5-2018

Can Probabilistic Models Reveal Characteristics of Voting Population Segments?

Alexander Lucien Wang

Can Probabilistic Models Reveal Characteristics of Voting Population Segments?

CAN PROBABILISTIC MODELS REVEAL CHARACTERISTICS OF VOTING POPULATION SEGMENTS?

By

Alexander Lucien Wang

An Undergraduate Thesis submitted in partial fulfillment of the requirements for the

JOSEPH WHARTON SCHOLARS

Faculty Advisor:

Peter Fader

Frances and Pei-Yuan Professor, Marketing Department

THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA

MAY 2018

WH 399: Thesis

Can probabilistic models reveal characteristics of voting population segments?

Alexander Lucien Wang

Introduction

In a modern, hyper-partisan America, predicting — or at least explaining — voting patterns has become an imperative for parties, lobbyists, and the general populace alike. In the aftermath of the 2016 United States Presidential Election, never had there been so many attempts to explain or justify the voting choices that the public made. From news articles to blog posts, the media was saturated in the aftermath of the election with countless pieces struggling to "demystify" the variables and factors that ultimately led to the choices of voters at the booths. These sentiments have likewise festered across the Atlantic, where a series of extreme right-wing parties have surged in popularity across Europe. From the United Kingdom Independence Party (UKIP) to Germany's Alternative für Deutschland (AfD, "Alternative for Germany"), to France's Front National (FN, "National Front"), all have had a significant impact on legislative and federal elections. Populism has gripped Western liberal democracies, where traditional center-leaning parties have been forced out of power; reactionary parties have taken their place in government, from Hungary's Fidesz – Magyar Polgári Szövetség ("Fidez-Hungarian Civic Alliance"), to Poland's Prawo i Sprawiedliwość (PiS, "Law and Justice"), to the recent electoral victories by the Czech party ANO 2011.

This motivated the exploration of this thesis and the specific thesis question, particularly concerning the voting patterns of multi-party democracies. Using tools such as probabilistic models first developed for analysis of customer purchasing (i.e. a choice process that is analogous to voting; buying a particular good that is offered by multiple brands is similar to voting in a mutually exclusive choice for a party amongst many), this thesis hopes to determine whether polarization and segmentation of voting choices can be characterized as parameters within the model. Specifically, this thesis hopes to explore whether the Dirichlet-Multinomial ("D-M") probabilistic model can be used to reveal the innate characteristics of polarization, determined as variables across the model. Furthermore, it hopes to explore whether covariates such as unemployment, education level etc. are even needed to explain the the segmentation of voting populations within these multiparty democracies. The focus is on multiparty rather than two-party democracies as it is easier to identify the specific subdivisions of ideology if there are more formalized and institutionalized parties.

Ultimately this thesis looked at the 2017 and 2013 German federal elections, and analyzed the constituency-level voting data. Only the secondary vote was taken into account for the model parameter prediction.

Literature Review

In order to better understand human choices, and with the knowledge of this mutually exclusive choice structure, it is easy to turn to the most obvious and common method of attempting to link seemingly correlated results — that of the linear regression.

A multiple or multivariable linear regression defines a relationship that is calculated between various independent explanatory variables and a single dependent variable. However, it is clearly not the optimal model in predicting categorical response variables, such as a vote for different parties, as its dependent variable is continuous. Furthermore, a linear regression has the potential to produce nonsensical probabilities, such as those above one, or negative in value. The functional form also assumes an initial marginal increment in an explanatory variable has same effect on decision as an increment on different range (i.e. an underlying assumption of homoscedasticity). This might not be true in the case of voting, where a wealth effect may influence voting between different social strata in different magnitudes.

Multiple papers thus turn toward the multinomial logit or probit model in an attempt to model the election process, and there have been recent discussions on this. These models use a vector of explanatory variables, in order to determine a categorical outcome, where $Y_i^m = X_i\beta_m + \epsilon_m$; and the m decision is associated to max (Y_i). A logit model takes a standard logistic distribution of the errors, while the probit assumes a normal distribution of errors.

Dow and Endersby (2004) compare the multinomial logit and probit models in attempting to determine which functional form is better at predicting voter responses in multi-party elections. The paper immediately turns away from binary models which implies a vote is either to the governing or opposing party, which is untrue in most cases of modern democracies. The paper argues that the simpler logit model is more ideal than the probit models, but ultimately concedes that there is "considerable uncertainty about the use of qualitative choice models in the study of voter choice and related applications."

Alvarez and Nagler (1998) considered the context of regression models by comparing the conditional logit and multinomial logit, arguing that the conditional logit (a variation of the logistic regression where instead of having the individual characteristics driving a particular choice, there is an evaluation of the characteristics of all the different alternatives proposed) is more representative of the voting process. In other words, a vote for a party is more akin to looking at the characteristics of each party first (the consideration of the different alternatives), then selecting the one most akin to the individual's belief (where all options are presented and a preferred one indicated), rather than a choice that is based on your personal characteristics (such as employment, demographics etc.). The papers also concede on the limits of the logit model and its imposition of the property of Independence of Irrelevant Alternatives on individual voters, whereby an entry of a third party would not affect the voting propensities of an individual amongst the initial two. In a simpler analogy, in a democracy with a center-left and center-right party, the entrance of a left-leaning party would not affect the voting propensities for the initial two parties, although this is certainly not the case. As parties enter and leave the political landscape over several elections, logit models have the inherent weakness of leading to incorrect estimates based on this assumption. Alvarez and Nagler (1998) nevertheless admit that the probit model is also not ideal in the lifting of the problem of IIA.

However, fundamentally, the papers rely on the notion that the data and the assumed variables should be driving the model. The very pillars from which the model is being created from are not questioned, where the necessity and importance of the explanatory variables are not probed. These models may also not be parsimonious, as the more variables and characteristics you introduce in the model before taking a maximum likelihood can "over-analyze" the data.

The purpose of this paper is to investigate into alternative modelling strategies in order to segment, investigate and potentially predict voter choices in **a parsimonious manner with as few explanatory variables as possible**. As Professor Fader consistently remarks in lectures in his seminal class STAT 476, "data is disgusting." Or rather, regression models that heavily rely on data in order to make any predictions are inherently "dirty." Regression models are extremely dependent on the underlying data in order to forecast the presence and value of distinctly identified and assumed variables. As such, their future estimations are based on parameters that are derived from predetermined assumptions of existence of particular variables. Aside from the idea of parsimony, there is no leeway within the model for idiosyncrasies of the decision-making processes; as in reality, even simple purchase decisions cannot be predicted on an individual-to-individual level basis. There must be some adjustment from the randomness of people's individual behavior — not everything can be predicted by the presence of a predetermined variable — which is particularly true if the causal integral variable is frequently unobservable and instead inferred from a proxy variable.

For the messy choice processes of purchase and voting decisions, it might be more prudent to turn toward probabilistic models. Probability models incorporate random variables and probability distributions into their models on both the individual and population level, representing various potential outcomes for uncertain events across a population. From this, probability models can be used to predict patterns of behavior and parameters can be extrapolated to provide information about the characteristics of the model-at-large, such as the polarization of the segments. The choices and patterns of behavior are predicted in the aggregate. Covariates can also be introduced to test and experiment the potentiality for causal variables that affect the robustness of the model. There is no requirement to include the covariates, and the data takes a backseat in the predictive capability of the model. The covariates exist to merely complement the "story" offered by the data. The beauty of probability models is also that the lack of data in and of itself is not a problem, unlike multinomial logit or probit models (Dow and Endersby remark that even with "a sample of 1500 observations on voter choice," it is not enough to inform a robust-enough model).

Probability models already have a wide range of marketing and customer choice applications, ranging from evaluating advertising effectiveness to assessing brand strength. More broadly, it can summarize patterns of market-level behavior, where predictions of future periods can be made from this aggregate past data. Summary measures can be derived, in order to make inferences about behavior and to profile the behavioral propensities of certain segments. Applying this market framework to the voting process, an analogous "story" can be told: A "household" (constituency) "purchases" (votes) in a mutually exclusive choice process across several "brands" (parties).

Preface to Probabilistic Models

As such, this analysis is a departure from regular regression or logit models that aim to predict voting patterns, as it removes the assumption of homogeneity of preferences amongst the public. A traditional multinomial regression model attempts to describe the relationship between one dependent normal variable against several independent variables; in other words, there is a common dependent variable such as voting choice that can be "predicted" through a set of independent variables such as race, highest degree attained etc. This thesis argues that multinomial logistic regression models used in previous papers (Dow and Endersby 2004; Alvarez and Nagler 1998) do not fully capture the nuance of a population, especially one as large and necessarily diverse in opinion as a voting public, while at the same time overcompensating with the messy intrusion of an "over-analysis" of data. A D-M model attempts to resolve those failings.

The goal and results of this thesis can be useful in the general realm of election modelling and prediction. Hopefully, the model's parameters can be used to answer questions on the heterogeneity of the voting population, and whether covariates can explain the polarization within the public. For example, a covariate that significantly reduces the polarization index of the segments could imply that some of the heterogeneity and disparate choices that the public made could be explained by that very universal factor. An intuitive example might be unemployment, where disgruntled citizens could vote for more extreme left- or right-wing parties. As such, the target audience can range from the professional pollster attempting to forecast an election result, to the curious general public wanting to seek clarity in a divisive election process. Nevertheless, these hypotheses remain very nebulous and are conjectures based upon an ideal result. The exploration in an alternative probabilistic modelling of voting may reveal that the choice process of market decisions is simply not analogous to that of voting. Factors such as the nature of the process may lend itself to a differing choice strategy that would not be well-captured by the model.

Methodology

The model and methodology itself can be best explained from a market-based decision process. Common choice models and customer lifetime value models operate on the assumption that the customer is homogenous in preference, that one individual is not clearly differentiable from another, and that there is no evolution of choice patterns over time. A probabilistic model as espoused by this paper differs from this by accounting into the fact of the distinct heterogeneity amongst the individuals themselves, as well as the individuals across a population.

The key idea behind this probabilistic model is that any decision can be modeled through an individual-level and population-level distribution. When making a choice or action, there is a certain intrinsic motivator or behavior pattern that influences the individual to do so — in other words there is an underlying individually based propensity of doing something. In a market setting, this may be a consumer choosing to purchase a product out of a range of brands, or visiting a certain website multiple times. These individual level behaviors are choice or count processes at the very fundamental level, where there is distinct heterogeneity between individuals, who make varying choices. A probability distribution is selected on the best description of the potential outcomes. For example, a binary outcome, such as flipping a coin, would potentially utilize a Binomial distribution: there is a p underlying propensity of purchase given an x number of purchase opportunities. An outcome with multiple choices (e.g. a process like rolling a die or voting in an election) could utilize a multinomial distribution, a generalization of the Binomial distribution with k-possibilities. Either way, a distribution is needed to describe these probabilities with each of their outcomes.

On a higher level though, there exists another distribution that best models how these distributions of individual level probabilities are likely to vary at the population level. In other words, this best describes the variance of the individual-level distributions, and are used to "mix" them at the population level: i.e. a "mixing" distribution. From there, an appropriate aggregate model can be derived that describes the distributions of probabilities on these two levels. Parameters can be derived describing this model, and covariates can be inserted to determine whether they affect the parameters. These parameters can ultimately be used to make predictions that help answer questions about the decision process of the individuals. The basic formulation of these concepts and methodology was introduced to a wider audience by Fader and Hardie (2007) in an analysis of customer retention rates (and ultimately customer lifetime value) using the Pareto/NBD model.

In the context of this thesis, a one-choice process is best modeled through a Beta-Binomial distribution. The derivation of the "simpler" Beta-Binomial model, can then be used to supplement the multi-choice version of the D-M model. A Beta-Binomial attempts to explain how many affirmations an individual makes (*x*) given a number of opportunities (*m*). As aforementioned, the individual-level process is best modeled by a Binomial distribution for a given population segment as:

$$P(X_s = x_s | m_s, p_s) = \binom{m_s}{x_s} p_s^{x_s} (1 - p_s)^{m_s - x_s}$$

Where the expected value, or $E(X) = m_S p_s$

This describes the probability of someone in segment *S* making *X* purchases, conditional on a given number of opportunities (*M*) and an underlying unobservable propensity to buy (p_s).

As people either make the choice or do not, an appropriate mixing distribution would be the Beta distribution, which is bound. In this case,

$$g(p_s) = \frac{p_s^{\alpha-1}(1 - p_s)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}$$

Where,

$$E(p_s) = \frac{\alpha}{\alpha + \beta}$$

$$Var(p_s) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

A polarization index can also be formed: $\phi = \frac{1}{1+\alpha+\beta}$

In other words, if the parameters of α, β are low, there results in high polarization and there is a large degree of heterogeneity. If α, β are large, then the population is very homogenous, and there is little weight in posterior mean.

From this, the mixture model as an integration of the two distributions for each segment is:

$$P(X_s = x_s|m_s) = \int_{p_s=0}^{1} \binom{m_s}{x_s} p_s^{x_s}(1-p_s)^{m_s-x_s} \frac{p_s^{\alpha-1}(1-p_s)^{\beta-1}}{B(\alpha,\beta)}\ \mathrm{dp_s}$$

$$= \binom{m_s}{x_s} \frac{1}{B(\alpha,\beta)} \int p_s^{(a_s+x_s)-1}(1-p_s)^{(\beta+m_s-x_s)-1} = \binom{m_s}{x_s} \frac{B(\alpha+x_s, \beta+m_s-x_s)}{B(\alpha,\beta)}$$

With this as a generalized formula:

$$P(X = x|m) = \frac{m!}{x!\,(m-x)!} \times \frac{\Gamma(\alpha+x)\Gamma(\beta+m-x)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+m)\Gamma(\alpha)\Gamma(\beta)}$$

Using the Gamma function property of the factorial, where $n! = \Gamma(n+1)$, this also equates to:

$$P(X = x|m) = \frac{\Gamma(m+1)}{\Gamma(x+1)\Gamma(m-x+1)} \times \frac{\Gamma(\alpha+x)\Gamma(\beta+m-x)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+m)\Gamma(\alpha)\Gamma(\beta)}$$

The Dirichlet-Multinomial distribution is in essence a multidimensional version of the Beta-Binomial. Specifically, it is a multibrand extension of the original choice process. Whereas the Binomial distribution characterizes the individual level choice across only one brand, a Multinomial distribution can involve multiple "brands." The individual has a choice and propensity to vote across a field of parties, a vector of choices. In a Multinomial distribution, this is defined as:

$$P(\vec{X} = \vec{x}|n, \vec{p}) = \frac{n!}{x_1!\,x_2!\,\ldots x_k!} \times p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

Using the Gamma function property:

$$P(\vec{X} = \vec{x}|n, \vec{p}) = \frac{\Gamma(n+1)}{\Pi\{\Gamma(x_i+1)\}} \times \Pi p_i^{x_i}$$

Where *n* is number of purchase opportunities; $\vec{p}$ is vector of propensities to purchase respective products (i.e. which also ultimately has to sum to one across all the products). This is analogous to rolling *k*-sided die *n*-times, where x_i is the number of purchases in brand *i*, rather than the "coin flip" process of a Binomial distribution. Instead of $\frac{m_s!}{(x_s)!(m_s-x_s)!}$ as the combinatoric term, the generalized combinatoric allows for multiple different types of possibilities of "purchases" given the range of "brands" (i.e. the different voting combinations across the various parties).

To model the population level, a Dirichlet distribution is used, a "multibrand" equivalence.

Taking a look first at the Beta distribution:

$$g(p_s) = \frac{p_s^{\alpha-1}(1-p_s)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} = \frac{1}{\frac{\Gamma(\alpha,\beta)}{\Gamma(\alpha)\Gamma(\beta)}} p_s^{\alpha-1}(1-p_s)^{\beta-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha,\beta)} p_s^{\alpha-1}(1-p_s)^{\beta-1}$$

The initial scaling constant can be expanded to a more generalized form, and so can the product of the probabilities to:

$$g(\vec{p}) = \frac{\Gamma(S)}{\Pi\{\Gamma(\alpha)_i\}} \times p_1^{\alpha_1-1} p_2^{\alpha_2-1} \ldots p_k^{\alpha_k-1}$$

Where $\Pi\{\Gamma(\alpha)_i\}$ is the product of the gamma function of each alpha; this is comparable in structure to the Beta distribution. *S* is the sum of all alphas. Note when k=2, the population level model reverts to Beta distribution.

Each α_i describes heterogeneity of preferences for each brand, which is essentially a strength parameter similar to α and β in a Beta distribution.

From this, the derivation of the Multinomial Dirichlet can be compared to that of the Beta-Binomial.

Recall that Beta-Binomial Distribution is:

$$P(X = x|m) = \frac{m!}{x!\,(m-x)!} \times \frac{\Gamma(\alpha+x)\Gamma(\beta+m-x)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+m)\Gamma(\alpha)\Gamma(\beta)}$$

$$= \frac{m!}{x!\,(m-x)!} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+m)} \times \frac{\Gamma(\alpha+x)\Gamma(\beta+m-x)}{\Gamma(\alpha)\Gamma(\beta)}$$

Hence, a generalized version would be:

$$P(\vec{X} = \vec{x}|n) = \frac{n!}{x_1!\,x_2!\ldots x_k!} \times \frac{\Gamma(S)}{\Gamma(S+n)} \times \frac{\Pi\{\Gamma(\alpha_i + x_i)\}}{\Pi\{\Gamma(\alpha_i)\}}$$

Which is essentially the multidimensional beta binomial. The polarization index now becomes is $\phi = \frac{1}{1+S}$, where S is the sum of all alphas (i.e. the preferences). Further, the term $\frac{\alpha_i}{S}$ would give an implied market share of the specific brand *i*, based on the parameters.

Once again, converting this into a form of Gamma functions:

$$P(\vec{X} = \vec{x} | n) = \frac{\Gamma(n+1)}{\Pi\{\Gamma(x_i+1)\}} \times \frac{\Gamma(S)}{\Gamma(S+n)} \times \frac{\Pi\{\Gamma(\alpha_i + x_i)\}}{\Pi\{\Gamma(\alpha_i)\}}$$

The addition of covariates is analogous to determining whether the heterogeneity of the individual level distributions can in some way be explained by exogenous variables. As aforementioned, a possibility could be the unemployment rate of a given voting region, where low job prospects may force citizens to approach more radical parties. Far-right and far-left parties usually espouse positions that enable an autarky or a full-welfare state, usually by rapidly deploying labour under state enterprises and ventures, driving down unemployment rates. Therefore if a covariate was introduced that factored in the unemployment rates across these areas, the polarization index of the model might decrease, highlighting the covariate as an explanation for the heterogeneity across the people. Regardless, the model should consider parsimony: whether the addition of another variable is necessary to tell the "story," or merely unnecessarily complicates the model in order to better fit the data.

The implementation of the covariate into the D-M model is indicated in an unpublished paper by Professor Fader, "Integrating the Dirichlet-Multinomial and Multinomial Logit Models of Brand Choice." A vector of covariates can be added into the equation of the D-M probabilities.

Moving forward with data

There are multiple analyses that can be done with regards to the data. Looking at Europe's two largest multi-party democracies — namely France and Germany — there are various methodologies that can reveal different things about voting propensities, and the covariates that might explain their choices.

There are three key areas that can be explored: 1) The choice of country, 2) the scope of the time periods used, and 3) the consideration of a "household-level" data point.

With regards to the choice of country, the France and Germany serve as ideal starting points simply because of the plethora of data available. Their large populations give multiple data points from which to derive functions and analyses. For France, it makes sense to focus on either the legislative **or** the **first round** of the presidential election. This is as there is a second-round election that occurs if no candidate receives more than 50% of the vote in the first round of a presidential election (which is usually the case).

A Germany case study would focus purely on the federal election, and only on the constituent seat voting. This is as Germany employs a mixed-member proportional representation system, where there are several hundred overhang seats that are allocated amongst parties to fill the exact proportion of voting. Simplistically, each German citizen has two votes, one for their preference of candidate within a particular geographical constituency (of which there are 299 in Germany) and another for their support of a party. The secondary vote can be for a party who has not fielded a candidate in that geographical constituency.

This paper specifically takes the 2017 and 2013 federal level elections in Germany, and applies the aforementioned probabilistic models on them. Only secondary round voting was considered, as the primary round requirements of physical candidacy within geographical constituencies meant that for constituencies where there were no candidates for certain parties, the model would have treated the voter count for that particular party as zero, which would have distorted parameter prediction in the probabilistic model. Looking at the primary vote count would be essentially discounting voter preference for specific parties if they did not field candidates, even if voters within a constituency leant most closely to that party's values. The zero would be a large distorting factor in the model's predictions, especially as the magnitude of the other vote counts in the constituencies would be relatively large.

Secondly, the time period in which the analysis is conducted is important. As evidenced by recent elections, truly unanticipated political events have come to shape recent elections, leading to lopsided and inconsistent voting from the past, which may not work well when presented to this model. The older, more traditional outcomes of the French and German elections (i.e. a split between the Socialists and Republicans, the SPD and the CDU/CSU) would perhaps be more amenable. There is also a question of whether to conduct the analysis over multiple time periods and elections. From this, the change in the parameters can be determined, which has implications on the heterogeneity of the population over time. With the introduction of covariates, this could potentially indicate whether the covariate has had a weakening or strengthening effect on voting patterns over time as well.

Finally, it is imperative to delineate what is considered as the fundamental "household" that is making the choice. Each specific region or area a voting (i.e. purchase) preference across the parties (i.e. brands), which forms the individual level distribution. Groups of these areas then serve as the population (in other words, the country). For France, this may be as small as the commune level (there are 36,681 communes in France), or the higher-level departments (101 in France). For Germany, only the constituency seat level voting is ideal.

Sourcing the data itself should not be too much of an issue. The French election data is available from the Ministry of the Interior, and the German election data is readily available from the Office of the German Federal Returning Officer.[1]

Data Processing

The initial compilation of the data included the input of 299 different geographical constituencies, and their resulting vote count. The aforementioned document broke down each constituency by vote count for

[1] The data can be found here: https://www.bundeswahlleiter.de/en/bundestagswahlen/2017/ergebnisse.html
The Federal Returning Officer. 2017. "Heft 3: Endgültige Ergebnisse nach Wahlkreisen". [English: "Issue 3: Final results by constituencies"] Retrieved December 15, 2017.

the respective parties, as well as voter eligibility and turnout. For the model analysis, the main groupings for the model parameters were the CDU/CSU, SPD, DIE LINKE, GRÜNE, FDP and AfD parties. These are the parties that are currently represented in the German lower house — constitutional stipulations require parties to either achieve more than 3 geographical constituency victories in the primary vote, or a larger than 5% overall national vote in order to be represented in parliament. Other votes were grouped into a bucket "brand" of Others. This was relatively small in proportion to the other parties within each constituency, never gaining more than 5% of the total vote count. The first ten rows of the aggregated results table for each geographical constituency is shown in Figure 1.

Ultimately, three models were created for the 2017 election results. These are, respectively, the D-M model, a multinomial distribution model, and a multiple Beta-Binomial model that assumed independent Beta-Binomial distributions for each "brand". Gammalog functions ("GAMMALN.PRECISE") were used in Excel to calculate the different components of the probabilities, before using the Maximum Likelihood Estimation ("MLE") procedure to determine optimum values for the parameters. The in-built Excel Solver was used to find this optimal solution. As the numbers can get exceedingly large, the analysis was done in log-space, with log-likelihoods calculated. Logarithm properties were used to break down the equations into negotiable components. For example, the D-M model's probability function:

$$P\left(\vec{X} = \vec{x} \middle| n\right) = \frac{\Gamma(n+1)}{\Pi\{\Gamma(x_i+1)\}} \times \frac{\Gamma(S)}{\Gamma(S+n)} \times \frac{\Pi\{\Gamma(\alpha_i + x_i)\}}{\Pi\{\Gamma(\alpha_i)\}}$$

$$\ln P\left(\vec{X} = \vec{x} \middle| n\right) = \ln\left(\frac{\Gamma(n+1)}{\Pi\{\Gamma(x_i+1)\}}\right) + \ln\left(\frac{\Gamma(S)}{\Gamma(S+n)}\right) + \ln\left(\frac{\Pi\{\Gamma(\alpha_i + x_i)\}}{\Pi\{\Gamma(\alpha_i)\}}\right)$$

$$\ln P\left(\vec{X} = \vec{x} \middle| n\right) = \ln(\Gamma(n+1)) - \ln(\Pi\{\Gamma(x_i+1)\}) + \ln(\Gamma(S)) - \ln(\Gamma(S+n)) + \ln(\Pi\{\Gamma(\alpha_i + x_i)\}) - \ln(\Pi\{\Gamma(\alpha_i)\})$$

Similar breakdowns were used in the calculations for the multinomial distribution and multiple Beta-Binomial models. A 2013 D-M model was also created, to see if there were any parameter or polarization shifts within the population.

To compare the robustness of the different models, the log-likelihood ("LL") amounts can be compared, as the D-M model is a nested model within the larger multinomial distribution. The Bayesian information criterion (BIC) can be used as a comparator between the different models as well. This is calculated as: $BIC = -2 \times LL + \#\ of\ variables * \ln(\#\ of\ constituencies)$, where LL is the log-likelihood. A higher LL and lower BIC imply a more robust model.

Further, a D-M two (or multiple) segment model can be created, by doubling the number of parameters and implementing two segments across the data. A reasonable explanation for this could be two inherently different types of populations, with one having a rigorous preference for certain parties, and the other having diverging preferences from the first segment. In a similar fashion, a multiple-spike D-M model can be created. This assumes that there is an inherent core population that strictly supports one party, and that the

Data Analysis

The initial multinomial model can be seen in Figure 2. It offers a LL of -2,353,847 and a BID of 4,707,729. In comparison, the D-M model (Figure 3) offers a LL of -17,994 and BIC of 36,028, a marked improvement. This comes at the expense of one additional variable. This showcases the robustness of the D-M model in using minimal amounts of parameters to estimate a distribution that fits the data. Interestingly and tellingly, the sum of alphas is quite high, at ~61. This implies that there is little heterogeneity between geographical constituency voting in 2017. Rather, the population overall seems very homogenous. As the models indicated that the population was rather homogenous in preference, there was no reason to conduct a further test in adding covariates to the model. If this were done, this would create parameters that would imply an even more homogenous population.

A multiple Beta-Binomial model was also created for the 2017 data (Figure 4). By using the parameters derived from the D-M model, an implied multiple Beta-Binomial model can be created (Figure 5), where the implied beta (under the definition of a Beta-Binomial model) is the sum of all alphas ("S") minus the alpha value. This generates an implied LL for each of the groupings/parties/"brands." By comparing these against the actual LLs calculated for the multiple Beta-Binomial model, determinations can be made on the over- or under- estimation of the D-M model.

In this case, the model reveals an over prediction in the FDP (11.0% implied vote share, versus 10.7% actual vote share) and for the "Other" parties grouping (5.5% implied vote share, versus 5.0% actual vote share). The reasonings for the FDP could potentially be the political climate of Germany. The FDP, a traditionally centrist party, has come under criticism for being having indistinct party policies that failed to invigorate a voter base. News publications such as Der Spiegel — a preeminent German weekly news magazine — noted at the time of the 2013 election that "apart from low taxes and deregulation, the party's agenda has become less than clear." (Weiland 2013). In fact the party did so poorly that they failed to reach the 5% national voting threshold to even send parliamentarians into the Bundestag. The 2017 election result could be a continuation of the voter apathy to the FDP. There is an intrinsic depression in voter support for the FDP in the 2017 election cycle. The lower than expected voter share for the other parties could simply be explained as there is more brand recognition for the more established parties. Hence voter support for the other fringe parties or candidates is naturally lower than expected.

A further treatment was also conducted on the D-M model for 2017, looking at the possibility of spikes or secondary segments. Interestingly, the two segment D-M model collapses into one segment. The model rejects the creation of a second segment by allocating close to zero percentage likelihood for the second segment when solving for the highest LL. This reinforces the narrative of a homogenous population in terms of voting preference. Similarly, the introduction of spikes (specifically for the CDU/CSU and the SPD; this was done as they were by far the consistently largest parties within the parliament) to account for die-hard voters of the center-right and center-left parties respectively did not resolve to a differentiated model. The model once again broke down to a one segment model.

A D-M model for 2013 was also made (Figure 6). The parameters did not greatly vary from the 2017 results, implying a lack of voting heterogeneity within the constituent population.

Conclusions

From the analysis, it is clear that the models tend to converge to parameters that indicate a homogenous voting population, across the geographic constituencies. There is no particularly polarized constituency that votes specifically for a right-wing or left-wing group. Instead, the constituencies, which number around 150,000-200,000 voters in each can be seen as microcosms of Germany at-large. The D-M model implies that, for Germany at least, views are held in relatively similar amounts across the country. From this homogeneity across constituencies, it seems that unemployment rates and other covariates are not necessary to explain segmentation — of the little that there is — within Germany.

This may be different from a two-party system such as the United States. Nevertheless, a similar analysis could be conducted on a county-level basis across the United States to determine polarization.

References

Alvarez, R. Michael, and Jonathan Nagler. "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science* 42, no. 1 (1998): 55-96. doi:10.2307/2991747.

Dow, Jay K., and James W. Endersby. "Multinomial probit and multinomial logit: a comparison of choice models for voting research." *Electoral studies* 23, no. 1 (2004): 107-122.

Fader, Peter S. "Integrating the Dirichlet-Multinomial and Multinomial Logit Models of Brand Choice" Unpublished.

The Federal Returning Officer. 2017. "Heft 3: Endgültige Ergebnisse nach Wahlkreisen". [English: "Issue 3: Final results by constituencies"]. Retrieved December 15, 2017. https://www.bundeswahlleiter.de/en/bundestagswahlen/2017/ergebnisse.html

Weiland, Severin. "Free Democrats Reel From Election Fiasco." *Spiegel Online*. Retrieved April 29, 2018. http://www.spiegel.de/international/germany/german-free-democrats-fail-to-enter-parliament-a-923816.html

Appendix

Note: if unspecified, all parameters are in the order of party columns: CDU/CSU, SPD, DIE LINKE, GRÜNE, FDP, AfD, and Others

Figure 1.

Aggregated Constituency Results | Secondary Vote Analysis | 2017

Actual voting share						
0.3292955	0.2050797	0.0923836	0.0893982	0.1074792	0.126369	0.0499948

Constituency Number	Constituency Name	CDU/CSU	SPD	DIE LINKE	GRÜNE	FDP	AfD	Others	Total
		Voting Proportion, only counting valid votes | Secondary Vote							
1.3.1	Flensburg - Schles	58320	40388	14002	22304	18955	11653	4843	170465
1.3.2	Nordfriesland - Di	52928	31120	8589	15144	18050	9030	3210	138071
1.3.3	Steinburg - Dithm	47366	29756	8732	12960	17298	11180	3586	130878
1.3.4	Rendsburg-Eckern	56585	35766	9962	19337	19071	11578	3968	156267
1.3.5	Kiel	40736	36208	15546	26143	17804	10504	5128	152069
1.3.6	Plön - Neumünster	43778	31013	8503	16350	16481	11161	3228	130514
1.3.7	Pinneberg	63863	42729	13111	21336	24735	15977	4621	186372
1.3.8	Segeberg - Storma	66367	43027	13237	21010	26043	17166	5095	191945
1.3.9	Ostholstein - Stor	48898	33764	8303	13493	18147	11782	2908	137295
1.3.10	Herzogtum Lauent	66031	42815	12480	20826	26163	18792	4864	191971

Figure 2.

Multinomial Distribution Model Results | Secondary Vote Analysis | 2017

	D-M Model						
ln(theta)	1.8850206	1.4114648	0.6140102	0.5811698	0.765367	0.927274	0
p	0.3292949	0.2050795	0.092383	0.0893984	0.1074795	0.1263691	0.0499955
Actual voting shar	0.3292955	0.2050797	0.0923836	0.0893982	0.1074792	0.126369	0.0499948

Sum of exp(ln(theta))	20.001798
LogLikelihood	-2353847.4
BIC	4707729

Figure 3.

Dirichlet-Multinomial Distribution Model Results | Secondary Vote Analysis | 2017

	D-M Model						
Alpha	19.980225	12.360587	5.5127568	5.2344711	6.712733	7.628328	3.3401738
a/S	0.3287883	0.2034019	0.0907162	0.0861368	0.1104626	0.1255294	0.0549648
Actual voting shar	0.3292955	0.2050797	0.0923836	0.0893982	0.1074792	0.126369	0.0499948

S:	60.769275
LogLikelihood	-17994.076
BIC	36028.054

Figure 4.

Multiple Beta-Binomial Distribution Model Results | Secondary Vote Analysis | 2017

	D-M Model						
alpha	20.030229	8.4956553	4.2717825	4.3613408	13.975195	5.7576563	8.9459024
beta	40.893631	32.64814	41.63418	45.33957	117.13402	39.225289	169.97626
Mean from parame	0.3287748	0.2064869	0.0930551	0.0877517	0.106592	0.1279964	0.0499988
Actual voting shar	0.3292955	0.2050797	0.0923836	0.0893982	0.1074792	0.126369	0.0499948

Figure 5.

Multiple Beta-Binomial Distribution versus Dirichlet-Multinomial Distribution Model Comparison | Secondary Vote Analysis | 2017

	D-M Model						
alpha	19.980225	12.360587	5.5127568	5.2344711	6.712733	7.628328	3.3401738
Implied beta	40.789049	48.408688	55.256518	55.534804	54.056542	53.140947	57.429101
Resulting loglikeli	**-3152.646**	**-3174.312**	**-3039.339**	**-3016.662**	**-2944.347**	**-3091.893**	**-2816.342**
Original loglikelih	**-3152.645**	**-3161.114**	**-3032.939**	**-3013.473**	**-2909.671**	**-3084.386**	**-2754.797**
Mean from implied	0.3287883	0.2034019	0.0907162	0.0861368	0.1104626	0.1255294	0.0549648
Actual voting shar	0.3292955	0.2050797	0.0923836	0.0893982	0.1074792	0.126369	0.0499948

S:	60.769275
Sum of loglil	**-21235.54**
Sum of loglil	**-21109.03**

Figure 6.

Dirichlet-Multinomial Model Distribution Results | Secondary Vote Analysis | 2013

	D-M Model						
Alpha	24.959461	15.276938	4.5953745	5.0673366	3.153747	3.228469	4.0702812
a/S	0.4135675	0.2531323	0.0761434	0.0839636	0.0522562	0.0534943	0.0674428
Actual voting shar	0.41543	0.2573296	0.08589	0.0844803	0.0476488	0.0470417	0.0621797

S:	60.351608
LogLikelihood	-17684.012
BIC	35407.928

Excel models with granular constituency level data and other models attached below [electronic versions only].

SummaryDocVFinal.xlsx