

Chapter 2

Fuzzy Preferences: Extraction from Data and Their Use in Public Choice Models

Abstract We describe a method for extracting fuzzy preferences from the Comparative Manifesto Project (CMP) data that makes use of the bootstrap procedure designed by Benoit et al. (2009). We argue that fuzzy preferences are a better representation of the abstract concept of a player's preferences in public choice models. Instead of representing preferences as precise points, our fuzzy approach maps them as bounded areas in a subset of \mathbb{R}^k . In so doing, we eschew the conventional assumption that political actors have precise policy positions. Instead, fuzzy preferences permit us to conceive of actor's preferences as vague, but communicated accurately. We conclude the chapter by introducing our basic approach to using fuzzy preferences in fuzzy public choice models. We argue that a fuzzy public choice model satisfies some of the intuitive and practical problems faced by the conventional model. Moreover, a fuzzy public choice model allows us to shed the assumption that actors perceive shifts in utility in infinitely precise increments at the same granularity across and infinite policy space.

2.1 Introduction

A number of methods have been designed for the extraction of preference measures from empirical data. These measures represent players' preferences as precise "crisp" points in an infinite Euclidean space. As a consequence, the predictions made in these "crisp" models suffer under the assumptions of traditional \mathbb{R}^k space. In contrast, fuzzy public choice models represent players' preferences as bounded areas in a subset of \mathbb{R}^k . Such models require a new method for extracting fuzzy preference measures from raw data.

This chapter presents a method for deriving fuzzy preference measures from raw text data. Rather than starting at square-one and collecting a new, raw dataset, we "fuzzify" an existing dataset, Comparative Manifesto Project (CMP) data. By doing so, we avoid the problems of selecting and collecting appropriate data. We may also

be able to avoid some of the problems that have already been resolved in existing datasets.

We select CMP data because of its breadth and availability. It has also been applied broadly and shown to perform better when estimating the preferences of political actors in democratic systems outside of the United States. Moreover, the CMP offers an interesting opportunity to extract fuzzy preferences because it derives crisp preference measures from inherently fuzzy text data.

The CMP project involved human coding of text data in the form of manifestos published by political parties that detail the issues and policies the party promotes in an election year. It is in the stochastic process of writing the text, or text generation, that Beniot, Laver, and Mikhaylov (Beniot et al. 2009) find non-systematic (unbiased) error introduced into preference measures derived from text data. Language is inherently vague, and it is possible that, in the drafting and re-drafting of a political text, an author may not represent his preferences “precisely.” Beniot, Laver, and Mikhaylov develop a bootstrapping method that replicates this stochastic process to calculate error scores for the CMP data. They find that much of the variation in parties’ preferences over time can be accounted for by error, and that even the preferences of parties competing in the same country in the same election year may be statistically indistinguishable, making decision-making analysis based on CMP data impossible.

However, a fuzzy approach allows us to salvage the CMP data for application in decision-making analysis. Instead of conceiving actors’ preferences as precise positions in policy space from which each incremental movement is less preferred, the fuzzy public choice model represents actors’ preferences as bounded subsets of policy space over which players perceive shifts in utility. Instead of treating uncertainty as “noise” that eliminates the CMP’s ability to identify precise policy positions, the fuzzy model uses uncertainty in a party’s preferences to determine how the party perceives shifts in utility over a subset of policy space.

We utilize the bootstrapping method designed by Beniot, Laver, and Mikhaylov to derive fuzzy preference measures from the CMP data. This method projects a distribution of the possible outcomes of text generation. We then use this distribution of policy positions to determine the bounds of α -levels of the fuzzy numbers representing each party’s preference profile. Finally, we apply these empirically-derived fuzzy preference measures in a public choice model. In later chapters, we further develop the basic model to permit predicting the outcome of the government formation process in parliamentary democracies.

2.2 Extracting Preference Measures from Empirical Data

There is a broad range of data and methods for collecting data available for the purpose of measuring the preferences of political players. All of these data and extraction methods have their benefits and disadvantages, but some data seem to be more amenable to the extraction of fuzzy preference measures than others. The most

commonly used data for the purpose of measuring the preferences of political actors are roll-call votes, expert surveys, legislator surveys, and political texts.

The use of roll-call data to extract political actors' policy preferences has a number of advantages and disadvantages. Roll-call votes are the spoken "Yea or Nay" votes of individual legislators given on request for a piece of legislation. Data on roll-call votes is widespread and accessible. Processes such as NOMINATE (Poole and Rosenthal 2007) and Optimal Classification (Poole 2005) use roll-call data to locate players' ideal points in policy space. The more roll-call data available to such programs, the greater the precision of their estimates.

Despite its widespread use, a number of criticisms have been levied against the use of roll-call data to identify players' preferences. First of all, there are a number of institutional, political, and strategic factors that may influence the way a player votes. Votes may not reflect players' sincere preferences, but may reflect their preferences as they are translated by the institutional and political constraints within which they are voting. There is a further causal problem inherent in using roll-call votes as inputs to describe the behavior of actors in political situations when it is the same political context from which the vote was a behavioral output (Laver 2001, p. 239).

Furthermore, roll-call data may not be amenable to the extraction of fuzzy preference measures. Roll-call data is based on a system of "Yea," "Nay," or "Abstain." The values "Yea" or "Nay" are dichotomous and represent crisp positions comparable to those used in traditional set theory from which Euclidean models are derived. A vote "Yea" puts a policy fully within the set of a player's preferences (i.e., $\alpha = 1$) and a vote "Nay" puts a policy strictly out of a player's preferences (i.e., $\alpha = 0$). While such data can be used to identify the core and the support of a fuzzy number, the ability to identify other discrete α -levels is limited.

Nevertheless, Potter (2007) extracts fuzzy measures of individual legislators' preferences from roll-call data. Potter assigns a vote 'Abstain' to an α -level of $\alpha = 0.5$, assuming that abstaining from a vote is the same as being indifferent. However, such an assumption is problematic in that the choice to abstain from a vote can be a sign of disapproval, and even an attempt to undermine the quorum necessary to hold a vote. When applied to NOMINATE and OC, a vote "Abstain" is coded as missing data. Without the ability to define α -levels other than $\alpha = 1$ and $\alpha = 0$, much of the power and nuance of the fuzzy model is lost.

A better alternative may be to derive preference information from expert or legislator surveys. Experts can respond to survey questions that provide an intensity scale that can easily be applied to the discrete α -levels derived from players' fuzzy preferences. However, expert surveys are subject to the same criticism as roll-call data in that experts' judgments are derived from the same actions they are seeking to explain. Beyond this potential problem of tautology, expert survey data can be expensive and difficult to gather, making its accessibility limited.

Furthermore, survey data is not replicable. Once a survey has been taken at a specific time t_1 , it cannot be taken at that time again. Instead any survey taken at a later time t_2 , has to be understood to be different and informed differently than the original survey. This is of a particular problem if we decide to expand our set of questions. For example, consider if we would like to add a new policy dimension

“Immigration” to our models of political decision-making. Without past questions about immigration policy, we cannot hope to compare surveys at t_1 and t_2 . Since it is impossible to go back and add questions about immigration to the survey at t_1 , we are not able to compare the policy positions of actors at these two different times. Due to the constraints upon gathering survey data and the continuing problem of tautology, expert data is less than ideal.

Alternatively, surveys of actual political actors allow us to capture the same kinds of scaling as expert data and avoid the problem of causal tautology. By surveying political actors directly, scholars can collect data about their preferences does not result from actions within a specific political context. However, legislator surveys are subject to the same problems of resource-intensiveness, time-sensitivity, and lack of replicability as expert surveys, and therefore are impractical for the extraction of preference measures.

Political texts are a good alternative to the other forms of data used for measuring players’ preferences. Political texts published by political actors are widely available and accessible in most developed democracies, and can be obtained with limited resources. The causal tautology is escaped, because the inputs for the prediction of players’ behavior are no longer the outputs of their behavior in the same political context. Also, the policy emphases of political parties in political texts tells us which issues they consider salient, giving us a sense of what issues we should consider plausible dimensions at that time. Therefore, we know what issue dimensions to consider, or can infer a player’s indifference to an issue dimension, without having to turn to direct questions about the issue. This releases us from the time-sensitivity and replicability problem of survey data. Finally, differing degrees of intensity of preferences can be derived based on the policy emphases in the language of the texts, making it particularly amenable to the measurement of fuzzy preferences.

2.2.1 Extracting Preferences from the CMP Data

Among the most widely used data sets of political texts from which preferences measures have been derived is the Comparative Manifesto Project (CMP). The CMP project was undertaken by a research group concerned with locating the policy positions of political parties based on the policy emphases within the parties’ electoral manifestos. An electoral manifesto is a document released by a political party prior to an election that contains a wide range of authoritative and representative statements about the policy positions of the political party as a whole (Budge et al. 2001; Klingemann et al. 2006). The CMP represents its cases as country-party-year. All salient parties from every free democratic election from 1945 to 2005 are included. A salient party is any political party that is either likely to be a member of a governing coalition or that can affect the outcome of a governing coalition. Thus, each case represents a salient political party in a country each year it competes in a free democratic election. The CMP dataset includes 780 political parties from 54 countries in 529 election years for a total of 3,108 cases (country-party-years).

Each party manifesto in the CMP database is parsed into “quasi-sentences”—or phrases that reflect complete thoughts—by human coders and placed numerically into one of 56 policy dimensions, or designated as “unknown” in the case of insufficient data (Budge et al. 2001; Klingemann et al. 2006). Insufficient or missing data are addressed by collecting and analyzing supplemental party materials. When election manifestos released by the political parties themselves are unavailable, the CMP seeks published documents fulfilling the above criteria. The CMP collects data only for significant political parties which are identified based on their potential for being members of a governing coalition or their ability to affect the tactics of parties competing for a position in the government.

Quasi-sentences are the basic unit of meaning identified and coded by the CMP. They can be understood as “complete thoughts” either in the form of sentences or sentence fragments (Budge et al. 2001; Klingemann et al. 2006). CMP coders placed each quasi-sentence on the policy dimension on which it best fits. The sum of the quasi-sentences assigned to each party is that party’s score on that dimension.

Subsets of the 56 dimensions coded in the CMP database, or *riles*, can be created to compare parties on any number of political concepts. Among the most popular of such *riles* is a single-dimensional left-right political scale developed by Laver and Budge (1992, pp. 23–30). They constructed the left-right scalar using exploratory factor analyses to determine whether issues that are theoretically related share a common dimension empirically. They then construct the *rile* by adding scores on common dimensions. The left-wing policy categories include an emphasis on democracy, state intervention, and peace and cooperation as well as a positive orientation toward social services, state education, and labor groups.¹ The right-wing policy categories include an emphasis on freedom, domestic human rights, the military, capitalist economics, and social conservatism.² The position of political parties on the *rile* dimension is determined by the sum of the proportion of the manifesto devoted to right-wing policy references minus the sum of the proportion devoted to left-wing references.

A number of criticisms have been levied against the process employed by the CMP. The strongest criticism of the CMP is the problem of inter-coder reliability. The inter-coder reliability problem refers to the dilemma that two different coders or

¹ State intervention and Peace and cooperation are additive dimensions established by exploratory factor analyses of theoretically similar CMP policy categories. State intervention includes the original CMP categories Regulation of capitalism (PER403), Economic planning (PER404), Protectionism: positive (PER406), Controlled economy (PER412), and Nationalization (PER413). Peace and cooperation is constructed from the categories Decolonization (PER103), Military: negative (PER105), Peace (PER106), and Internationalism: positive (PER107).

² Capitalist economics and Social conservatism are additive dimensions established by exploratory factor analyses of theoretically similar CMP policy categories. Capitalist economics includes the CMP categories Free enterprise (PER401), Incentives (PER402), Protectionism: negative (PER407), Economic orthodoxy and efficiency (PER414), and Social services expansion: negative (PER 505). Social conservatism includes the categories Constitutionalism: positive (PER203), Government effectiveness and authority (PER305), National way of life: positive (PER601), Traditional morality: positive (PER603), Law and order (PER 605), and National effort, social harmony (PER606).

the same coder at two different times may code the same text differently (Mikhaylov et al. 2012). Error resulting from the identification and coding of quasi-sentences reflects uncertainty about the accuracy of the coders' reading of the text, but it does not tell us anything about the political actor's preferences. Moreover, it is specific to the CMP coding process.

Benoit, Laver, and Mikhaylov (BLM) (2009) argue that error may result as well from published texts failing to communicate the preferences of the party accurately. This type of error, error introduced by the process of authorship or "text generation" (Benoit et al. 2009), reflects uncertainty about the relationship between the author's intended message and communicated message and is problematic for any process extracting preference measures from text data. In order to get at this problem, BLM develop a method to measure uncertainty related to this type of error.

BLM (2009) describe text generation as a stochastic process by which an author attempts to communicate his policy position by writing a text. The authors start with an intended message μ that he seeks to communicate. BLM assume that the authors intended message μ is the author's actual policy position, and that the author is not dissembling. We can imagine that the process of authorship involves a number of different versions of any text, including intended drafts in the mind of the author, drafts written and stored by the author, written and thrown away, and the actual text that is published coded by the CMP. Each one of these texts communicates a policy position that is an approximation of the author's intended message μ . The process of communicating the intended message μ with some text τ is stochastic and therefore allows for random error to enter into the information communicated in τ . Through the stochastic process of text generation, authors may accidentally misallocate quasi-sentences to the different policy categories they mention, thus overemphasizing some policies and underemphasizing others. This results in the author communicating a policy position not in-line with their intended message. Therefore, the policy position communicated by the text may be another policy position on the relevant dimension with some random error differentiating it from the author's actual position. By this logic, any attempt to derive the true preferences of a political actor from a political text will be affected by some random error.

BLM use a bootstrapping procedure to derive uncertainty scores based on the error in text data-derived preference measures caused by the stochastic process of text generation. Salience theory, the theoretical grounding of the CMP, assumes that authors emphasize certain policies by making reference to them.³ Authors only mention policies they support, and the more an author mentions a policy (i.e., the more quasi-sentences the author commits to the policy), the more strongly the author prefers it. The bootstrap procedure simulates the stochastic text generation process by taking the number of quasi-sentences in a document and the categories to which those quasi-sentences are assigned and then randomly reassigning the quasi-sentences to the set of categories. This process is repeated 1,000 times to produce a standard error of policy uncertainty, which is assumed to represent the error, or noise, in the CMP data.

³ Budge et al. (2001), Klingemann et al. (2006).

As an illustration of the bootstrap procedure, assume a document of 20 quasi-sentences assigned to four different categories from the CMP; we will call these categories *a*, *b*, *c*, and *d*. Assume that the quasi-sentences in the text are distributed equally across the four issue categories, such that five quasi-sentences are devoted to each category. The bootstrap procedure would then hold the number of quasi-sentences in the document constant at 20, but reassign them randomly to each of the four issue categories. The bootstrap does not assign quasi-sentences to any issue category not included in the original coding, assuming that, by not mentioning an issue category, a party is explicitly not showing support for that policy. So an example of a bootstrap reiteration of the text in this case would be the reassignment of the quasi-sentences from $a = 5$, $b = 5$, $c = 5$, and $d = 5$ to $a = 6$, $b = 4$, $c = 3$, and $d = 7$. The bootstrap procedure simulates the stochastic process of text generation and reiterates it 1,000 times. The procedure then calculates the standard error from the points projected from the set of iterations to derive an uncertainty score for the data.

BLM conclude that much of the perceived change in parties' ideal policy positions over time calculated by the CMP can actually be accounted for by unbiased error introduced by the stochastic text generation process. In fact, BLM find that observed changes in the parties' policy positions are statistically significant in only 28.1% (less than one-third) of the cases they investigate.⁴ Thus, BLM conclude that much of the observed changes in parties' policy positions over time in the CMP data is "noise." Similarly, this causes problems for comparing parties' policy positions in a specific country and election year in order to make inferences about political outcomes such as government formation and legislation. BLM demonstrate that, due to unbiased error introduced by the stochastic text generation process, many parties' policy positions are no longer statistically distinguishable. Because Euclidean public choice models are based on differences between actors' policy positions measured in distance, such error renders the use of CMP scores to determine decision-making outcomes impossible. In fact, even if parties' policy positions are still statistically distinguishable, the inability to determine the precise position of each parties' ideal point in policy space makes determining the distance between parties' ideal points impossible. Thus, the unbiased error introduced into the CMP data by the stochastic text generation process that is calculated by BLM demonstrates that the CMP data is unusable for Euclidean public choice models.

The bootstrap procedure created by BLM (2009) is a major contribution to studies utilizing not only CMP data but all methods of extracting preference information from texts. However, their analysis of how error is introduced into preference data by the stochastic text generation process relies on some fundamental assumptions about how humans think and communicate. First, BLM assume that the text author's intended message (and, by assumption, his actual policy position) is a single precise point in policy space. Second, they assume that each text communicates a single

⁴ BLM only consider cases without the extended categories add in the second-generation CMP data (Klingemann et al. 2006) and do not consider cases in which 99.99% of quasi-sentences are uncoded (Sweden from 1948 to 1982 and all Norway).

precise point in the policy space. By these assumptions, it would be possible for an author to communicate his intended message μ in a text τ if he used the right language. However, BLM assume that the text generation process is stochastic and thereby introduces error into the communication of preferences.

Moreover, Benoit et al.'s bootstrapping procedure creates a significant problem for the conventional public choice models. If, as they assume, the error derived from the bootstrapping represents the writer's inability to adequately communicate their precise policy position in Euclidean space, then the further assumption is that a party's policy position does not change between election years. Instead, any change in quasi-sentence allocation is error. However, without variability in party policy positions, the CMP data becomes less than ideal for estimating preferences as the basis upon which outcomes will be predicted by public choice models.

2.2.2 A Method for Extracting Fuzzy Preference Profiles from CMP Data

The assumptions of BLM are essentially Euclidean in nature. In the Euclidean tradition, BLM conceive of each author's real policy position, intended message, and policy position communicated by the author's text as precise points in policy space. Therefore, error introduced into the CMP data by the stochastic text generation process comes from a statistically significant difference between the ideal point position of the author's intended message and the text's message.

A fuzzy approach to estimating preferences would permit us to conceive of the uncertainty caused by the text generation process not as error between the text's message and the author's intended message, but rather as uncertainty within the author's actual preferences. Thus, rather than treat actor's policy positions as precise points in policy space as in Euclidean models, uncertainty in actors' preferences can be represented as subsets of policy space over which the actors perceive shifts in utility. In this way, vagueness about the "precise" position of an actor's ideal point is not problematic, and in fact is information about the actor's preferences and the likelihood that the actor will agree with other actors. In essence, a fuzzy approach invites us to interpret the stochastic text generation process other than authors erring in their emphasis of certain policy positions. Rather, we see them as flexible in the priority they place upon certain policies.

The bootstrap method designed by BLM estimates error in the CMP data by simulating the stochastic process of text generation for each case in 1,000 reiterations. Each simulation predicts a point meant to represent a possible policy position communicated in a text following the text generation process. This process produces a distribution of points around the point estimated by the CMP (treated as the median of the distribution). From this distribution, a density function can be derived from which BLM estimate error in the CMP data introduced by the stochastic text generation process (depicted in Fig. 2.1).

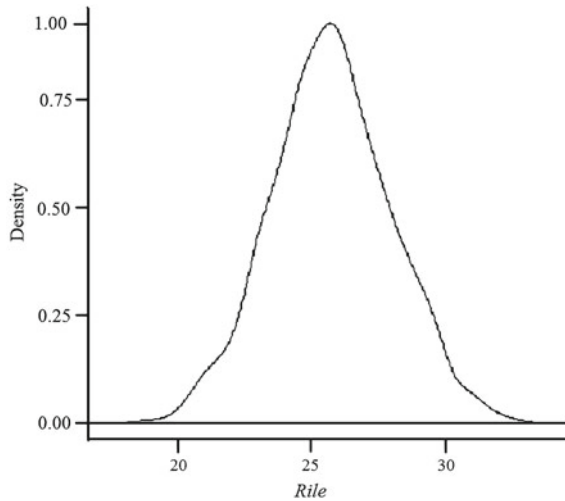


Fig. 2.1 Density function derived from bootstrapping procedure

In Fig. 2.1, there are two axes. The x -axis is the *rile* issue dimension upon which the bootstrap procedure estimates the placement of policy positions communicated by the stochastic text generation process. The *rile* dimension (x -axis) is used for the categories of random distribution of quasi-sentences per manifesto. The y axis represents the density distribution of quasi-sentences after 1,000 iterations. The y -axis is the density of the predicted policy positions at a certain point on the x -axis. The density function is single-peaked and resembles a normal curve.

We use the density function resulting from BLM's bootstrap procedure to derive a fuzzy preference. Again, we place the density function in two-dimensional space. We use the y -axis of the function, the density of the points projected on the x -axis, as a surrogate for the party's utility if policy is located on that position of the issue dimension. Therefore, we assume the denser the distribution of points at a position, the more that position is preferred by a party. We identify the highest density position as the set of policies perfectly in the set of the party's preferences with an α -level of $\alpha = 1.00$. The boundary of the density function, where distribution of points ends, we identify as the boundary of the fuzzy number, beyond which the party's utility is $\alpha = 0.00$. Fuzzy preference profiles are created with the introduction of cut points at density levels 0.25, 0.5, and 0.75 to begin creating the alpha levels of indifference. Once the alpha cuts are made to the density functions, we further adjust the preference profile to represent areas of indifference in policy space. This is done by reshaping the density function between alpha cuts into rectangles with the same area as the original curve.

Figure 2.2 illustrates the process. Between $\alpha = 1.00$ and $\alpha = 0.00$ we make four equidistant cuts (α -cuts) perpendicular to the y -axis; $\alpha = 0.75$, $\alpha = 0.50$, and $\alpha = 0.25$ (depicted in Fig. 2.2a). The area and center of gravity for the segment

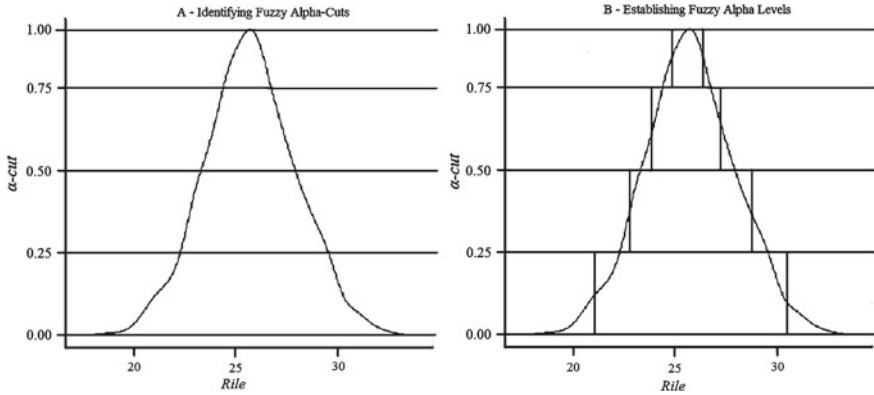


Fig. 2.2 Deriving fuzzy numbers from bootstrap density function

of the distribution sandwiched between each pair of α -cuts. The function is then re-shaped into a rectangular shape with the same area and positioned at the same center of gravity (see Fig. 2.2b). This process allows us to shape the density function into a discrete fuzzy number with minimal loss of information. Each party's fuzzy preference profile was then placed on a *rile* dimension with other parties from its country and the year it competed in the election.

2.3 The Conventional Public Choice Model

Now that we have a means to extract fuzzy preferences from CMP data, we are in a position to employ them in public choice models to see if we can improve our predictions. We conclude this chapter by introducing the book's approach to using fuzzy preferences in public choice models. We begin by giving considering the conventional public choice model. We then juxtapose our fuzzy public choice model against the conventional approach.

Traditional Euclidean public choice models are based on the assumption that there exists a universal set of policies X , such that, for all policies x and y in X , any player i has a binary preference relation R over these policies Austen-Smith and Banks (1999, pp. 1–3). In other words, X is a set of all conceivable policies and x and y are policies.

Any political player perceives x and y and has some preference relation R between them, for example xRy . The statement xRy means that “ x is at least as good as y .” If, for some player i , xRy and yRx we say that “ x and y are at least as good as each other” or “ i is indifferent between x and y ” and we use the binary relation I to write the statement xIy . If, instead, xRy and not yRx then we say “ x is strictly preferred to y ” and we use the binary relation P to write the statement xPy (Austen-Smith and Banks 1999, p. 2).

Implicit in the statement we just made is the assumption of completeness: that for all policies x and y in X , a player i holds some preference relation over them such that either xRy or yRx or both (p. 3). In addition, the traditional model necessitates the assumption of reflexivity, xRx , that x is at least as good as itself.

When developing a public choice model, we are concerned with the preferences of individual players only in so far as their preferences determine the group preference. When seeking a group preference we are interested in the existence of a maximal set. Given some decision-rule, the maximal set is the set of all policies x in X such that, for all policies y in X , x is at least as good as y , or xRy . In other words, the maximal set is the set of all policies for which there is no other policy that the group strictly prefers.

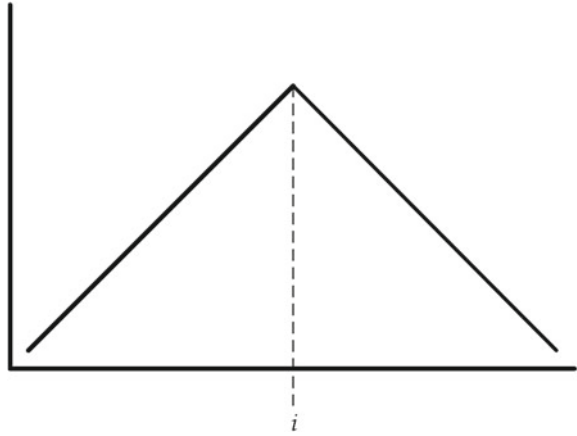
The existence of a maximal set requires that the group preference be complete and reflexive. In much of the literature on models of public choice, transitivity is considered a requirement for a maximal set to exist. When we assume rational individuals, we assume that their preferences are transitive. That means, if player i prefers x to y (xRy) and y to z (yRz), then i must also prefer x to z (xRz), for a complete preference profile of $xRyRz$.

However, the rational choice assumption applies to individuals, not groups (see Arrow 1951). Therefore, we cannot assume that the group's preferences are transitive. This leaves the possibility that a group's preferences may cycle. Cycling occurs when the group's preferences are intransitive; for example, the group prefers x to y and y to z , but prefers z to x ; or the group's preferences are $xRyRzRx$. In such a case, there is no element in the maximal set; the maximal set is empty. Austen-Smith and Banks (1999, pp. 4–6) demonstrate that transitivity is a sufficient but not necessary requirement for a maximal set to exist. A weaker assumption, that group preferences are acyclic, is sufficient for a maximal set to exist. Acyclicity assumes that for all policies x, y, z, w in X , if $xPyPzPw$ then xRw . If a group's preferences are complete, reflexive, and acyclic, then the maximal set is said to be nonempty.

While acyclicity is a minimal requirement for a nonempty maximal set, there is nothing in the definition of group preference to guarantee that they will be acyclic. To guarantee acyclic preferences, the Euclidean model assumes that the preferences of individuals in a group are single-peaked and monotonic. If a player's preferences are single-peaked and monotonic, this means that there is some point in the policy space that the player holds as ideal and that every incremental movement away from this point in the policy space is strictly less preferred than any point closer to the player's ideal. This is also called the Euclidean distance assumption.

The Euclidean distance assumption is based on the idea that players view preferences as if they existed in an n -dimensional space, where n is any number of dimensions. The simplest Euclidean models feature two dimensions in which the x -axis is a continuum of possible policies and the y -axis is the player's utility at that given policy. In this book, we follow the convention in political science of labeling these models single-dimensional because only one *policy* dimension is depicted. Figure 2.3 depicts a single-dimensional model in which the x -axis is some policy and the y -axis is the player's utility. The single-peaked preference profile for a single player is depicted in the model. Along the x -axis are a number of points in the

Fig. 2.3 Example of a Single-peaked preference profile

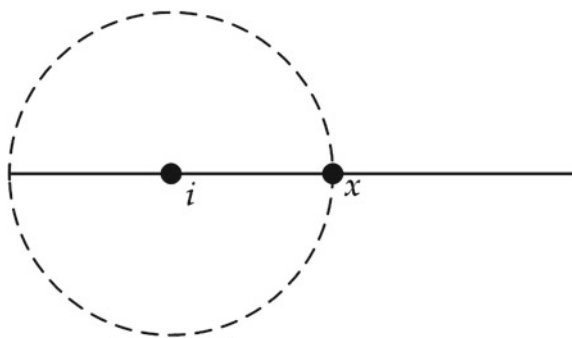
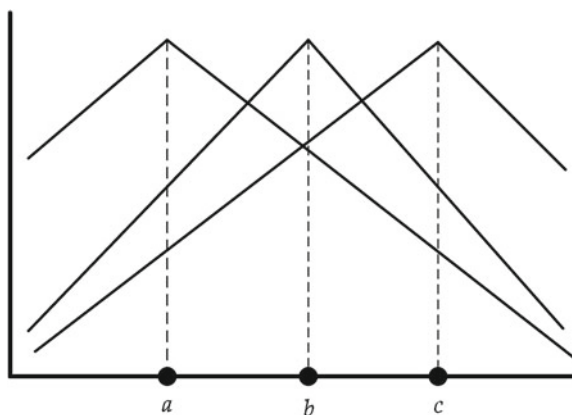


policy space. Point i is the player's ideal point, and is where the player's utility peaks. Notice that as the policy points move away from the player's ideal point i the player's utility decreases, i.e., they are less preferred by the player.

Figure 2.3 is an example of a *single-peaked preference profile*. This term implies that as policy alternatives move away from a player's ideal point, the player's perceived utility decreases monotonically. Traditional Euclidean models assume that all players' preference profiles are single-peaked.

Another way to demonstrate a player's preferences over a policy is by drawing the player's circular indifference curve. Assuming the player's preferences are single-peaked, we know that each incremental move from the player's ideal point is less preferred. If we measure the distance between the player's ideal point and a policy point x , and then draw a circle that captures all points y equidistant from the player's ideal point, then all points y on the circle are indifferent to x ($x R y$ and $y R x$, therefore $x I y$). Any point bound by the circle is preferred to the points on the circle, this is called the winset of x . Any point beyond the bounds of the circle are less preferred than the points on the circle and are referred to as the inverse winset of x . In a single-dimensional model, there are only two policy points on the indifference curve where the circle intersects the dimension (Fig. 2.4). However, as the number of dimensions expands to two or greater, the number of points on the indifference curve becomes infinite.

In the single-dimensional Euclidean model, given that players all have single-peaked preferences and decide by a simple majority voting rule, there is always a nonempty maximal set; i.e., there exists some policy point x such that for all other policy points y , the group holds x to be at least as good as y , or y is not strictly preferred to x , or $x R y$. This is demonstrated by what is called Black's Median Voter Theorem (Black 1958), which states that, given an odd number of players with single-peaked preferences along a single dimension, the group outcome will always be the ideal point of the median voter. This can be seen in Fig. 2.5 below. Figure 2.5 represents three players each with a single-peaked preference profile; such

Fig. 2.4 Circular indifference**Fig. 2.5** Black's median voter theorem

that for Player 1 aR_1bR_1c , for Player 2 bR_2cR_2a , and for Player 3 cR_3bR_3a based on Euclidean distance. Therefore, for either Player 1 or Player 3, Player 2's ideal point b is preferable to the ideal point of the further player. In a tournament, either Player 1 or Player 3 would always agree to Player 2's ideal point b rather than accept their least preferred option.

Once the number of dimensions in the Euclidean model exceeds one, the outcomes of a simple majority game become less certain. In fact, except under very specific conditions,⁵ the set of possible outcomes of a simple majority game can be the entire Euclidean space (Austen-Smith and Banks 1999, p. 180). This is often referred to as McKelvey's Chaos Theorem. This is the result of cycling throughout the Euclidean space, and predicts unstable outcomes as the result of simple majority games in which there exist more than two-issue dimensions.

⁵ Under conditions of perfect symmetry in two dimensional space, called the Plott conditions, the maximal set is nonempty (Austen-Smith and Banks 1999, pp. 142–149).

Besides the failure of traditional Euclidean models to produce stable outcomes in multi-dimensional space, there are a number of assumptions integral to stable outcomes even in a single-dimensional game that seem counterintuitive, if not theoretically and empirically restricting.

The first is the assumption that political actors can locate a precise point in an infinite space that they consider to be ideal and that they can distinguish any shift in policy away from that point to the most minute detail (Potter 2007, pp. 8–9). This is tantamount to assuming that in, say, a \$700 billion budget allocation, a player would notice an increase or decrease of \$100, \$10, \$1, or even \$0.01, and would prefer that policy less than a policy allocating exactly \$700 billion.

The Euclidean model's assumptions extend further as we consider that any player must also hold preferences over the entire infinite policy space (pp. 9–10). For example, take the same political actor who has requested a \$700 billion budget allocation. Not only does the Euclidean model require the actor to hold distinct preferences between \$700 billion and \$700 billion and one cent, but it also requires the same actor to be able to perceive the difference between a budget allocation of \$10 and \$11. While we may expect an actor seeking a \$700 billion budget allocation to consider a policy allocating him \$600 billion or perhaps even \$500 billion, we would not expect him to accept a policy offering him \$100 because his options were that or \$90. We would expect him to either laugh or stomp out of the room furious. However, in the Euclidean model, the actor is expected to have preferences over the entire infinite policy space. In this sense, players in traditional Euclidean models lack a preference horizon, a limit to the set of preferences over which they perceive shifts in utility.

Beyond the requirement that political actors perceive differences in policies with infinite acuity over an infinite set of possible policies extending infinitely far from their ideal point, the Euclidean model also requires that political actors evaluate policies using the same metric or granularity (pp. 10–11). In the Euclidean model, all players are assumed to perceive shifts in policy at the same intensity, with no account for possible differences in the actors' perception of policy space. Returning to our \$700 billion budget allocation, the player requesting the budget allocation, we assume, has a deep understanding of the requirements of his program, and therefore understands the great necessity of the money he has requested. While he may be able to tolerate some minor shifts in the amount allocated to his program, say a few hundred thousand or a few million dollars, major shifts from his budget request would be significantly less preferred by him, perhaps even lie outside his preference horizon. Alternatively, another player in the budget allocation process who only sees the money as tax-payer dollars may not sense a big difference between \$700 billion and \$200 billion. What seems to be marginally as satisfying a policy to the latter player would be disastrous in the eyes of the former. Clearly, these players would view policy shifts in different granularities, but the Euclidean model cannot account for this.

2.4 A Fuzzy Public Choice Model

The application of fuzzy set theory to public choice models allows us to escape many of the pitfalls of traditional Euclidean models. By allowing for indifference in large areas of players' preferences, the fuzzy model avoids the strong assumptions of the traditional Euclidean model of public choice that players perceive shifts in policy with infinite acuity. Rather, the fuzzy model proposes that there are large areas of policy in which a player will experience no change in preference.

In reference to our previous example, Fig. 2.6 depicts the fuzzy preference profile of our player requesting a budget allocation of \$700 billion dollars may actually be happy with anything in a range from \$650 to \$800 billion dollars. Anything less than \$650 billion or more than \$800 billion would then be less satisfactory for him. Let us say that our player's preference for a budget allocation decreases as it moves from \$650 billion to \$400 billion and below \$400 billion our player is completely dissatisfied. Similarly, our player's satisfaction decreases as the budget allocation exceeds \$800 billion and he is completely dissatisfied if the budget allocation exceeds \$950 billion, seeing the allocation as exorbitant and wasteful.

We then refer to the range of \$650 billion to \$800 billion as the core of the fuzzy number representing our player's preferences at which the membership value of a policy in our player's set of preferences equals 1. In Fig. 2.6, the core of the fuzzy preference profile is the flat top of the trapezoidal fuzzy number. Notice that the crisp point of the Euclidean model has been replaced by a flat area of indifference, where the player perceives no difference between the policies. The support of the fuzzy number representing the player's preference profile is then the base of the fuzzy number where the membership value of a policy in the player's preferences is greater than zero. The continuous lines from the core of the trapezoidal fuzzy number to the base of the support are then the monotonically decreasing preferences of the player. Like in crisp, player's preferences decrease incrementally along these continuous slopes.

As can be seen in Fig. 2.6, we can draw lines through the slopes that descend between the core and support of the fuzzy preference profile that we refer to as alpha-cuts. An alpha-cut, or α -cut, selects those elements whose membership values are at least equal to α . A strong α -cut selects only those elements whose membership values are strictly greater than α . The α -cut at one, or $\alpha = 1$, is the core and the strong α -cut at zero is the support.

The core is always a subset of the support. That means that the continuous lines running from the core to the support always decrease monotonically. However, as can be seen in Fig. 2.6, the rate of decrease on either side of the fuzzy number need not occur at the same rate. The fuzzy model does not measure player's preferences by Euclidean distance but by the shape of the fuzzy number representing the player's preferences. The fuzzy model thereby abandons the Euclidean distance assumption, and we no longer need to assume that player's perceive shifts in policy at the same metric. Furthermore, the fuzzy set A is a subset of the set of alternatives X , therefore the support of the fuzzy number is necessarily a subset of the policy space. By

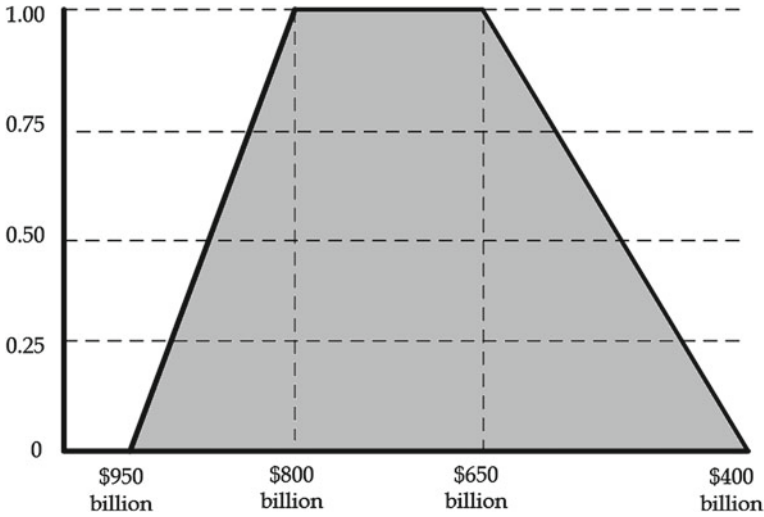
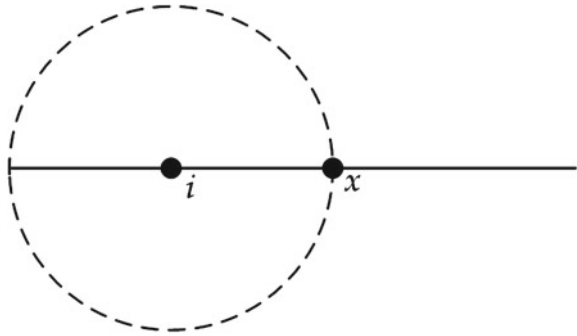


Fig. 2.6 Continuous fuzzy preference profile

Fig. 2.7 Circular indifference



representing player's preferences as a fuzzy number, we define a player's preference horizon; a point beyond which the player derives no utility.

While the fuzzy model alleviates these two problems associated with the Euclidean assumptions of traditional models, the continuous lines decreasing monotonically from the core to the support retains the problem that player's are expected to make infinitely fine judgments between policies. However, we can alleviate this problem by representing players' preferences as discrete fuzzy numbers rather than continuous fuzzy numbers, as depicted in Fig. 2.7.

In Fig. 2.8, the preference profile of our player requesting a budget allocation of \$700 billion is represented as a discrete fuzzy number. The player's preferences are now represented by a set of stacked boxes rather than a smooth curve. Each block

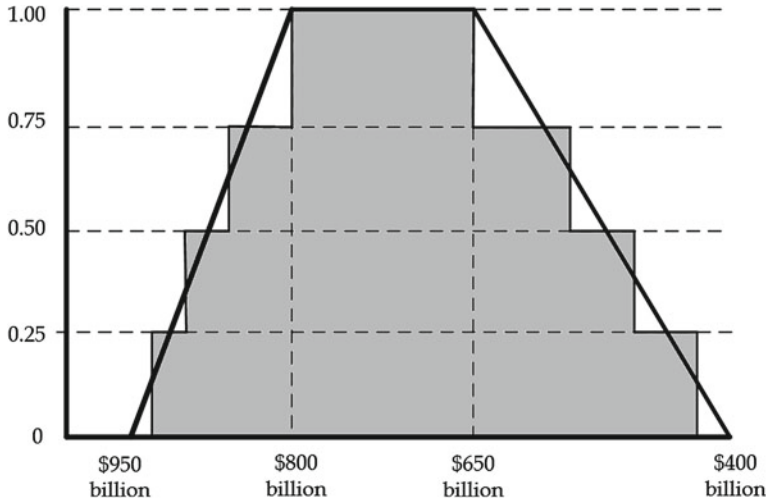


Fig. 2.8 Tiered fuzzy preference profile

represents an α -cut. The difference between representing a player's preferences as a discrete fuzzy number and representing them as a continuous fuzzy number is granularity. A continuous fuzzy number assumes that the player can perceive shifts in policy along the slope descending from the core to the support with infinite acuity. A discrete fuzzy number assumes that the player has areas of indifference at which he cannot determine the difference between policies in terms of his utility. Where a player formerly perceived shifts in utility from 0.84 to 0.8397 to 0.80 to 0.76, he now groups these together and sees them all at a single-level; $\alpha = 0.75$. As can be seen in Fig. 2.8, $\mu_A = 0.75$ if $0.75 \leq \mu_B < 1.00$. So, while our player may perceive the difference between \$625 and \$600 billion conceptually, the difference means nothing to him in terms of utility, as his utility for both is $\alpha = 0.75$.

We can understand the discrete α -cuts of any element's inclusion in the set of a player's preferences. We already know that at $\alpha = 1.00$ we say that a policy is perfectly in the set of a player's preferences, or is ideal. If $\alpha = 0.00$ for a certain policy, the policy is completely out of the player's set of preferences, or completely not preferred by the player. Between these two extremes of ideal preference and complete rejection are areas of ambiguous preference. At $\alpha = 0.75$ we may say that a policy is more in the set of a player's preferences than out, or the player prefers the policy more than does not. At $\alpha = 0.25$ we may say that a policy is more out of the set of a player's preferences than in the set, and therefore the player more dislikes than likes the policy. At $\alpha = 0.5$ a policy is neither in nor out of the set of a player's preferences, and therefore we can say his preferences over these policies is ambiguous.

One possible criticism of the discrete fuzzy preference profile is that, while it assumes players perceive policies with large areas of indifference, it also assumes that there are crisp points at which a player perceives a sudden loss (or gain) in utility. While we concede this problem does exist, we argue for the scholarly utility of being able to account for large areas of indifference in player's preferences. This problem of sudden shifts in utility at the bounds of α -levels can be alleviated by assigning fuzzy boundaries to the discrete levels. This can be achieved through the development of the relevant set theory.

2.5 Advantages of the Fuzzy Public Choice Model

The fuzzy public choice model offers a number of intuitive and practical benefits over the traditional Euclidean model. It allows researchers the opportunity to measure player's preferences without assuming that they each evaluate the entire infinite policy space with the same level of acuity. By loosening this assumption, researchers can actually tailor fuzzy numbers specifically for the preferences of the actors they are studying. Furthermore, they can find areas in which no compromise is possible between actors; where their preference horizons end before they overlap. Most importantly, however, researchers can use the fuzzy model to make stable predictions about group choice without importing further assumptions such as institutional procedures or sophisticated voting.

However, the most important advantage of a fuzzy public choice model is that it predicts stable outcomes more consistently than does the traditional Euclidean model. We will demonstrate the ability of several fuzzy public choice models to predict government formation outcomes using the CMP empirical data. In Chap. 4, we consider two single-dimensional fuzzy public choice models. The first makes use of the fuzzy maximal set, and the second will make use of the fuzzy Pareto set to make predictions. In Chap. 6, we focus on the ability of a fuzzy two-dimensional model to predict the same government formation outcomes.

All of the fuzzy public choice models that we present begin by using fuzzy preferences that conceive of preferences as subsets of policy space over which the actor assigns different utility values. Benoit et al. assume that a party preference profile contains a large amount of error and a potential overlap in preferences does not equal a willingness to agree between parties. Our approach argues that overlaps, or intersections, between actors' preference profiles represent subsets of policy over which actors can agree. Moreover, our fuzzy public choice models can explain seemingly detrimental shifts in policy on one dimension in favor of gains on another dimension. In conventional public choice models, a political actor will only give way to a loss in utility on one dimension if it entails an overall gain in utility in multidimensional space. A fuzzy public choice model, on the contrary, allows actor's to make compromises on one dimension for gains in another without necessarily experiencing a loss in utility on the former. In this way, scholars can predict whether an actor may be more likely to make a compromise for a gain, when the actor perceives no loss

in utility on the compromised dimension, while the actor may be more resistant to such a trade-off if some real loss in utility occurs as a result.

In the next chapter we address issues related to fuzzy single-dimensional public choice models. The discussion will help us to present and test our fuzzy single-dimensional public choice models of government formation in Chap. 4.

References

- Arrow, K.: Social Choice and Individual Values. Wiley, New York (1951)
- Austen-Smith, D., Banks, J.: Positive Political Theory I. The University of Michigan Press, Ann Arbor (1999)
- Benoit, K., Laver, M., Mikhaylov, S.: Treating words as data with error: Estimating uncertainty in the comparative manifesto measures. *Am. J. Polit. Sci.* **53**(2), 49–513 (2009)
- Black, D.: The Theory of Committees and Elections. Cambridge University Press, Cambridge (1958)
- Budge, I., Klingemann, H.D., Volkens, A., Bara, J., Tannenbaum, E. (eds.): Mapping Policy Preferences: Estimates for Parties, Elections, and Governments: 1945–1998. Oxford University Press, Oxford (2001)
- Klingemann, H.D., Volkens, A., Bara, J., Budge, I., McDonald, M.: Mapping Policy Preferences II: Estimates for Parties, Electors, and Governments in Eastern Europe, European Union, and OECD 1990–2003. Oxford University Press, Oxford (2006)
- Laver, M.: How should we estimate the policy positions of political actors. In: Estimating the Policy Position of Political Actors. Routledge, London (2001)
- Laver, M., Budge, I. (eds.): Party Policy and Government Coalitions. MacMillan Press, London (1992)
- Mikhaylov, S., Laver, M., Benoit, K.: Coder reliability and misclassification in the human coding of party manifestos. *Polit. Anal.* **20**(1), 78–91 (2012)
- Poole, K.: Spatial Model of Parliamentary Voting. Cambridge University Press, Cambridge (2005)
- Poole, K., Rosenthal, H.: Ideology and Congress. Transaction Publishers, New Brunswick, NJ (2007)
- Potter, J.: Extracting fuzzy preferences from empirical data (2007), unpublished thesis

Fuzzy Social Choice Models

Explaining the Government Formation Process

C. Casey, P.; B. Gibilisco, M.; A. Goodman, C.; Pook, K.N.;

N. Mordeson, J.; J. Wierman, M.; D. Clark, T.

2014, XIII, 183 p. 29 illus., Hardcover

ISBN: 978-3-319-08247-9