

8/1/2001

Measuring Fuzzy Set Membership Functions: A Dual Scaling Approach

by Jay Verkuilen

jayv@uiuc.edu

Institute of Government and Public Affairs
Department of Political Science
Department of Statistics
University of Illinois Urbana-Champaign

Prepared for Presentation at the Annual Meeting of the APSA, San Francisco, CA,
August 30-September 2, 2001.

The author would like to thank Jim Kuklinski for an enormous amount of aid well above and beyond the call of duty in the process of writing this piece. In addition, discussions with, comments by, and encouragements from Michael Coppedge, Jan de Leeuw, Michael Greenacre, John Marden, Gerardo Munck, Steven Seitz, Michael Smithson, and William Stout—and probably a number of people I have neglected to mention here—were essential. Thanks also to Charles Ragin, whose book has made presenting research on fuzzy sets a lot easier as well as providing a foil for my own work.

Abstract

Charles Ragin's (2000) recent book opened up a recent dialogue on fuzzy set methods in social science data analysis. Membership functions are measures of partial set membership and are normalized to be in the unit interval and have been used to characterize situations of non-probabilistic vagueness. One of the deficiencies of the fuzzy set literature has been a lack of a firm basis in measurement for membership functions, even though leading fuzzy set theorists have noted that any practical applications depend strongly on the quality of membership assessment. In this paper, I discuss the use of dual scaling, a method from nonlinear multivariate analysis that scales multiple categorical items, for the assignment of fuzzy set membership functions. The paper also includes an example of the scaling applied to some data from Russett (1964) that are very typical of the sort encountered in medium N cross-national research.

I. Introduction

When Lotfi Zadeh (195) first proposed them in the mid 1960's, fuzzy sets looked quite promising for social scientists. Although Zadeh hoped that fuzzy sets would provide a mathematics appropriate to social science, both fuzzy set theorists and social scientists interested in fuzzy sets lament the lack of applications of fuzzy set theory to social science. Indeed, according to Smithson (1999), fuzzy set theory has hardly made a dent in social science research circles in over thirty years of existence. Occasionally, scholars see fuzzy sets as an inspiration and there have been sporadic review pieces, for instance, Cioffi-Revilla (1981). A few scholars have made real use of fuzzy set techniques to solve social scientific problems, see for instance Smithson (1987) working in social psychology; Manton et al (1992) working in statistical demography; or Seitz (1994), Sanjian (1987), and Taber (1992), working in international relations.

What sort of promise does fuzzy set theory offer? I cannot make a complete case for fuzzy set theory here, but will sketch a basic argument for why it would be a useful addition to the social scientist's methodological toolbox. For a long time there has been a critical gap between social scientific theorizing on one hand and the methods used to test these theories on the other. Theory is most often expressed in logical or, equivalently, set theoretic terms, but quantitative testing is done using methods that do not deal with the statements as they are, but instead assume other structures. For instance, Barrington Moore's classic theory of democratization, put rather schematically, states that unless a country has a strong bourgeoisie, it will not be democratic. This is essentially a set-theoretic statement, which is to say, if a country is not in the set of countries with strong

Verkuilen, “Measuring Membership”

bourgeoisies, it will not be in the set of democracies.¹ Many scholars would use, without thinking about it much, some kind of correlation or regression to measure how associated democracy and bourgeoisie were. The problem is that correlation is fundamentally symmetric, since $r_{XY} = r_{YX}$. However, Moore’s argument itself contains no such symmetry and importing the assumption that underlie correlation is not innocuous. Causal arguments say “If X then Y,” *not* “X is correlated with Y” but too often the latter is used to test the former.

Fuzzy set theory provides *a* basis for scholars to have a mathematical language in which to express quite sophisticated theories in a precise and testable manner. For instance, fuzzy set theory has been used to make expert system models, which are fairly complex computer systems that model decision making processes by a interlocking system of logical statements. The Taber and Seitz models cited above are both expert systems. It can also provide a number of tools for empirical data analysis. The example at the end of the paper will provide a taste of data analysis based on fuzzy sets. However, it is important to recall that “many behavioral and social scientists have been led down the garden path by persuasive methodologists’ promises of problems solved and data made intelligible” (Smithson, 1987, p. 288). Thus I do not claim that fuzzy set theory will be *the* basis for a ground-shaking revolution in the practice of social scientific theorizing or data analysis or that it will replace everything that we have all learned in the past, only that it can shed some light on corners that have so far been barely illuminated by existing techniques.

As interesting as the work cited above has been, none of it has had a resounding impact. A recent major book by prominent sociologist Charles Ragin (2000) has sought

¹ The equivalence of Boolean logic and set theory is a well-known mathematical fact.

Verkuilen, “Measuring Membership”

to raise the battle standard of fuzzy sets in the social sciences again and provided an opportunity to examine what fuzzy set theory could provide to social scientists. Much is positive about Ragin’s effort at making fuzzy set theory accessible to a broader audience. His book goes a long way to showing the working social scientist that there is indeed a payoff, in relatively simple English, not lots of intimidating equations.²

However, readers of Ragin’s book have noted a crucial deficiency: it is quite unclear how one is to obtain membership functions.³ If membership functions are, as Ragin asserts, the fundamental connection between empirical data on one hand and one’s fuzzy set models on the other, this is a telling weakness. According to George Klir, one of the most prominent fuzzy set theorists, “... its [mathematics based on fuzzy sets] usefulness depends critically on our capability to construct appropriate membership functions for various given concepts in various contexts.” (Klir and Yuan, 1995, p. 15) Ragin seems to say that it is best to let theory be the guide but this only begs the question if, as is normally the case, we want to use membership functions to *test* theory. He offers little guidance on how one would do this rigorously and, more importantly, how independent scholars could critique and/or replicate results of membership functions that are based almost entirely on subjective judgment of a single investigator.

This problem is not one unique to Ragin, but, rather is endemic to the literature (Smithson, 1987). To be sure, fuzzy set theorists have paid some attention to obtaining membership functions (Klir and Yuan, 1995; Bardossy and Duckstein, 1995). However, most of the approaches do not make use of insights from psychometrics, where

² Unfortunately it will not be possible to avoid equations in this piece, though I endeavor to explain them as much as possible in plain English.

³ When this book was assigned in Jim Kuklinski’s graduate methods class at the University of Illinois Urbana-Champaign in Spring 2001, all the students brought this up without any prompting (Kuklinski, personal communication).

specialized methods for creating low dimensional indexes from many observed variables have been developed. As a consequence, most of the most of the more sophisticated existing methodologies have not been brought to bear on the problem of membership assignment. This seems to have worked fairly well in control systems engineering, where the sets are frequently well understood physically and extensive experimental validation is possible, but social science generally—and cross-national research particularly—is not in the same situation. As Zimmerman (1991) notes, there is a fundamental difference between empirical sciences, physics, biology, most social sciences, on one hand and technological disciplines, such as computer science, management science, or control systems engineering. Thus, fuzzy set theory, which primarily developed in the latter disciplines probably needs some adaptation to be fruitful in the empirical sciences.

In this paper, I propose the use of dual scaling to generate membership functions from multiple categorical indicator variables. Dual scaling is a technique with a long history in multivariate analysis and is related to principal component analysis, canonical correlation, and other “eigenvalue” techniques.⁴ It has proven to be useful for analyzing many categorical variables with a minimum of distributional assumptions. Since most of the data that social scientists encounter *are* categorical—or could be thought of as categorical—this focus is warranted.

There are a total of six sections that follow the introduction. In section II, I will discuss the basic properties of the membership function, giving a brief overview of what

⁴ I adopt the term *dual scaling* for what has been variously called optimal scaling, optimal scoring, canonical analysis of indicator matrices, Hayashi’s quantification method, multiple correspondence analysis, homogeneity analysis, factorial analysis of the Burt table, seriation by indicator matrices, etc. The term dual scaling was coined by Shizuhiko Nishisato, and his book has a review of the history of the technique (Nishisato, 1994). The main advantages of the term are: first and foremost, unlike its polysyllabic competition, dual scaling is compact, simple, and easy to say, and second, it is descriptive, since the mathematics employed depend on a duality relationship between the rows and columns of the original data matrix. I do not mean to imply an association with a particular school of thought.

Verkuilen, “Measuring Membership”

it represents, focusing on the properties relevant to the measurement problem. In section III, I will discuss measurement models in general, providing a brief overview of the concept as well as a review of the existing literature on measurement as it applies to fuzzy sets. I will concentrate particularly on the kinds of problems a social scientist would be likely to face when making use of fuzzy set theory and argue that dual scaling is an appropriate approach to handle the kind of data that is common in most problems of cross-national data analysis. In section IV, I will discuss dual scaling and derive the basic equations. In section V, I will discuss issues of solution quality, focusing on three issues: multidimensionality, solution stability, and missing or ambiguous data. Section VI I will provide an example using some cross-national data from a *World Politics* article Russett (1964). Finally, section VII is the conclusion, where I will discuss some future directions for the scaling of membership functions.

II. Properties of the Membership Function

Here I will review the basic properties of the membership function, particularly those related to the measurement problem. There has been a great deal of confusion about these points and I will attempt to lay out what I believe is the defensible position on the measurement theoretic properties of membership functions. I will not, however, discuss the interpretation of the membership function in great detail except insofar as it is necessary to understand the current problem. Instead I refer readers to sources such as Cioffi-Revilla (1981) or Ragin (2000) for gentle reviews aimed at social scientists; Smithson (1987) for a quite thorough treatment from a social scientist’s perspective; and

Verkuilen, “Measuring Membership”

Klir and Yuan (1995), Bardossy and Duckstein (1995), or Zimmerman (1985) for more traditional engineering and mathematics approaches.

Formally, the membership function μ_x is a function over some domain, or property space, \underline{X} , mapping to the unit interval, $[0,1]$. What does the membership function measure? It is an index of “sethood” that measures the degree to which object A with property x is a member of a particular defined set. The usual definition of a classical set uses properties to determine strict membership or nonmembership. The main difference between classical set theory—which is a boundary case—and fuzzy set theory is that the latter admits to partial set membership.

This makes fuzzy set theory very useful for modeling situations of vagueness, that is, nonprobabilistic uncertainty. For instance, there is a fundamental ambiguity about the term “power.” It is difficult to put many nations unambiguously into or out of the set of powerful nations. Few people would argue whether the United States is powerful or that Burkina Faso is not. However, is Serbia? What about Brazil? These cases are much more difficult to classify (and of course depend heavily on one’s meaning for “power,” which will depend on application). It is not a matter of whether either country has a probability of being in a well-defined set, it is that the set itself does not have firm boundaries.

Gender is another example, which may seem somewhat controversial. For the social scientist (but perhaps not for the physiologist), biological gender is clearly divisible into two classes, male and female. The set of complement set of males is females and vice versa. However, it is not clear that in a social sense gender can be partitioned so easily. If one considers a hypothetical continuum of behavior ranging from stereotypically male

Verkuilen, “Measuring Membership”

to stereotypically female there is clearly going to be substantial within-group variation in both groups, even if there is separation between the centers of both distributions. The same criticism holds for many other social scientific concepts, such as ideology, democracy, racism, etc., and the structure of the 2000 Census, in which respondents could choose more than one race, makes the problem of how to categorize respondent race a relatively urgent one.

In my view, it would be largely pointless to construct a “one size fits all” fuzzy set of powerful nations that would be of much use in empirical research and the same holds true for most of the other concepts mentioned above. Measures *must* be embedded in a research program to be meaningful at all and they obtain what validity they have *from* this research program. Fuzzy sets provide a means to quantify the vagueness inherent in terms and fuzzy set theory provides a mathematical language for the manipulation of these terms. There is no mystery beyond this.

\underline{X} is usually—but not always—taken to be a subset of real space. Actually, lying behind this space is a measurement model which identifies observables with real numbers, analogous to the formal definition of a random variable. The domain \underline{X} may have numerical properties implied by the field properties of real numbers, e.g., ordering, or it may not. The Stevensian classification of measurement levels identified different levels of measurement by the class of transformations that could be applied to the domain without affecting the information contained. “Higher” levels of measurement, for instance the ratio scale, were identified with more restrictive groups of transformations (scalar multiplication), while “lower” levels of measurement such as the ordinal scale admitted broader groups of transformation (monotonic) without any information loss.

What a researcher is willing to assume about the measurements in question depends on the structure of the domain, which is a matter for substantive knowledge and empirical investigation (Jacoby, 1999). For instance, many classification schemes, that is nominal scales, are coded using numbers, but no one seriously believes that the numbers *mean* anything beyond an arbitrary categorization. That is, no one believes that the numbers assigned to the categories actually have the additional properties of numbers such as ordering or addition. We may as well relabel the categories any other way. So long as we have the same number of categories, nothing is lost. Thus, in perhaps more familiar psychometric terms, a membership function is a one dimensional latent variable that takes some vector of properties—possibly ordered, possibly numerical, possibly not—into the unit interval. Larger values indicate stronger membership in a latent, i.e., unobservable, abstract, set, while smaller values indicate weaker membership in this set. Interpretation depends, fundamentally, on one’s purposes (Collier and Adcock, 1999). Ragin’s admonition of researchers to consider carefully what they mean when they index something is quite relevant (Ragin, 2000).

Unfortunately, there has been substantial confusion about the measurement level of the membership function in the literature. Levels as strong as absolute to as weak as ordinal have been claimed by different scholars at different times. Sometimes the same scholars have even claimed one thing at one time and changed their minds later on. Ragin, for one, claims that the membership function is a ratio scale with two endpoints (Ragin, 2000, p. 154). However, Norwich and Turksen (1984, pp. 2-5) proved that for a continuous, order-dense domain, that is one in which there is—at least in principle—

Verkuilen, “Measuring Membership”

always an object possible between any two given objects and thus no “gaps” in the domain, the membership function is in fact *no stronger* than interval.

The simplest way to understand this is to note that a membership function, unlike a probability measure, does not fulfill the requirement of a meaningful concatenation operator that underlies the ratio scale (Roberts, 1979). It is meaningful to add the probability of the union of two mutually exclusive events A and B: $P(A) + P(B) = P(A \text{ and } B)$ because a probability measure is a ratio scale. It is not, however, meaningful to add the membership values of two objects in a fuzzy set. The sum $\mu_A + \mu_B$ may be arithmetically *possible*, but it is certainly not *interpretable* in terms of fuzzy sets and there does not seem to be another concatenation operator that is meaningful in general (Norwich and Turksen, 1984, p. 7). For instance, if one were to add together two masses, it makes sense to say that the mass of the combined system is the sum of the two masses. However, if one were to take two objects that are elements of fuzzy sets and “put them together” there is no natural way to combine them, unlike masses or probability. Turksen (1991) has additional discussion on these points, and extends the results into conjoint measurement over a multidimensional Cartesian product space, of which dual scaling is a variety.

However, I do not want to belabor the axiomatics and instead prefer a more pragmatic view. A number of important scholars, Louis Guttman for one, have attacked axiomatic measurement theory on the grounds that it is fundamentally about “giving permission” to do analysis which he believes methodologists should never do. “There is a widespread folklore concerning mythical statistical ‘rules’ that forbid or permit calculations involving ‘scales,’ these ‘rules’ being independent of context.” (Guttman, 1977). Abraham Kaplan

Verkuilen, “Measuring Membership”

notes, “The point of view which I have been presenting with regard to the validity of measurement comes to this, that you can do what you like with the numbers assigned, provide that you know what you are doing and are prepared to accept the consequences” (Kaplan, 1964, p. 205). The Gifi group—intellectual descendants of Guttman in many ways—consider levels of measurement to be testable empirical hypotheses, not a priori statements of meaningfulness or meaninglessness, as the axiomatic measurement theory school views them (Gifi, 1990, Chapter 1). Furthermore, the measurement level of a variable is not a “once and forever” thing. No variable is *inherently* at *any* level of measurement. Instead variables have levels that can only be established in comparison to other variables of interest.

Nevertheless, however one considers the issue, it is useful to establish relatively strong metric information for a membership function. For a number of reasons, the most important being the need to employ power transformations, set complements, and certain union and intersection operators, all of which depend on the numerical structure of the membership function, is desirable that a membership function be *no weaker* than the interval level (Smithson, 1987, Chapter 4). Thus, my discussion below will concentrate on providing an interval-level scale, given some data. The fact that the generated scale is interval is important for resolving some identification issues in the scale equations as well as addressing an important substantive dilemma about the meaning of the membership function.

III. Measurement Models

Social scientists are rarely interested in directly observable quantities. The concepts used for theorizing are, by and large, set at a fairly high level of abstraction and are related to other abstract concepts. The observable world only impinges on these concepts indirectly. However, social scientists *are* scientists, not philosophers, and so it is essential that they make a connection to empirical evidence. While statistics concerns itself with the relationship between data and inference, measurement theory considers the relationship between observation and data. The two are complementary and measurement is a precondition for statistical analysis, though good measurement is too often ignored. There are two issues I will discuss in this section. First I will discuss some desirable criteria that a method for scaling membership in a social science context should possess, namely that it use multiple indicators, that it be sensitive to the fact that most social scientific data are categorical, and that it not assume too much about the interrelationships of the indicators. Second, I will discuss some of the existing methods used to scale membership functions.

Some Criteria for Scaling Membership

There are always many different ways to measure our concepts in use. It is particularly desirable, because we do not believe that our observations are either noise-free or bias-free, to use multiple measures, combining their values through a measurement model (Bollen, 1989; McDonald, 1999). By combining more than one measure into an index, we “borrow strength” from all the separate measurements. Easily the most familiar example of this is the construction of a composite grade from many different items on a test or the construction of a final grade from several tests. We

Verkuilen, “Measuring Membership”

recognize that any one test item is too error-laden to be the basis of an accurate or nuanced assessment of a student’s performance; the solution is to use many indicators, hopefully ones that err in different ways, which leaves the composite stronger than the individual parts.

Furthermore, most social science data is *categorical*, even though many analyses do not note this fact (Gifi, 1990; Gleditsch and Ward, 1997; Jacoby, 1999). Survey questions, for instance, are frequently analyzed using linear regression but a few points on a Likert scale are clearly categorical and importantly, bounded at two ends. I do not mean to imply that one should *never* analyze survey data this way, but it is important to be aware of this fact. Thus a scaling method should be able to accommodate categorical data.

Finally, it is important that the scaling method does not assume too much about the interrelationship of the indicators. In social science we typically have “bad” data in the sense that it is not metric but, at best, ordinal, or at least—with relatively few exceptions—we have little reason to assume our data is better than ordinal. Even something like per capita GNP, which seems on the surface obviously to be ratio, may be, in a particular research context, at most ordinal or perhaps even nominal (Lieberman, 2000). Fairly often in practice one uses the logarithm, i.e., an order-of-magnitude scale, rather than per capita GNP itself, as this sort of comparison is considered more meaningful substantively in addition to being desirable statistically by deskewing the observations and so stabilizing the variance. The relationship between per capita GNP and some variables of interest might well be nonmonotonic, as, for instance, some researchers in democratization have claimed (Diamond, 1992). Thus we need a method

Verkuilen, “Measuring Membership”

that works on nominal or ordinal data and cannot in general assume that our input data will be “nice.”

Review of Existing Methods

Smithson (1987, Chapter 3) has an excellent review. Klir & Yuan (1995) have a chapter on the topic. Bardossy and Duckstein (1995) devote but a chapter section, as do Dubois and Prade (1986). Most fuzzy set theorists have been interested in what can be done with already existing membership functions and not how one would get membership functions in the first place. Thus the topic has suffered for a certain lack of attention. Dubois and Prade make a brief discussion but largely dismiss the topic on the grounds that most of the mathematical procedures of fuzzy set theory are robust to errors, a move not unlike that made by many social scientists with regards to the sometimes problematic assumptions of linear regression and similar procedures.

By far the most common method for assigning membership is based on direct, subjective judgments by one or more experts. This is the method recommended by Ragin. In this method, an expert rates objects on a membership scale, assigning membership values directly and with no intervening transformations. One or more than one expert may be used for this task. For instance, Norwich and Turksen (1983) use direct rating of several judges to scale “tall man.” For relatively conceptually simple sets such as “tall man,” this method often does the job quite well and so it should not be neglected as a means of obtaining membership values. In a very important way all other methods require fundamental understanding of pragmatic/theoretical concerns. If nothing else, only previously existing theory and investigator insight can choose appropriate

indicators and investigator intuition will be essential to ensure that derived ratings make sense.

However, this methods has many shortcomings. First, experts are often better at simpler tasks, e.g., paired comparison or generating ratings on several more concrete indicators, than they are providing values for one membership function “fully formed from Zeus’ brow,” as it were, for a relatively complex set. Asking them to rate objects on membership in a highly abstract set directly may be too difficult and thus resources would be better spent having expert judges rate several more concrete scales which are subsequently combined into one index by an aggregation procedure (Klir and Yuan, 1995). Saaty (1986) noted this point when he proposed using eigenvalue decompositions of matrices of paired comparison data to scale membership functions, which is not at all unlike the procedure I propose. Second, investigator bias—even unconscious bias—or inconsistency can creep in when the ratings need to be made for the conceptually complicated sets which abound in social science. Freedom House’s ratings have been criticized on both grounds, for instance (Bollen, 1993). It is much harder to defend a membership rating that comes solely from expert judgment when there is little to point to backing up the procedure besides the expert’s status as an expert. It is better to have a very clear procedure in place that is as transparent as possible and well documented.

Another approach uses parametric functions and optimizes a loss function to find estimates given some data. Parametric functions are ones that have some, usually small, say one to five, unknown parameters. The function itself is chosen because it has desirable properties, e.g., it is monotonic, has a sigmoid or bell shape, is analytically tractable, etc. The task of estimation is to provide good values for the parameters based

on data. On the positive side, statisticians have developed numerous techniques for parametric estimation, including the very familiar least squares and maximum likelihood methods, which estimate the parameters by optimizing a loss function, either analytically or numerically. In the case of estimating values for membership functions, the function itself would be chosen because it meets the bounding conditions, say the two parameter logistic function $f(x) = 1/(1 + a e^{-bx})$, for $x > 0$, thus ensuring that all assigned membership values would lie in the unit interval, and that other the axioms of fuzzy set theory were met because the function satisfies these properties. In addition, the function chosen should be shaped properly to represent the idea in mind. If, for instance, the set in mind is represents a concept subject to diminishing returns, something that allows for rapid change early on but slow convergence to an asymptote later on would make a lot of sense.

Parametric approaches are also nice because they give one parameters, which often have useful substantive interpretations. However, they tend to be relatively inflexible and have relatively strong requirements regarding the input data’s measurement level, usually needing interval or ratio data. Often there is little solid justification for the choice of one function versus another: There are many functions that have sigmoid shape, why did you pick the logistic? Most theories in social science do not say enough to specify functional form with any real precision (Beck and Jackman, 1998), so the functional form we choose is just a convenience, but in many cases it might well not be *all that* convenient. Model artifacts can be a real problem, hiding as much as they reveal.

I consider a nonparametric approach more fruitful, though it is important to specify what one means by nonparametric, as there are numerous methods that fall under this

Verkuilen, “Measuring Membership”

basic name, ranging from the “sort-of” parametric approaches like dual scaling (perhaps better called “many” parametric) to the more truly nonparametric approaches like kernel density estimation. This approach starts with some basic considerations, for instance, about what shape the computed function should have, and allows the computer to find a representation that fits well according to some given loss. For instance, we might want a monotonically increasing function and want to use least squares, and so we use a generalized additive model or loess (a local linear smoother) constrained to monotonic and in the unit interval.

One of the more interesting results of the “nonmetric revolution” has been the finding that many ordinal constraints often are enough to specify metric information to a high degree of accuracy, *presuming that one is willing to make some assumptions*. For instance, ordinal multidimensional scaling often nearly reproduces the configuration of points metric multidimensional scaling produces when the true distances—ratio level measurements—are known (Borg and Groenen, 1997). Similar results apply to the marginal frequencies of crosstabulations and the bivariate marginals, which can be used given a loss function, to “metrize” nominal data. Thus it is possible to define techniques that, at the cost of fairly commonly made, though not uniformly innocuous auxiliary assumptions, generate a solution very close to metric. Dual scaling is one such approach.

IV. Dual Scaling

As mentioned previously, I propose to use dual scaling to generate membership values from multiple categorical indicators.⁵ Dual scaling has a long history, having been

⁵ I should note that there are a number of other methods that might prove fruitful to examine. Polytomous Rasch models are one possibility. There are a number of Bayesian methods based on Markov Chain Monte

Verkuilen, “Measuring Membership”

reinvented under different names in slightly different flavors by investigators in different disciplines over the last several decades. There are excellent historical reviews in Nishisato (1994; Chapter 1), Greenacre (1984; Chapter 4), and Gifi (1990), although each approaches the material from a slightly different perspective than the others. There have not been many political science applications, but the algorithm used is similar to that discussed by Poole (1998) and Jacoby (1999). Broadly, dual scaling scales objects (cases, subjects, respondents, countries, etc.) over multiple variables by finding a system of weights for both, relying on a dual relationship between scaling objects and variables. The procedure uses spectral decomposition of data matrices very similar to those used in principal components, factor analysis, or canonical correlation. In fact, one way of looking at the technique is as an n -set canonical correlation applied to indicator matrices. In this section, I will first try to provide some basic intuition about the procedure I propose. Then I will derive the basic dual scaling equations. Finally I will discuss the adjustments necessary to meet the requirements of a membership function.

Intuition

Hill (1974) presents a solid discussion of the intuition behind the basic procedure, although in an ecological context.⁶ It is worth summarizing the discussion so the equations that will be given below will make sense. Social scientists should easily recognize how similar this problem is to ones they regularly face, for instance, establishing an ideological continuum from several survey items. In empirical studies in

Carlo estimation that could be very useful as well. Johnson and Albert (1999) have a useful discussion of these models as well as citations. This is an important area for future research. (Thanks to Michael Smithson for reminding me to make this point.)

⁶ <http://www.okstate.edu/artsci/botany/ordinate/> is an excellent Internet reference to ordination in ecology.

Verkuilen, “Measuring Membership”

ecology, it is often necessary to establish a gradient, for instance, from a wet region to a dry one or from high altitude to low. In each habitat, different sets of species are observed; looking across habitats, some species may be common to different habitats.⁷

For instance, for three species A, B, and C, and four habitats, 1, 2, 3, and 4, we may observe only species A in habitat 1, A and B in habitat 2, B and C in habitat 3, and only C in habitat 4. See Figure 1. This pattern would be indicative of a clear one-dimensional gradient from habitat 1 to habitat 4, though interpreting what this gradient represents is a matter for substantive theory. A species that lives in all the habitats in question obviously tells the investigator nothing about a possible gradient and should be removed from analysis.

[Figure 1 about here]

In real life, where we measure many more species and many more sites such a neat pattern rarely emerges. Instead, it is necessary to quantify departure from the ideal model by using a loss function. The original approach to this problem was to assign each species a preference weight. For instance, a wet-preferring species might get a weight of 10, a wet-neutral species a weight of 5, and a wet-adverse species a weight of 0. By observing to see what species live at each sampling point, the ecologist could form an average preference score at the sampling point. (Note that the unit of analysis here is the sampling point, to which we want to assign a score.) If the average scores sorted into an order, particularly one that could be validated on other criteria, the presence of a gradient would be established. On the other hand, given a knowledge of a previously established

⁷ Archeologists face the same task given a set of artifacts and strata in a dig. The earliest nonmathematical specification of the problem came in 1877 from Sir William Flinders Petrie, the founder of scientific archeology, when he invented stratigraphic excavation and subsequently the seriation technique based on permutations of incidence matrices which forms the foundation of the entire literature.

Verkuilen, “Measuring Membership”

gradient, it would be possible to derive the preference weights for each species by reversing the process, turning the unit of analysis into the species and finding the species’ average score at each site with a known level of the gradient.

The problem, of course, is that these weights are fundamentally arbitrary, as would be any gradient scores. By making the preference weight large or small enough, it is possible to bias the scores substantially and the same would be true for preference weights derived from subjectively generated scores. Dual scaling derives the weights and scores *simultaneously*, using the observed relationships in the data to form optimal weights and scores and using an explicit loss function—least squares—to quantify departures from the ideal model. The scores that are generated will, if the objects are sorted on this score, “petrify” the data matrix into one that is as diagonal as possible, which establishes the gradient. This makes it superior to the subjective weight technique described above, which requires that the investigator assign points based on intuitions which should be the subject of empirical testing, not assertion.

Derivation

This derivation is based on the works on dual scaling cited above and uses the properties of the singular value decomposition, or SVD. There is a derivation based on Lagrange multipliers in some sources, for instance, Nishisato (1994; Chapter 6), that is quite informative. I take the basic task to be the assignment of membership values to N objects with Q attributes.⁸ Let \mathbf{X} be an $N \times Q$ matrix, with each row vector \underline{x}_i containing

⁸ Though it is an interesting topic, I will not discuss three- or higher way data, such as multivariate time series, here.

the i -th object’s attribute profile across the attributes. An optimal approximation to \mathbf{X} , in the least squares sense, comes from taking the singular value decomposition of \mathbf{X} :

$$\mathbf{X} = \mathbf{U}\mathbf{\Gamma}\mathbf{V}'$$

Here $\mathbf{\Gamma}$ is a diagonal matrix of the singular values, $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_K \geq 0$, where K is the rank of \mathbf{X} ($\min\{N, Q\} \geq k$), and \mathbf{U} and \mathbf{V} are orthogonal matrices of the left and right singular vectors with dimension $N \times K$ and $K \times Q$, respectively. We take them to be ordered with respect to the singular values in $\mathbf{\Gamma}$, so the first column of \mathbf{U} and \mathbf{V} corresponds to the first singular value γ_1 , the second corresponds to the second singular value γ_2 , and so forth. To approximate \mathbf{X} by a matrix of rank $J < K$, use the singular vectors associated with the first J singular values (van de Geer, 1986).

However, in this application it is desirable, for two main reasons to be discussed to use the SVD on a function of a somewhat different matrix, which will be constructed as we go. We will first need a matrix called a super-indicator matrix, which is a matrix of dummy-coded values for the items in question. Let $n = 1, \dots, N$ index the objects, $i = 1, \dots, Q$ index the number of items, and $j(i) = 1, \dots, J(i)$ index the number of options per item i . For typical social science examples, Q will be no larger than ten or so and usually fewer, $J(i)$ will be no larger than six or so,⁹ and N should, of course, be as large as possible, though the author has observed reasonable results with relatively small sample sizes. Within each item, the data matrix is coded 1 if the object in question is positive on that item category, 0 otherwise; each object will have at most one 1 per item.¹⁰ A direct

⁹ Too many categories within a variable leads to problems with the solution. See the selection on Solution Stability for a discussion.

¹⁰ It is possible to relax this requirement, instead requiring that the sum within each item must be 1. This is somewhat confusingly called “fuzzy coding”—*codage flou en français*—and is an important extension of the basic technique, particular for the handling of missing values. The computation in this case is the same.

consequence of this is that all rows in this matrix will sum to Q , the number of items.

Table 1 shows a sample with $N = 3$, $Q = 3$, and $J(i) = 3, 3, \text{ and } 2$.

[Table 1 about here]

Let \mathbf{Z} be a super-indicator matrix constructed as above and let $\mathbf{B} = \mathbf{Z}'\mathbf{Z}$, the cross-product matrix, also known as the Burt matrix. By construction of the super-indicator matrix \mathbf{Z} , the Burt matrix has an important block structure defined by the items. In each diagonal block, the Burt matrix is strictly diagonal with the diagonal entries indicating the frequencies of each item option. In off-diagonal blocks, the Burt matrix has the two-way tables crossing items. That is, the Burt matrix crosses all items with all other items. For Q items, there are ${}_QC_2 = Q(Q-1)/2$ pairs of tables including the frequency tables, which are of course really just the items crossed with themselves. The Burt matrix contains all the information that the decomposition will use¹¹ and it plays the same role in this technique as the covariance matrix does in principal components analysis, factor analysis, or regression. Table 2 shows a sample Burt matrix for data from Healy and Goldstein (1976) taken from a survey of 12,232 mothers of young children. I have partitioned the Burt matrix below into the appropriate blocks to make the structure more apparent, so the upper leftmost block shows the crossing of A with A , i.e., the frequencies of A , while the upper rightmost block shows the crossing of A with C , etc.

I will discuss using *codage flou*, which I prefer to call probabilistic coding, with regards to missing data and aggregating multiple coders' ratings below.

¹¹ This is an important limitation to the analysis, as only second-order information is considered even though the \mathbf{Z} matrix contains all the higher order information. As de Leeuw (1984) notes, the analysis is really “joint bivariate,” not truly multivariate. I should point out, however, this is precisely the same tradeoff made by ordinary regression, principal components, and numerous other commonly statistical techniques. All ignore higher order interactions unless they are specifically singled out by the investigator. To do otherwise requires sample sizes that are simply impossible if one is to estimate higher order terms with any degree of reliability. This may be less of a cost than one might suppose. According to McDonald (1981), though it is possible to make examples in which variables are pair-wise but not jointly independent, in practice, higher order interactions are relatively rare when the weaker, second-order conditions are found not to hold.

[Table 2 about here]

We will decompose \mathbf{B} , or rather, a transformation of \mathbf{B} , to get category weights. Once we have the category weights it is a simple thing to score objects as the sum of the relevant category weights. Let \mathbf{D} be $\text{diag}(\mathbf{B})$, that is, the matrix having the diagonal elements of \mathbf{B} , $\mathbf{1}$ be the unit vector of appropriate dimension, and n the table sum of \mathbf{Z} (which is also equal to $\mathbf{D}\mathbf{1}/Q$), the super-indicator matrix. Then:

$$\mathbf{B}^* = \mathbf{D}^{-1/2} (\mathbf{B} - \mathbf{D}\mathbf{1}\mathbf{1}'\mathbf{D}/n) \mathbf{D}^{-1/2}$$

is the centered and standardized Burt matrix, which has dimension $J \times J$, where J is the sum of the number of categories across all the variables. This is the matrix we will decompose by the SVD, or, in this case, the more usual eigenvalue decomposition, which are equivalent since \mathbf{B}^* is symmetric. There are two main reasons to analyze \mathbf{B}^* instead of \mathbf{B} (or \mathbf{Z}) directly. The first is that it naturally incorporates an important identification constraint that avoids a trivial solution in which all category weights are set equal to each other, which is clearly uninformative. The second is that it can be viewed as the standardized residual matrix of the independence model from the chi-square table, which provides useful geometric insight into the meaning of the solution as well as a means for conducting statistical tests (presuming random sampling holds).

Taking the eigenvalue decomposition, we decompose \mathbf{B}^* as $\mathbf{B}^* = \mathbf{U}\mathbf{\Lambda}\mathbf{U}$. Here \mathbf{U} is the matrix of eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. Because \mathbf{B}^* is positive semi-definite, the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_J$ will all be real and nonnegative; there will be J of them. By convention, the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_J$ are ordered from greatest to least; the eigenvector \mathbf{u}_1 that has the most variation is associated with the first eigenvalue λ_1 , the dimension with the most variance orthogonal to the first dimension, \mathbf{u}_2 ,

is associated with the second eigenvalue, and so forth, λ_2 . By convention, eigenvectors are of unit length, that is, $\|\underline{\mathbf{u}}_j\| = 1$ and by construction they are orthogonal, that is, $\mathbf{u}_i' \underline{\mathbf{u}}_j = 0$. Each eigenvector represents the linear combination of the columns of \mathbf{Z} that are associated with the maximal variation represented by the associated eigenvalue, orthogonal to all previous combinations. The category score vector in the k -th dimension $\underline{\mathbf{c}}_k$, which is $J \times 1$ because there are a total of J categories, is:

$$\underline{\mathbf{c}}_k = \lambda_k \mathbf{D}^{1/2} \mathbf{u}_k$$

This equation adjusts the eigenvector by the frequency with which a particular category appears; the eigenvalue is a constant of proportionality.

By construction, the category scores within a variable have zero mean and unit variance. All of the books cited above have useful guidelines on how to interpret these scores but it is worth mentioning here that they have mean zero and unit variance; the scaling scales the categories, not the variables themselves. Categories within a variable that are far apart are relatively dissimilar and vice versa. Positive values indicate that the presence of the category contributes to the object score on the dimension while negative values indicate the presence of the category reduces the object score. The fact that categories are scaled is a big advantage of this procedure over ones that assume linearity in items (such as linear principal components) because it is not infrequent that an item does not in fact behave in a linear (or monotonic) manner, even if one believes it should. The decomposition itself *is* linear but linear in the categories. Now for the object scores. Let $\underline{\mathbf{r}}_k$, $N \times 1$ be vector of object scores in dimension k . We can compute this matrix by taking $\underline{\mathbf{r}}_k = \mathbf{Z} \underline{\mathbf{c}}_k$. That is, the score of the i -th object in a dimension is just its row profile multiplied by the column scores.

Rescaling to the Unit Interval

However, the solution given does not meet the criterion of a membership function as it is not normalized to unit interval, but instead is on the real line. Therefore, it is necessary to rescale. The simplest method is to rescale all the object scores by setting the lowest to 0 and the greatest to 1. If \mathbf{r}_1 is the vector of raw object scores in the first dimension and $\underline{\mu}$ is the rescaled vector, the following equation gives the membership value for the i -th object:

$$\mu_n = \frac{v_i - \min(\underline{v}_1)}{\max(\underline{v}_1) - \min(\underline{v}_1)}$$

For a more sophisticated (which is not necessarily to say better) method, one can take the approach of fixing the endpoints of each item, so the minimal category in each item chosen on substantive grounds is set to 0 and the response profile with maximal category within each item chose is set to have a value of 1, rescaling all intermediate values (Healy and Goldstein, 1976). (This calculation will be illustrated in the example.) After this rescaling, the solution is no longer optimal in the least squares sense (because the transformation is not linear), though it is much more interpretable. The Healy and Goldstein rescaling is quite similar to the identification constraints used by Manton et al’s approach to estimation of membership by MLE, in which the user selects anchoring objects that are definitely in or out of the set which are identified with perfect membership or nonmembership, with the estimation procedure positioning intermediate objects in the unit interval. This is the normalization I will adopt, though others are, of course, possible depending on particular substantive purposes in mind. In my view, it provides a balance of compensatory combination of items through an additive scale that

Verkuilen, “Measuring Membership”

is close to the least squares optimal values from the SVD with the additional necessary constraint of the score being in the unit interval.

This, however, raises an important dilemma, and not one that, in my view, is purely a statistical one, though the observed data can tell us something about the problem in conjunction with substantive knowledge. Namely, is there an object with full membership or nonmembership? Norwich and Turksen (1983) suggest that subnormality, that is fuzzy sets that do not have objects attaining full membership (or nonmembership), is in fact quite common empirically. It is thus essential to think carefully about what the membership function means in terms of substantive theory, as Ragin (2000) stresses. The simplest rescaling procedure discussed above implicitly assumes that at least one object attains full membership and at least one object attains full nonmembership in the set. This may not be reasonable, for instance, if a case with maximal membership is considered a nonexistent prototype. For instance, Robert Dahl (1971) developed the more restrictive concept *polyarchy* because he viewed democracy as being one that is not met by any empirical cases. As Smithson (1987, Chapter 2) notes, there are a number of philosophical difficulties about fuzzy set theory related to things like prototype theory, “folk” categorization, etc. Most of these difficulties have not been resolved and it seems that fuzzy set theory does not provide a natural model for human cognition in general.¹²

Can the data itself help resolve this issue? In some ways, the Healy-Goldstein rescaling does so, since it assigns the value of 1 only to objects possessing the purely maximal profile, i.e., all values in the profile are the maximal values in each item, and 0

¹² My aim is much less ambitious. The pragmatic question “Is it useful?” trumps the question “Is it a true representation of reality?” No model is. Models are never true, only useful.

only to objects possessing the purely minimal profile, i.e., all values in the profile are the minimal values in each item. If no observed objects actually possess either pure profile, subnormality emerges quite naturally, as the object score is the sum of the item weights. With the simpler rescaling, the investigator simply picks the desired endpoints for the objects, say 0.1 and 0.9, and scales accordingly, though this procedure does beg the question of what these endpoint values should be. This issue represents a potentially important locus of substantive concern, one that requires careful consideration by the investigator.

V. Assessing the Quality of Solutions

It is, of course, essential to examine the quality of any solution computed. The mathematical theorems underlying the method guarantee that *a* solution will be found, but this solution may not be at all good. There are three main issues to discuss: (1) multidimensionality, (2) solution stability, and (3) data that is missing or ambiguous for some reason. I will discuss each in turn, offering some advice on how to address each issue.

Multidimensionality

First, the eigenvalue decomposition generates many solutions, but for the purposes of scaling a membership function, we are only interested in one. It is well known from the use of factor analysis and other multiple indicator measurement models that multidimensional solutions are both common and can be problematic. Dual scaling is no different in this regard. If one hoped to find a one dimensional gradient but instead finds

that three dimensions are necessary to capture the variation in a set of indicators, this is an indication that one membership function will not summarize the distinctions found in the data. It may be that three different fuzzy sets are required. The primary cure for multidimensionality, of course, is for the investigator to exercise careful judgment in selecting indicators. However, *too much* investigator judgment has another name: bias. In addition, it is possible to capitalize on chance if one is not careful and disciplined.

One artifact of the estimation procedure is a tendency to inflate the dimensionality (Greenacre, 1994). The source of this inflation is the perfect association that is present in the diagonal blocks of the Burt matrix. These diagonal blocks are totally uninformative about the association structure of the variables with each other, but contribute variation to the decomposition. A number of different methods for correcting this have been proposed, from rescaling the eigenvalues in certain ways to a new solution algorithm that completely eliminates the diagonal blocks from consideration. All of these are discussed extensively in Greenacre (1994; Chapter 7). The simplest correction for this flaw rescales the eigenvalues as follows. Let Q be defined as above, the number of variables and let λ_k be the k -th eigenvalue. Furthermore, assume that $\lambda_k > 1/Q$. Then:

$$\tilde{\lambda}_k = \left(\frac{Q}{Q-1} \right)^2 \left(\sqrt{\lambda_k} - 1/Q \right)^2$$

provides a rescaled k -th eigenvalue. This eliminates all eigenvalues less than $1/Q$, which are artifacts of the solution, and rescales the rest of the eigenvalues appropriately. The computed fit of the solution will usually improve substantially.

How many dimensions in a solution are important? One determines this by examining the rescaled eigenvalues. By far the simplest way to do this is via the scree plot, which plots the eigenvalues in order against their index. One looks for an “elbow”

Verkuilen, “Measuring Membership”

in the eigenvalues; where it bends indicates the number of dimensions needed to get an adequate picture of the solution. It is also possible to normalize the eigenvalues by dividing by the sum of all the eigenvalues, which expresses each as a proportion of the total variation in the solution. I refer readers to the usual sources on factor analysis or principal components analysis, e.g., Rummel (1970), for details. Heiser and Meulman (1994) has a discussion of some innovative, more formal tests based on permutations that have proven useful in determining the dimensionality of solutions.

There is another potential method artifact to be aware of. The “horseshoe effect,” also known as the “Guttman effect,” occurs when a strongly dominant first dimension “bends” the higher dimension components, which is easy to see in a plot of the dimension two and higher category scores of the variables on the dimension one category scores. The second component will appear as a quadratic function of the first component, the third component will appear as a cubic function of the first component, and so forth. It is caused by the solution’s inability to contrast cases sufficiently that have very different profiles (Greenacre, 1984; pp. 226-232)—the metric produced by the solution breaks down at the edges. This issue is, in fact, common to many methods that use rank order information or less (Borg and Groenen, 1997) and can prove to be a nuisance in some circumstances, particularly those encountered in ecology. However, as Greenacre notes, in most cases the steps one needs to take to ameliorate the situation are not worth the costs. Viewed from another perspective, the presence of a horseshoe could be taken as an indicator of scale quality: “We have observed oscillatory principal components in almost all properly constructed attitude and aptitude scales, usually in the form of a horseshoe”

Verkuilen, “Measuring Membership”

(Gifi, 1990, p. 314). Thus, users are advised to plot the first few category weights to look for the presence of a horseshoe.

Solution Stability

One difficulty with all these methods is a paucity of analytic results for confidence intervals. In the traditional theory of statistical inference, one uses confidence intervals (or, equivalently, hypothesis tests) to assess statistical significance. This is a valid concern, but I want to broaden the notion to what Greenacre (1984) calls “stability,” which requires a brief digression about the nature of scientific inference. Many of the applications of this scaling procedure, or indeed statistical procedures in general, in the social sciences—particularly in comparative politics—cannot be thought of as being applied to a random sample from a population. Rather, in many cases the entire relevant population is to be examined and the purpose is to draw meaningful conclusions from the data, for instance, what effects are large and stably so, thus perhaps not due only to one or two influential cases but rather seem robust to perturbations. In the arena of designed experiments, the test of statistical significance is the *sine qua non*, but the assumptions of most statistical tests are rarely met in many areas of social science. Bayesian statistics provides one way of going about this task (Western and Jackman, 1994) without recourse to the assumptions underlying conventional statistical tests. Correspondence analysis/dual scaling is another tradition, less specifically tied to probabilistic considerations, that can address a broader notion of inference (Greenacre, 1984). Bayesian statistics has a connection with modeling, but only a relatively tenuous one with multivariate analysis (this may be changing) on one hand and logic on the other (see my

companion paper as well as Ragin, 2000, on the use of fuzzy sets in concept modeling and logical models in the social sciences). Correspondence analysis has a deep connection with multivariate analysis of course, but lacks a solid core of modeling.

Return to the matter of a confidence interval. Regardless of whether one thinks of confidence intervals as representing a stability analysis in the broad sense or the stronger but narrower notion of statistical significance, they are still essential. Unfortunately, while there are a number of analytic results for the distribution of eigenvalues of sample covariance matrices assuming multivariate normality, far fewer results hold more generally. As multivariate normality is a rather dubious assumption for discrete data, we are left looking for alternatives. The primary method that has been used in the literature is the bootstrap.

There are a number of different alternatives in applications of bootstrapping depending on one’s sampling model (Efron and Tibshirani, 1986). As we are interested in providing one score for each case based on the case’s total profile, the appropriate method here is to bootstrap entire cases, the so-called “Type I” bootstrap, which is the simplest form of the bootstrap (Michaelides and de Leeuw, 1998). That is, we create bootstrap samples (typically 250 to 1000 of them) by sampling *with replacement* from the original cases, which will result in some of the cases being duplicated and some omitted in each bootstrap sample.¹³ Each bootstrap sample is then analyzed in the usual way to generate scale quantifications.¹⁴ The distribution of the computed statistics is then used

¹³ It is important to note that bootstrapping cannot make information that is lacking in the original dataset, so it does not provide protection against selection bias.

¹⁴ There is an important nuisance effect that can occur when bootstrapping with procedures based on spectral decompositions. The spectral decomposition calculates the singular values and singular vectors, which are used to form an optimal basis for the original matrix. However, the singular vectors—which are, as discussed above where the object scores and item weights come from—are defined only up to an arbitrary change of scale and reflection; the signs are unidentified. We fix the scale by imposing

to assess how stable the computed weights are (because the scores are functions of the weights, they can also be computed from the weights). The easiest way to do this is to use a variation of Tukey’s five number summary: plot for each category (1) the weight from the original data (instead of the median), (2) q_1 and q_3 , the upper and lower quartile cutoffs from the bootstrap samples, and (3) some measure of extreme variation from the bootstrap samples, such as Tukey’s fences, that is, $q_1 - 1.5(q_3 - q_1)$ and $q_3 + 1.5(q_3 - q_1)$, or the minimum and maximum. A good solution will have the original category weight in the middle of the bootstrap quartiles and not have too much extreme variation.

It may be that a small number of observations are responsible for most of the “pull,” indicating that there are outliers in one’s data. Just as in ordinary least squares, outliers can cause a great deal of trouble. By the way the solution is defined, categories with very few observations tend to have more extreme quantifications, while more numerous categories tend to be close to zero (Gifi, 1990, pp. 112-113). In some cases these barely filled categories should be merged with adjacent categories to stabilize the solution. According to Nishisato, too many categories in an item cause a great deal of trouble on their own and can generate spurious results due to a phenomenon he calls “overquantification,” which occurs because the solution can have too much freedom to improve fit (Nishisato, 1994; pp. 321-330). He also suggests using the method of trimmed means in the calculation of the solution when outliers are suspected, though I do

identification constraints, but the constraints do not fix the sign, which would require an additional constraint. These reversals provide no real information about the variables but can mask it by generating a clearly bimodal distribution of quantifications, since the relative position of the values is what is important, not the absolute value. For instance, an M-shape and a W-shape are basically the same upside down, reflected about 0 by the algorithm converging from the opposite direction. By multiplying all quantifications within a variable by -1 where necessary, it is possible to match scaled items by their dominant feature, in effect turning all the M’s into W’s or vice versa. That is, if the dominant features are made sign-consistent with no regard to the other quantification values, which are reflected if the dominant feature is reflected. A dominant feature could be the maximal category value, for instance. This removes a source of nuisance variation but avoids alterations based on any other information.

not discuss this here. One way of detecting outliers is to use jackknifing. That is, by computing the solution space with each point deleted and examining it to see if it changes, it is possible to get a sense of the relative influence of the point. If the computed solution changes substantially when a point is deleted, this is a good indication that the point is influential. (Of course, “influential” does not imply “bad.” Throwing out data is rarely a good idea.) Cross-validatory schemes that are more complicated than simple jackknifing are also possible (Greenacre, 1984, pp. 218-219).

Missing or Ambiguous Data

We cannot always expect (a) to have complete data for all cases or (b) have complete agreement on ratings made when we have multiple coders. In the first case, it may simply be unknown what a survey respondent’s selection was due to illegibility in the response. In the second case, two coders may not arrive at the same conclusion. While these are two separate situations, the method proposed for handling these situations is fundamentally the same, namely probabilistic coding.¹⁵

Consider the super-indicator matrix \mathbf{Z} , which, recall, has N rows and J columns. We can partition \mathbf{Z} into Q blocks each with $J(i)$ columns, corresponding to the items: $\mathbf{Z} = (\mathbf{Z}_1|\mathbf{Z}_2|\dots|\mathbf{Z}_Q)$. The construction of each block ensures that the row sum within a block equals 1, while the row sum of the entire super-indicator matrix \mathbf{Z} is Q , since each row within a block will have at most one 1 with the rest of the entries being 0, and there are Q blocks total. Probabilistic coding relaxes this constraint by requiring that each block row

¹⁵ As mentioned previously, the literature sometimes refers to this technique as “fuzzy coding.” I prefer the term probabilistic coding because the system is based on the probabilistic notion of normalization with probability as a measure of membership in a well-defined set, unlike the fuzzy set, in which the set itself is fuzzy. There are certain resemblances to the fuzzy set complement, but I feel it is unwise for a term to do double duty.

sum to 1, but allowing more than one nonzero column entry per block. This assignment allows for representation of uncertainty in a way that is useful for handling both missing data and coder disagreement while preserving the mathematics so far described.

First I will discuss missing data. Greenacre (1984) has an extensive discussion of missing values in dual scaling, as does Gifi (1990) and Nishisato (1994). Their guidelines are quite pragmatic. While the general thrust requires a substantial input of investigator labor, there is no real substitute to knowing one’s data intimately. First, any case that is missing on a substantial portion of the items should be dropped from the analysis entirely. Second, for missing values on a particular item the investigator should think carefully about the meaning and assign coding accordingly. It may be desirable to consider “missing” to in fact be its own category. For example, consider the treatment of abstention in voting. Is an abstention a missing vote expressing *no* preference, or is it the expression of a preference to abstain? If it is the former, it makes sense to assign weights of 0.5 to yea and 0.5 to nay, indicating that each vote is weighted equally in this scheme. If it is the latter, it makes sense to assign it to its own category, in effect making the vote item have three categories (yea/nay/abstain), though it may not be clear what weights will be assigned to particular votes in this case. If one were to treat voting as having three components and a vote was genuinely missing, say particular legislators were absent for several months due to illness and there is uncertainty about how they would have voted in that time, it would make sense to assign each category a weight of 0.33. It is also possible to use more sophisticated imputation procedures that estimate the missing data as additional parameters based on the observed relationships within the data. Though I will not discuss this issue here, it is an interesting point for future research. Nishisato

Verkuilen, “Measuring Membership”

(1994; Chapter 15) has a very useful discussion of the tradeoffs of various methods of handling missing data in dual scaling applications.

The treatment of coder disagreement by probabilistic coding is similar to handling missing values. If there are C coders and $J(i)$ categories in an item on which they disagree, simply using the relative frequency within each item category gives a reasonable weighting that expresses the disagreement. If, given a two category item and six coders, four of the coders assign to one category and two to the other category, it makes sense to code the item as $(0.67, 0.33)$ —proportional representation, as it were—not as $(1, 0)$ —majority rule. The probabilistic coding more accurately expresses the disagreement of the coders than the majority rule assignment would. In addition, since the object scores and item weights are numerical (interval), it is possible to examine intercoder reliability by generating object scores and item weights for each coder individually and analyzing the relationships between them subsequently by the usual techniques.

VI. Example¹⁶

In this section I will briefly consider an example using dual scaling to create two membership functions to examine the data in Russett (1964) on the relationship between economic inequality and political instability. There are three reasons to use these data. First, it was analyzed in Gifi (1990, pp. 223-232) and thus is, in part, a replication of an existing analysis (though Gifi used a slightly different method than the one considered here, working with nonlinear canonical correlation). Second, the data set is small

¹⁶ Computations were performed with a mix of SPSS Categories 10.0, R 1.2.1, and Microsoft Excel 2000. The author is working on Splus/R code to compute scales of this sort.

enough—less than a fifty cases—to be highly manageable for expository purposes. Third, it illustrates the technique in a “medium N” setting which resembles many analyses in comparative politics, particularly since no sampling mechanism is described and probably not the case that the data are a random sample. In particular, Russett’s verbal exposition and his data analysis do not entirely match. I refer readers to Russett’s article for the original data and a discussion of each variable. The Gifi discretization¹⁷—which all computations here will be based on—is included in Table 3 (Gifi, 1990, pp. 225).¹⁸ For lack of space, unfortunately, I will not be able to use all the techniques discussed above, such as bootstrap confidence intervals or probabilistic coding, here. In addition, there are a number of additional examinations one might want to make if the analysis was “for real.” I hope it will be sufficiently illustrative to give a flavor of how the scaling proceeds, however.

[Table 3 about here]

Russett’s theory examines the relationship between economic inequality and political instability. His contention is that economic inequality causes political instability and he uses a relatively ad hoc correlation analysis to examine this hypothesis, correlating all variables against each other and reporting some of the high correlations. This article is, of course, nearly forty years old and the data analysis is fairly primitive by contemporary standards. However, Russett’s thinking in the article is very much in line with that expressed previously in this article. He examines several indicators of both inequality

¹⁷ A number of developments in dual scaling have made discretization less necessary. Current research examines the use of splines for scaling continuous data. These have some serious advantages in that they avoid the difficulties and potential biases of discretizing continuous variables. They also tend to generate smoother solutions, which can be useful. Implementations are available in the SPSS 10.0 Categories CATPCA program. It would be interesting to compare the spline solution computed on Russett’s original data to the one computed on the discrete data.

¹⁸ Three of Russett’s variables, GNP per Capita, % of Labor in Agriculture, and Democracy, are not used here, but are included in the data table for completeness.

Verkuilen, “Measuring Membership”

and instability, recognizing that none of the indicators given are perfect measures of the concepts he had in mind. Clearly, the toolbox of methods available to him in 1964 was not up to the task he set himself.

To illustrate the techniques discussed in this article, I will generate fuzzy set membership functions for “economic inequality” and “political instability” from his indicator variables. To generate the set “Inequality” I use three indicators, the Gini index of land owning, the percent of farms with one half of the land, and the percentage of farms that are rented. These are denoted GINI, FARM, and RENT, respectively and the Burt matrix for these items is in Table 4. Recall that the diagonal blocks show the marginal frequencies of each item and the off diagonal blocks show the crosstabulations of the items with the other items. To generate the set “Instability” I also use three indicators, personnel instability, which measures the amount of turnover in government, the Eckstein internal war index, and the death rate per million due to civil group violence. These are denoted INST, ECKS, and DEAT, respectively and the Burt matrix for these items is in Table 5. In Russett’s analysis (and indeed in the Gifi analysis) these variables are all *assumed* to be monotonic. That is, higher scores on these variables should indicate more of the concept measured. As we will see below, one of the items in the Instability scale does not seem to behave monotonically with respect to the other variables, which would explain the difficulties the Gifi group had in computing the solution.

[Tables 4 and 5 about here]

First the Inequality scale. Simple examination of the Burt table shows these items seem to go together fairly well, with most of the crosstabulations showing a classic one-to-one relationship pattern. With three variables that have six, five, and six categories

Verkuilen, “Measuring Membership”

each, there are $6^2 \cdot 5 = 310$ different possible response profiles, which of course exceeds the size of the sample. One subtlety is that for the variable RENT the last category (9) is really a missing value in the original Russett data. There are three cases that are missing on this variable. Unfortunately, we have no good expectation about the value this should have a priori; dual scaling will, in effect, impute a value for this category based on the observed covariation with the other items. Based on the performance it will show, I will treat it as the maximal category for use with the Healy-Goldstein rescaling, though this decision turns out to make little difference since its value and the next category’s value are fairly close.

It seems there are only a relatively few response profiles that are common, which is what we would expect given a unidimensional pattern. How well is this eyeball analysis borne out formally? The first three adjusted eigenvalues are 0.79, 0.34, and 0.078, which clearly elbow at the first eigenvalue; it accounts for 65% of the variation. Figure 2 shows the raw, i.e., unrescaled to fit in the unit interval, which we will do subsequently, category weights for the Inequality items. The horseshoe pattern is quite clear, though the RENT item does not seem to behave quite as well as the other two, because of the missing value category. This suggests that the multidimensionality present is due, by and large, to the inability of the solution to contrast very different profiles sufficiently, not because there is “genuine” multidimensionality here. However, this pattern is not perfect, so further examination would be wise.

[Figure 2 about here]

Table 6 has the category scores for the first dimension. One of the most notable properties of this table and the previous figure) is that the objects with very low values

Verkuilen, “Measuring Membership”

are contrasted strongly from the objects with very high values, all of which are assigned relatively similar scores, which suggests that a dichotomy on each variable set contrasting the minimal category with the other categories would not do the data injustice. The pattern is quite close to monotonic in all variables, with only slight deviations in GINI and RENT.

[Table 6 about here]

We will use the Healy-Goldstein rescaling here, which identifies the minimal profile (1,1,1) with membership 0 and the maximal profile (6,5,9) with membership 1. Within each item, we set the minimal category to zero by subtracting the minimal category’s weight from all weights. Then, summing the scores for the maximal profile, we divide all weights by this sum, which scales every value into the unit interval. There are cases that have both profiles, Yugoslavia and Poland for (1,1,1), and Austria for (6,5,9), so the matter of subnormality within the data itself does not show up here.¹⁹ Table 7 has these rescaled scores, which are no longer strictly optimal in the least squares sense described previously, but are now interpretable in terms of membership functions, which we will do after scaling the Instability indicators. Items that are “longer” contribute more to the membership than items that are “shorter;” this length is a measure of the relative importance of each item to the membership. For instance, RENT’s maximum value is 0.39 while GINI’s is 0.27 so RENT contributes a bit more to the membership. If we look at values of categories within an item, we can see the relative contribution to the membership from switching categories. Switching from RENT = 1 to RENT = 2 increases the membership by 0.18. By contrast, switching from RENT = 4 to RENT = 5

¹⁹ Because of the data at any rate—an investigator would be well advised to consider the derived scale on theoretical grounds.

Verkuilen, “Measuring Membership”

hardly changes the membership at all. It is often informative to plot these values against the categories but I do not do so here.

Now for the Instability scale. The analysis performed is similar to the one for Inequality, but it is much more problematic. Note there are $6*3*5 = 90$ different possible profiles. The adjusted eigenvalues are 0.26, 0.06, and 0.05, with the first eigenvalue accounting for 70% of the observed variation. However, while two of the three variables, ECKS and DEAT, fit a monotonic pattern reasonably well, INST does not fit well at all in any kind of monotonic fashion, as Figure 3 shows quite clearly. ECKS and DEAT both exhibit the horseshoe pattern, more or less, but INST does not nor are its category weights remotely monotonic. INST horseshoes in the wrong dimension and the pattern in the first dimension seems to be upside down quadratic, not monotonically increasing. Examining the crosstabulations ECKS×INST and DEAT×INST in the Burt matrix (Table 5) backs this up—INST does not move entirely in tandem with the other two variables.

[Figure 3 about here]

All this suggests that we would be wise to drop INST from the analysis. In fact, this sort of within-category item analysis is one of the principal selling points of the dual scaling approach, which, as mentioned previously, allows one to *test* the assumption of monotonic response rather than *assuming* it. The more usual additive scale approach would involve examining the correlation matrix of these three items and perhaps examining a reliability statistic such as Cronbach’s alpha.²⁰ Table 8 shows the correlation matrix with the Cronbach’s alpha scores for the scale with items deleted for the sum of standard scores of these three variables. This procedure also seems to indicate

²⁰ Cronbach’s alpha is, in fact, a monotonic function of the dual scaling eigenvalues for the first dimension transformed variables. The alpha for untransformed variables will, of course differ, in general being lower than the alpha for the transformed variables since the transformation also maximizes Cronbach’s alpha.

Verkuilen, “Measuring Membership”

that there is something “wrong” with INST, since alpha would climb by 0.1 if it were removed and drop substantially if any of the other variables were left in, though it is not entirely clear what it might be. Because alpha is fairly high and the two items to be used are monotonic we would not be far off to use a simple additive scale here.²¹

[Table 8 about here]

It is possible to use only two items to create a scale via dual scaling (this is the minimum number), though of course more is better. In this case, we rely on technique that is a special case of dual scaling, correspondence analysis. The mathematics is exactly the same and here we get a consistent pattern, except that the DEAT variable is not entirely monotonic. The first dimension accounts for over 85% of the variance in the table, which is an excellent solution, clearly one dimensional. The raw weights are in Table 9 and the rescaled weights are in Table 10.

[Tables 9 and 10 about here]

After some calculation we arrive at membership values for each scale, given in Table 11. Figure 4, a scatterplot of the two membership functions shows a lower triangle plot, the interpretation of which in terms of entailment is discussed in Ragin (2000, Chapter 8) or Smithson (1987, Chapters 3, 5, and 6). This sort of plot is quite common indeed in cross-national research.

²¹ This, of course, begs the question of why dual scaling should be used at all as opposed to the much more simple additive scale. First and foremost, the dual scaling used here demonstrates that the items are monotonic rather than assuming they are. It also makes more use of the information in the data because it provides the investigator the opportunity to examine how items that are rejected during scale construction are related to the others that are kept. Third, it is a relatively common occurrence in cross-national research to have an item that has three or four theoretically ordered categories and one category that is not, usually a residual category that is present to classify cases that do not fit into the other categories for some reason. It is rarely clear how to handle these cases and they can be quite frequent, so it is unwise simply to drop them from analysis, particularly when sample sizes are fairly small—as is typically the case in cross-national research. Dual scaling in effect imputes a value for these categories based on the interrelations of the cases' values on other items.

[Figure 4 about here]

The plot suggests that Inequality is a (fuzzily) necessary but not sufficient condition for Instability. Alternatively, one can interpret Instability as a fuzzy subset of Inequality. Correlating these two scales gives a Pearson-r of 0.30, which, while in the direction predicted, is hardly anything to write home about. Pearson-r is, of course, a measure of one-to-one association, and is symmetric since for any two variables X and Y $r_{XY} = r_{YX}$. Fuzzy set researchers have developed a number of methods for measuring one-to-many associations, which are asymmetric in the sense that X’s relationship to Y is not the same as Y’s relationship to X. Smithson’s coefficient of inclusion “measures the proportion of deviations away from the diagonal which fall into the appropriate triangular region” and is thus an index of the degree to which points fall on the “wrong” side of the diagonal, weighted by the distance from the diagonal (1987, p. 101). It is normalized to the unit interval, with higher values indicating stronger degrees of inclusion. In this case, the coefficient of inclusion is 0.96, which indicates a very strong, nearly perfect fuzzy subsethood relationship. Because most of the cases fall fairly far away from the reference line $y = x$, it is possible to consider making the relationship between these two sets stronger by considering more stringent reference curves that cut the unit square into two regions, one with no points and one with all points (or, more realistically, few points and most points). These reference curves have interpretations in terms of inclusion in new sets derived from the base fuzzy sets.

VII. Future Directions

In this paper I have discussed the scaling of membership functions for social science applications from a very general perspective. In addition I have proposed a particular technique, dual scaling, for accomplishing this task in a context that I believe is quite common in social scientific practice. I believe it is a useful addition to the toolbox of anyone using fuzzy set theory. Dual scaling is more general than that and there is no inherent connection between it and fuzzy set theory, but I believe this article is an illustration of one way in which they could touch fruitfully.

There are two different directions in which this research could go. One is to incorporate some of the more advanced techniques from the dual scaling literature. The scaling derived above is one of the most basic results of the scaling literature (it was derived in rudimentary form in the 1930's and more fully in the 1940's) and researchers in that area have made a number of advances, some of which I have mentioned in relevant places in the paper. While there have been advancements, the solution described above is still very much the workhorse of nonlinear multivariate analysis and the standard against which the rest of the methods are defined, so this statement should not be taken as an indicator of low quality. On the other hand, fuzzy set theory can provide insights into multivariate data analysis. For instance, Smithson (1987) used fuzzy set theory to define a continuous version of the Guttman scale and also made use of fuzzy set methods in a multidimensional scaling application. Others have used fuzzy set methods in clustering applications. Many of these methods rest, at the bottom, on correlations of some sort, so it may be that other relationships such as fuzzy inclusion or fuzzy similarity would provide useful bases for scaling. At minimum, it is necessary to make a systematic and thorough examination of how dual scaling performs versus alternative methods for

Verkuilen, “Measuring Membership”

scaling membershipo, ranging from simple direct ratings to additive scales to other sophisticated methods such as polytomous IRT models. With high quality data it is true that nearly any method will provide solid results, whereas it is relatively poor data—like that used in most applications in comparative social science—that requires more sophisticated tools. To quote Gary King, “Indeed, sophisticated methods are required only when data are problematic; extremely reliable data with sensational relationships require little if any statistical analysis” (King, 1989, p. 3).

What should this mean to comparative scholars? The qualitative-quantitative debate has been one of the most persistent and contentious ones in the discipline since the behavioral revolution. To some extent it is beyond resolution—it represents important differences of “intellectual style.” However, in a very real sense, the debate is fueled by the simple fact that many of the quantitative tools used by comparativists, mostly inherited from tools developed by scholars working in an experimental setting, have not been up to the demands posed by analysis in comparative applications and have in many cases been “stretched” into applications they were never designed for. For most qualitative scholars, I believe it is not a stretch to say that regression analysis is viewed as not simply one choice among many methods for quantitative analysis, but the *only* choice. Unfortunately some quantitative scholars considering qualitative research, e.g., King et al (1994), have done little to dispel this myth. Regression, especially used the way in which typically appears—additive models with several variables—does a poor job of translating theories that are specified (one hopes) in logical terms.

One consistent criticism that qualitative scholars make is that quantitative translations of their verbal theories have often been simplistic or simply wrong. To some extent, this

Verkuilen, “Measuring Membership”

is because qualitative scholars have not always done a very good job specifying their own theories in verbal terms that are precise enough to be empirically testable. Mathematics’ big advantage as a theoretical language is precision; it is hard to hide behind fancy but fundamentally empty wording and imprecision or contradictions in theory are often revealed in the process of mathematical translation. Mathematics also provides some useful methods for empirical testing that can extend a scholar’s mental grasp substantially. However, not all the blame for the mismatch can be apportioned to verbal theories. Much instead must be born by the lack of appropriate mathematical structures in use by quantitative scholars. Thus fuzzy set theory may prove to be a useful source of quantitative tools that can do a much better job of matching the interesting verbal theories of qualitative scholars with mathematical realizations and empirical data, particularly since, as Smithson (1994) notes, “fuzzy set theory can provide an alternative interpretation for standard statistics that link ‘qualitative’ (setwise) concepts with ‘quantitative’ measures.”

Tables and Figures

Figure 1: Three Species-Four Site Incidence Matrix

	Site 1	Site 2	Site 3	Site 4
Species A	+	+	-	-
Species B	-	+	+	-
Species C	-	-	+	+

Table 1: Sample Super-Indicator Matrix with $Q = 3$, and $J(i) = 3, 3,$ and 2

Item 1			Item 2			Item 3	
Option 1	Option 2	Option 3	Option 1	Option 2	Option 3	Option 1	Option 2
1	0	0	0	1	0	1	0
0	1	0	0	0	1	0	1
0	0	1	1	0	0	1	0
...

Table 2: Burt Matrix of Healy & Goldstein (1976) Data

	A1	A2	A3	B1	B2	B3	C1	C2	C3
A1	11440	0	0	5923	5134	383	5957	5254	229
A2	0	667	0	143	440	84	135	468	64
A3	0	0	125	22	62	41	18	70	37
B1	5923	143	22	6088	0	0	3896	2111	81
B2	5134	440	62	0	5636	0	2084	3387	165
B3	383	84	41	0	0	508	130	294	84
C1	5957	135	18	3896	2084	130	6110	0	0
C2	5254	468	70	2111	3387	294	0	5792	0
C3	229	64	37	81	165	84	0	0	330

Table 2: Gifi (1990) Discretization of Russett’s (1964) Data

Country	GINI	FARM	RENT	GNPC	LABO	INST	ECKS	DEAT	DEMO
ARG	5	5	4	4	2	5	3	4	2
AUS	6	5	9	7	1	4	1	1	1
AUT	4	5	2	5	2	5	2	1	2
BEL	2	3	5	6	1	6	2	2	1
BOL	6	5	3	1	4	6	3	5	3
BRA	5	5	2	4	4	6	3	2	2
CAN	1	2	2	7	1	4	2	1	1
CHL	6	5	3	3	2	5	2	2	2
COL	5	5	3	4	3	5	3	4	2
COS	5	5	2	4	3	5	2	3	2
CUB	4	5	5	4	3	5	3	5	3
DEN	1	1	2	6	2	5	1	1	1
DOM	4	5	3	3	3	4	2	3	3
ECU	5	5	3	3	3	6	3	2	3
EGY	4	5	3	2	4	6	3	2	3
ELS	5	5	3	3	4	6	2	2	3
FIN	2	3	2	6	3	6	2	1	2
FRA	2	3	4	6	2	6	3	2	2
GUA	5	5	3	3	4	5	3	3	3
GRE	4	5	3	3	3	6	2	2	2
HON	4	5	3	2	4	5	3	4	3
IND	2	3	5	1	4	2	3	2	1
IRQ	5	5	5	3	5	6	2	4	3
IRE	2	3	2	5	2	5	2	1	1
ITA	5	5	4	5	2	6	3	2	2
JAP	1	2	2	3	2	6	2	2	2
LIB	3	2	4	1	4	5	2	1	3
LUX	3	3	3	7	2	5	1	1	1
NTH	3	3	5	6	1	5	2	1	1
NWZ	4	5	4	7	1	5	1	1	1
NIC	4	5	9	4	4	5	2	2	3
NOR	3	3	2	6	2	5	2	1	1
PAN	4	4	3	4	3	6	2	3	3
PER	5	5	9	2	3	5	2	3	3
PHL	2	3	4	3	5	5	2	4	1
POL	1	1	1	5	3	3	2	2	3
SVI	3	4	3	2	4	4	3	5	3
SPA	4	5	5	4	3	1	2	2	3
SWE	2	3	3	7	1	3	1	1	1
SWI	1	2	3	7	1	3	1	1	1
TAI	3	4	4	2	3	1	2	1	3
UNK	4	4	5	6	1	5	2	1	1
USA	4	5	3	8	1	5	2	1	2
URU	5	5	4	5	2	5	2	2	1
VEN	6	5	3	6	3	5	2	4	3
WGE	3	4	2	6	1	2	2	1	2
YUG	1	1	1	4	4	1	2	1	3

Table 4: Burt Matrix of Russett’s (1964) Data, Inequality Items

	GINI1	GINI2	GINI3	GINI4	GINI5	GINI6	FARM1	FARM2	FARM3	FARM4	FARM5	RENT1	RENT2	RENT3	RENT4	RENT5	RENT9
GINI1	6	0	0	0	0	0	3	3	0	0	0	2	3	1	0	0	0
GINI2	0	7	0	0	0	0	0	0	7	0	0	0	2	1	2	2	0
GINI3	0	0	7	0	0	0	0	1	3	3	0	0	2	2	2	1	0
GINI4	0	0	0	12	0	0	0	0	0	2	10	0	1	6	1	3	1
GINI5	0	0	0	0	11	0	0	0	0	0	11	0	2	4	3	1	1
GINI6	0	0	0	0	0	4	0	0	0	0	4	0	0	3	0	0	1
FARM1	3	0	0	0	0	0	3	0	0	0	0	2	1	0	0	0	0
FARM2	3	0	1	0	0	0	0	4	0	0	0	0	2	1	1	0	0
FARM3	0	7	3	0	0	0	0	0	10	0	0	0	3	2	2	3	0
FARM4	0	0	3	2	0	0	0	0	0	5	0	0	1	2	1	1	0
FARM5	0	0	0	10	11	4	0	0	0	0	25	0	3	12	4	3	3
RENT1	2	0	0	0	0	0	2	0	0	0	0	2	0	0	0	0	0
RENT2	3	2	2	1	2	0	1	2	3	1	3	0	10	0	0	0	0
RENT3	1	1	2	6	4	3	0	1	2	2	12	0	0	17	0	0	0
RENT4	0	2	2	1	3	0	0	1	2	1	4	0	0	0	8	0	0
RENT5	0	2	1	3	1	0	0	0	3	1	3	0	0	0	0	7	0
RENT9	0	0	0	1	1	1	0	0	0	0	3	0	0	0	0	0	3

Table 5: Burt Matrix of Russett’s (1964) Data, Instability Items

	INST1	INST2	INST3	INST4	INST5	INST6	ECKS1	ECKS2	ECKS3	DEAT1	DEAT2	DEAT3	DEAT4	DEAT5
INST1	3	0	0	0	0	0	0	3	0	2	1	0	0	0
INST2	0	2	0	0	0	0	0	1	1	1	1	0	0	0
INST3	0	0	3	0	0	0	2	1	0	2	1	0	0	0
INST4	0	0	0	4	0	0	1	2	1	2	0	1	0	1
INST5	0	0	0	0	22	0	3	14	5	10	3	3	5	1
INST6	0	0	0	0	0	13	0	7	6	1	9	1	1	1
ECKS1	0	0	2	1	3	0	6	0	0	6	0	0	0	0
ECKS2	3	1	1	2	14	7	0	28	0	12	9	4	3	0
ECKS3	0	1	0	1	5	6	0	0	13	6	1	3	3	3
DEAT1	2	1	2	2	10	1	6	12	0	18	0	0	0	0
DEAT2	1	1	1	0	3	9	0	9	6	0	15	0	0	0
DEAT3	0	0	0	1	3	1	0	4	1	0	0	5	0	0
DEAT4	0	0	0	0	5	1	0	3	3	0	0	0	6	0
DEAT5	0	0	0	1	1	1	0	0	3	0	0	0	0	3

Figure 2: Inequality Scale Raw Category Weights

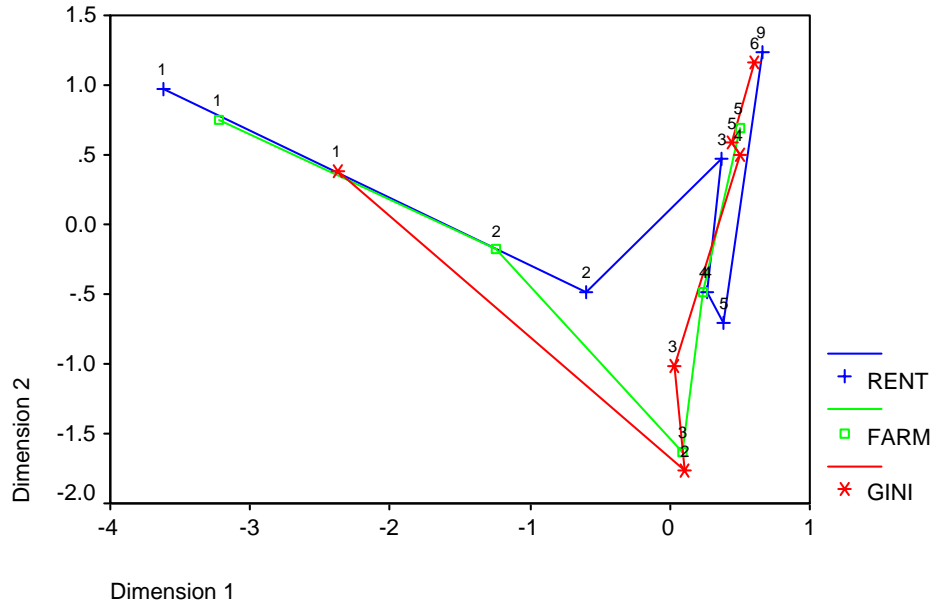


Table 6: Inequality Scale Raw Category Weights

Category	1	2	3	4	5	6/9
GINI	-2.38	0.11	0.03	0.5	0.45	0.61
FARM	-3.23	-1.24	0.10	0.24	0.50	--
RENT	-3.63	-0.59	0.37	0.27	0.39	0.66

Table 7: Inequality Scale Rescaled Category Weights

Category	1	2	3	4	5	6/9
GINI	0	0.226	0.22	0.26	0.26	0.27
FARM	0	0.181	0.3	0.32	0.34	--
RENT	0	0.276	0.36	0.35	0.37	0.39

Figure 3: Instability Scale Raw Category Weights

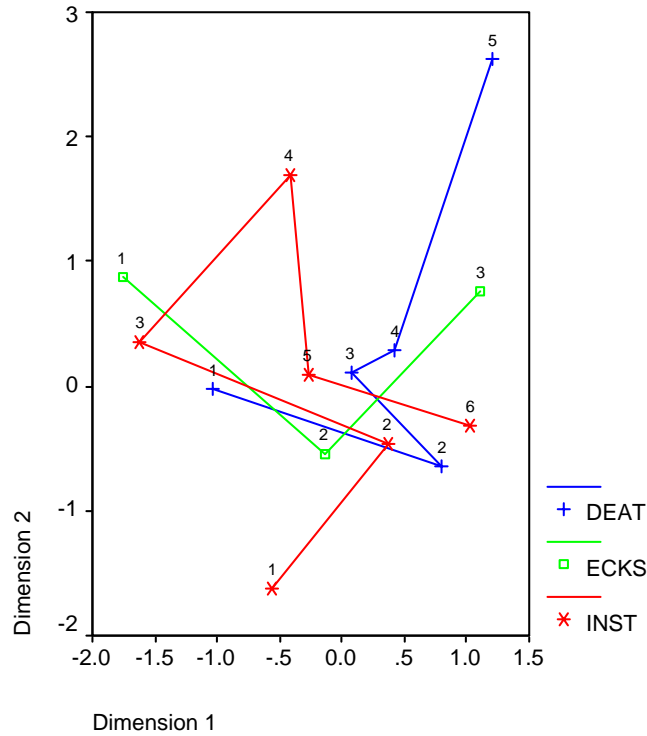


Table 8: Instability Additive Scale Approach (Current Alpha 0.62)

	INST	ECKS	DEAT	Alpha
INST	1.00	--	--	0.73
ECKS	0.23	1.00	--	0.41
DEAT	0.25	0.58	1.00	0.37

Table 9: Instability Scale Raw Category Weights (Dropping INST)

Category	1	2	3	4	5
ECKS	-1.40	-0.25	1.18	--	--
DEAT	-0.94	0.48	0.06	0.69	1.76

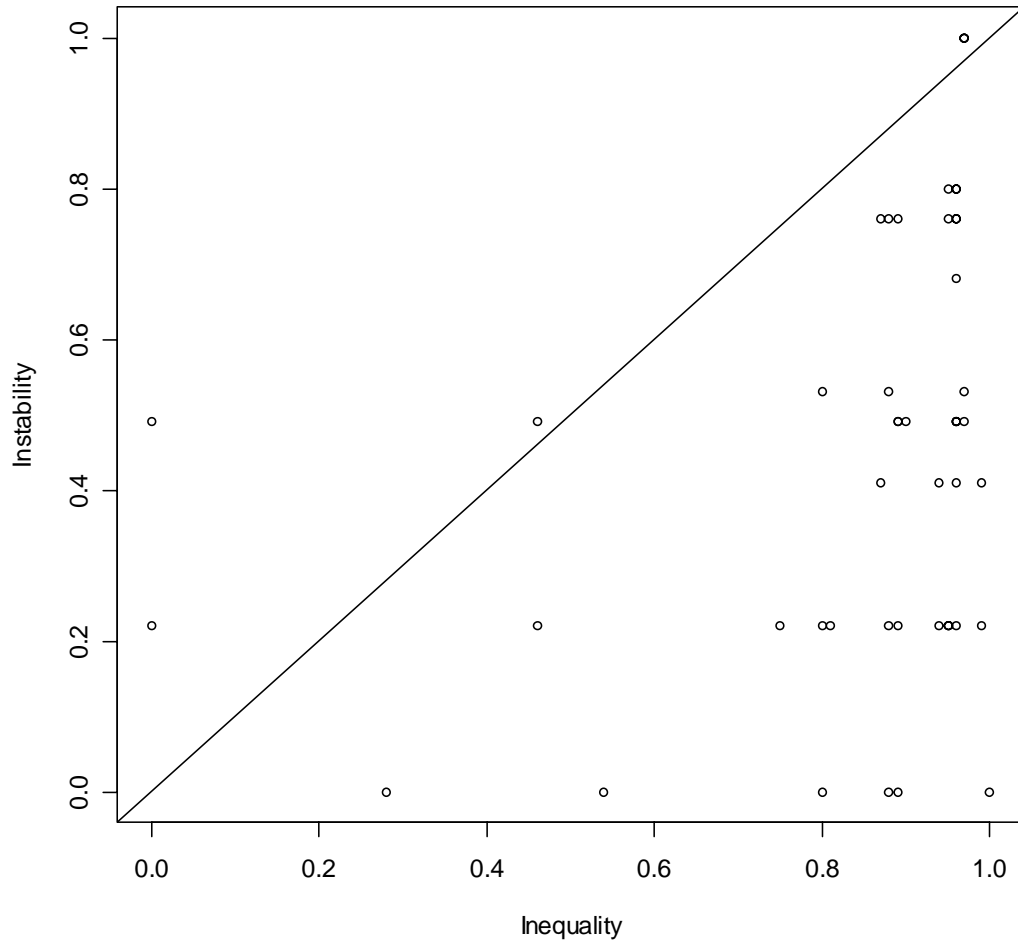
Table 10: Inequality Scale Rescaled Category Weights (Dropping INST)

Category	1	2	3	4	5
ECKS	0	0.22	0.49	--	--
DEAT	0	0.27	0.19	0.31	0.51

Table 11: Derived Membership Scores

Country	Inequality	Instability	Country	Inequality	Instability
ARG	0.95	0.80	ITA	0.95	0.76
AUS	1.00	0.00	JAP	0.46	0.49
AUT	0.88	0.22	LIB	0.75	0.22
BEL	0.89	0.49	LUX	0.88	0.00
BOL	0.97	1.00	NTH	0.99	0.22
BRA	0.87	0.76	NWZ	0.80	0.00
CAN	0.46	0.22	NIC	0.89	0.49
CHL	0.97	0.49	NOR	0.95	0.22
COL	0.96	0.80	PAN	0.94	0.41
COS	0.87	0.41	PER	0.99	0.41
CUB	0.97	1.00	PHL	0.88	0.53
DEN	0.28	0.00	POL	0.00	0.49
DOM	0.96	0.41	SVI	0.97	1.00
ECU	0.96	0.76	SPA	0.90	0.49
EGY	0.96	0.76	SWE	0.89	0.00
ELS	0.96	0.49	SWI	0.54	0.00
FIN	0.80	0.22	TAI	0.89	0.22
FRA	0.88	0.76	UNK	0.94	0.22
GUA	0.96	0.68	USA	0.95	0.22
GRE	0.96	0.49	URU	0.96	0.49
HON	0.96	0.80	VEN	0.97	0.53
IND	0.89	0.76	WGE	0.81	0.22
IRQ	0.80	0.53	YUG	0.00	0.22
IRE	0.96	0.22	Median	0.94	0.49

Figure 4: Scatterplot of Derived Inequality and Instability Membership Scores



Bibliography

- Bardossy, Andras and Lucien Duckstein (1995). *Fuzzy Rule-Based Modeling with Applications to Geophysical, Biological, and Engineering Systems*, New York: CRC Press.
- Beck, Nathaniel and Simon Jackman (1998). "Beyond Linearity by Default: Generalized Additive Models," *AJPS*, 42(2), pp. 596-627.
- Bollen, Kenneth (1993). "Liberal Democracy: Validity and Method Factors in Cross-National Measures," *AJPS*, 37(4), pp. 1207-1230.
- Borg, Ingwer and Patrick Groenen (1997). *Modern Multidimensional Scaling*, New York: Springer.
- Cioffi-Revilla, Claudio (1981), "Fuzzy Sets and Models of International Relations," *AJPS* 25(1), pp. 129-159.
- David Collier and Robert Adcock (1999). "Democracy and Dichotomies: A Pragmatic Approach to Choices about Concepts," *Ann. Rev. Polit. Sci.*, 2, pp. 537-565.
- Diamond, Larry (1992). "Economic development and democracy reconsidered," *American Behavioral Scientist*, 35, Mar./June, p. 450-99.
- Efron, Bradley & Robert Tibshirani (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Stat. Sci.*, 1(1), pp. 54-77.
- Geer, John P. van de (1986). *Introduction to Linear Multivariate Data Analysis*, Volume 1, Leiden: DWSO Press.
- Gifi, Albert (1990). *Nonlinear Multivariate Analysis*, New York: Wiley.
- Gleditsch, Kristian S. and Michael D. Ward (1997). "Double Take: A Re-examination of Democracy and Autocracy in Modern Polities," *JCR*, 41, pp. 361-82.
- Greenacre, Michael J. (1984). *Theory and Applications of Correspondence Analysis*, London: Academic Press.
- Greenacre, Michael J. (1994), "Multiple and Joint Correspondence Analysis," in Greenacre and Blasius (1994).
- Greenacre, Michael J. and Jorg Blasius (1994). *Correspondence Analysis in the Social Sciences*, London: Academic Press.

Verkuilen, “Measuring Membership”

Guttman, Louis (1977). “What Is Not What in Statistics?” *The Statistician*, 26(2), pp. 81-107.

Healy, M. J. R. and H. Goldstein (1976). “An Approach to the Scaling of Categorized Attributes,” *Biometrika*, 63, pp. 219-229.

Heiser, Willem J. and Jacqueline J. Meulman (1994), “Homogeneity Analysis: Exploring the Distribution of Variables and their Nonlinear Relationships,” in Greenacre and Blasius (1994).

Hill, M. O. (1974). “Correspondence Analysis: A Neglected Multivariate Method,” *Applied Statistics*, 23(3), pp. 340-354.

Jacoby, William G. (1999), “Levels of Measurement and Political Research: an Optimistic View,” *AJPS* 43(1), pp. 271-301.

Johnson, Valen E. and James H. Albert (1999). *Ordinal Data Modeling*, New York: Springer.

Kaplan, Abraham (1964). *Conduct of Inquiry*, San Francisco: Chandler.

King, Gary (1989). *Unifying Political Methodology*, New York: Cambridge University Press.

King, Gary, Robert O. Keohane, and Sidney Verba (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton, NJ: Princeton University Press.

Klir, George A. and Bo Yuan (1995). *Fuzzy Sets and Fuzzy Logic*, New York: Academic Press.

Lieberman, Evan S. (2000). “Cross National Measurement of Taxation,” paper presented at the 2000 Annual Meeting of the APSA.

Manton, Kenneth G., Max A. Woodbury, and H. Dennis Tolley (1994). *Statistical Applications Using Fuzzy Sets*, New York: Wiley.

McDonald, Roderick P. (1981), “The Dimensionality of Tests and Items,” *British J. Math. Stat. Psych.*, 34(1), pp. 100-117.

McDonald, Roderick P. (1999). *Test Theory: A Unified Treatment*, Hillsdale, NJ: Lawrence Erlbaum.

Michailides, George & Jan de Leeuw (1998), “The Gifi System of Descriptive Multivariate Analysis,” *Stat. Sci.* 13(4), pp. 307-336.

Verkuilen, “Measuring Membership”

- Nishisato, Shizuhiko (1994). *Elements of Dual Scaling*, Hillsdale, NJ: Lawrence Erlbaum.
- Norwich, A. M., and I. B. Turksen (1983). “A Model for the Measurement of Membership and the Consequences of Its Empirical Implementation,” *Fuzzy Sets and Systems*, 12, pp. 1-25.
- Poole, Keith T. (1998). “Recovering a Basic Space From a Set of Issue Scales,” *AJPS*, 42(3), pp. 954-993.
- Ragin, Charles C. (2000). *Fuzzy-Set Social Science*, Chicago: University of Chicago Press.
- Roberts, Fred S. (1979). *Measurement Theory*, Reading, MA: Addison-Wesley.
- Rummel, Rudolph J. (1970). *Applied Factor Analysis*, Evanston, IL: Northwestern University Press.
- Russett, Bruce M. (1964). “Inequality and Instability,” *World Politics*, 21, pp. 442-454.
- Saaty, Thomas L. (1986). “Scaling the Membership Function,” *European Journal of Operational Research*, 25, pp. 320-329.
- Sanjian, Gregory S. (1988), “Fuzzy Set Theory and U.S. Arms Transfers: Modeling the Decision-Making Process,” *AJPS*, 32(4), pp. 1018-1046.
- Seitz, Steven Thomas (1994), “Apollo’s Oracle: Strategizing for Peace,” *Synthese* 100, pp. 461-495.
- Smithson, Michael (1987). *Fuzzy Set Analysis for Behavioral and Social Sciences*, New York: Springer.
- Smithson, Michael (1994). *FuzzyStat Version 3.1 Tutorial Guide*, unpublished manuscript, James Cook University.
- Taber, Charles S. (1992), “POLI: An Expert System Model of U. S. Foreign Policy Belief Systems,” *APSR*, 86(4), pp. 888-904.
- Turksen, I. B. (1991). “”Measurement of Membership Functions and Their Acquisition,” *Fuzzy Sets and Systems*, 40, pp. 5-38.
- Western, Bruce and Simon Jackman (1994). “Bayesian Inference for Comparative Research,” *APSR*, 88(2), pp. 412-423.
- Zadeh, Lotfi A. (1965). “Fuzzy Sets,” *Information and Control*, 8, pp. 338-353.