to Practice what They Preach* (Edited by T. Gschwend and F. Schimmelfennig), 62-79. Palgrave Macmillan, New York, NY.

Lijphart, A. (1999). *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*, Yale University Press, New Haven, CT.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 281-297. University of California Press, Berkeley.

McLachlan, G. J. and Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*, Marcel Dekker, Inc, New York, NY.

McLachlan, G. J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*, Wiley-Interscience, Hoboken, NJ.

Milligan, G. W. and Cooper, M. C. (1988). A study of standardization of variables in cluster analysis. *Journal of Classification* **5**, 181-204.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology* **25**, 111-163.

Raftery, A. E. (1999). Bayes factors and BIC: Comment on "A critique of the Bayesian information criterion for model selection." *Sociological Methods and Research* **27**, 411-427.

Raftery, A. E. and Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association* **101**, 168-178.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (http://www.r-project.org/).

Rudra, N. (2007). Welfare states in developing countries: Unique or universal? *The Journal of Politics* **69**, 378-396.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461-464.

Scott, A. J. and Symons, M. J. (1971). Clustering methods based on likelihood ratio criteria. *Biometrics* **27**, 387-397.

Siaroff, A. (1999). Corporatism in 24 industrial democracies: Meaning and measurement. *European Journal of Political Research* **36**, 175-205.

Sokal, R. R. and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin* **38**, 1409-1438.

Spirling, A. and Quinn, K. (2010). Identifying intraparty voting blocs in the U.K. House of Commons. *Journal of the American Statistical Association* **105**, 447-457.

Steinley, D. (2004). Properties of the Hubert-Arabie adjusted Rand index. *Psychological Methods* **9**, 386-396.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **58**, 236-244.

Webb, P. (2008). The attitudinal assimilation of Europe by the conservative parliamentary party. *British Politics* **3**, 427-444.

Wolfson, M., Madjd-Sadjadi, Z. and James, P. (2004). Identifying national types: A cluster analysis of politics, economics, and conflict. *Journal of Peace Research* **41**, 607-623.

Zhong, S. and Ghosh, J. (2003). A unified framework for model-based clustering. *Journal of Machine Learning Research* **4**, 1001-1037.