Fuzzy Sets: Calibration Versus Measurement

Charles C. Ragin
Professor of Sociology and of Political Science
University of Arizona
Tucson, Arizona 85721 USA

Calibration is a necessary and routine research practice in such fields as chemistry, astronomy, and physics (Pawson 1989:135-7). In these and other natural sciences, researchers *calibrate* their measuring devices and the readings these instruments produce by adjusting them so that they match or conform to dependably known standards. These standards make measurements interpretable (Byrne 2002). A temperature of 20 degrees centigrade is interpretable because it is situated in between 0 degrees centigrade (water freezes) and 100 degrees centigrade (water boils). By contrast, the calibration of measures according to agreed upon standards is relatively rare in the social sciences.[1] Most social scientists are content to use uncalibrated measures, which simply show the positions of cases relative to each other. With an uncalibrated measure of temperature it is possible to know that one object has a higher temperature than another or even that it has a higher temperature than average, for a given set of objects, but still not know whether it is hot or cold.

This essay explores the connections between measurement and calibration in the social sciences and addresses its long-standing neglect. My starting point is the contrast between conventional approaches to measurement in quantitative and qualitative social research. After sketching common measurement practices in both types of research, I argue that a useful way for social scientists to incorporate measurement calibration into their research is through the use of fuzzy sets. In order to use fuzzy sets effectively, researchers must assess the degree of

membership of cases in well-defined sets (e.g., degree of membership in the set of "developed countries"). This requirement forces researchers to attend to the issue of calibration and provides additional motivation for them to explore the conceptual underpinnings of their measures. Fuzzy sets resonate with the measurement concerns of qualitative researchers, where the goal often is to distinguish between relevant and irrelevant variation--that is, to interpret it--and with the measurement concerns of quantitative researchers, where the goal is the precise placement of cases relative to each other.

The second half of this essay sketches a technique for calibrating conventional interval- and ratio-scale variables according to external standards. In the examples provided, the external standard used is a qualitative assessment of the degree to which cases with given scores on a conventional interval-scale measure are members of a target set. A simple estimation technique rescales the interval-scale measure so that it conforms to these qualitative assessments. The end product of this procedure is the calibration of the degree of membership of cases in sets, which in turn is suitable for fuzzy-set and other types of analysis. The examples illustrate the responsiveness of this calibration technique to the researcher's qualitative assessments of cases. While calibration in the social sciences is unlikely ever to match the sophistication of calibration in the physical sciences, the technique of qualitative calibration presented here is an important first step.

Common Measurement Practices In Quantitative Research

Measurement, as practiced in the social sciences today, remains relatively haphazard and unsystematic, despite the efforts and exhortations of many distinguished scholars (e.g., Duncan 1984; Pawson 1989). The dominant approach is the indicator approach, in which social scientists seek to identify the best possible empirical indicators of their theoretical concepts. For example,

national income per capita (in U.S. dollars, adjusted for differences in purchasing power) is often used as an empirical indicator of the theoretical concept of development, applied to countries. In the indicator approach the key requirement is that the indicator vary across cases, ordering them in a way that is consistent with the underlying concept. The values of national income per capita, for example, must distinguish less developed from more developed countries in a systematic manner.

In this approach fine gradations and equal measurement intervals are preferred to coarse distinctions and mere ordinal rankings. Indicators like income per capita are especially prized not only because they offer fine gradations (e.g., an income per capita value of $5,500 is $100 less than a value of $5,600), but also because the distance between two cases is considered the "same" regardless of whether it is the difference between $1,000 and $2,000 or between $21,000 and $22,000 (i.e., a $1,000 difference).[2] Such interval- and ratio-scale indicators are well-suited for the most widely used analytic techniques for assessing relationships between variables, such as multiple regression and related linear techniques.[3]

More sophisticated versions of the indicator model use multiple indicators and rely on psychometric theory (Nunnally and Bernstein 1994). The core idea in psychometric theory is that an index that is composed of multiple, correlated indicators of the same underlying concept is likely to be more accurate and more reliable than any single indicator. A simple example: national income per capita could easily overstate the level of development of oil-exporting countries, making them appear to be more developed than they "really are." Such anomalies challenge the face validity of income per capita as an indicator of the underlying concept.[4] However, using an index of development composed of multiple indicators (e.g., including such things as literacy, life expectancy, energy consumption, labor force composition, and so on) would address these

anomalies, because many oil-exporting countries have relatively lower scores on some of these alternate indicators of development. Ideally, the various indicators of an underlying concept should correlate very strongly with each other. If they do not, then they may be indicators of different underlying concepts (Nunnally and Bernstein 1994). Only cases with consistently high scores across all indicators obtain the highest scores on an index built from multiple indicators. Correspondingly, only those cases with consistently low scores across all indicators obtain the lowest scores on an index.

Perhaps the most sophisticated implementation of the indicator approach is through an analytic technique known as structural equation modeling (or "SEM"; see Bollen 1989). SEM extends the use of multiple indicators of a single concept (the basic psychometric model) to multiple concepts and their interrelationships. In essence, the construction of indexes from multiple indicators takes place within the context of an analysis of the interrelationships among concepts. Thus, index construction is adjusted in ways that optimize the hypothesized relationships among concepts. Using SEM, researchers can evaluate the coherence of their constructed indexes within the context of the model in which they are embedded. Simultaneously, they can evaluate the coherence of the model as a whole.

What all techniques in the "indicator" family share is a deep reliance upon observed variation, which in turn is almost always sample specific in its definition and construction. As mentioned previously, in the conventional approach the key requirement that an indicator must meet is that it must order cases in a way that reflects the underlying concept. It is important to point out that these orderings are entirely relative in nature. That is, cases are defined *relative to each other* in the distribution of scores on the indicator (i.e., as having "higher" versus "lower" scores). For example, if the U.S.'s national income per capita is $1,000 higher than Italy's, then

the U.S. correspondingly is considered *relatively* more developed. The greater the gap between countries, the more different their relative positions in the development hierarchy. Furthermore, the definition of "high" versus "low" scores is defined *relative to the observed distribution of scores*, usually conceived as a sample of scores drawn from a well-defined population. Thus, a case with a score that is above the sample's central tendency (usually the mean) has a "high" score; the greater this positive gap, the "higher" the score. Likewise, a case with a score that is below the mean has a "low" score; the greater this negative gap, the "lower" the score. Notice that the use of deviations from sample-specific measures of central tendency offers a crude but passive form of calibration. Its crudeness lies in the fact that the calibration standards (e.g., the mean and standard deviation) vary from one sample to the next and are inductively derived. By contrast, the routine practice in the physical sciences is to base calibration on external, dependably known standards.

At first glance, these conventional practices with respect to the use of indicators in the social sciences appear to be entirely straightforward and uncontroversial. It seems completely reasonable, for example, that countries should be ranked relative to each other and that some measure of central tendency, based on the sample or population in question, should be used to define "high" versus "low" scores. Again, the fundamental requirement of the indicator model is simply variation, which in turn requires only (1) a sample (or population) displaying a variety of scores and (2) a measure of central tendency based on the sample (or population). Note, however, that in this view all variation is considered equally *relevant*.[5] That is, variation in the entire range of the indicator is considered pertinent, with respect to what it reveals about the underlying concept. For example, the two countries at the very top of the income distribution are both "highly developed countries." Yet, the difference that separates them indicates that one is

5

still more "highly developed" than the other. In the indicator approach, this difference is usually taken at face value, meaning that there is usually no attempt to look at the cases and ask whether this difference--or any other difference, regardless of magnitude--is a relevant or meaningful difference with respect to the underlying concept.[6] The interpretation of scores relative to agreed upon standards is central to measurement calibration. These external standards provide a context for the interpretation of scores.

Common Measurement Practices in Qualitative Research

In conventional quantitative research, measures are indicators of concepts, which in turn are components of models, which in turn are derived from theories. Thus, the quantitative approach is strongly theory centered. Much qualitative research, by contrast, is more knowledge centered and thus tends to be more grounded in empirical evidence and also more "iterative" in nature--that is, there is an interplay between concept formation and measurement, on the one hand, and research strategy, on the other (see, e.g., Glaser and Strauss 1967). The researcher begins with orienting ideas and broad concepts and uses empirical cases to help refine and elaborate them (Becker 1958). This process of progressive refinement involves an iterative "back-and-forth" movement between ideas and evidence (Katz 1982; Ragin 1994). In this back-and-forth process, researchers specify and refine their empirical indicators and measures.

A simple example: macrolevel researchers often distinguish between countries that experience "early" versus "late" state formation (see, e.g., Rokkan 1975). Those that developed "early" had certain advantages over those that developed "late" and vice versa. David Laitin (1992:xi), for example, notes that coercive nation-building practices available earlier to monarchs (e.g., the draconian imposition of a national language) are not available to leaders of new states today, in part because of the international censure these policies might generate. But what is

"early" state formation? The occurrence of state formation, of course, can be dated. Thus, it is possible to develop a relatively precise ratio-scale measure of the "age" of a state. But most of the variation captured by this direct measure is not relevant to the concept of "early" versus "late" state formation. Suppose, for example, that one state has been around for 500 years and another for 250 years. The first is twice as old as the second, but both are fully "early" when viewed through the lens of accumulated substantive and theoretical knowledge about state formation. Thus, much of the variation captured by the ratio-scale indicator "age" is simply irrelevant to the distinction between "early" versus "late" state formation. "Age in years" must be adjusted on the basis of accumulated substantive knowledge in order to be able to calibrate "early" versus "late" in a way that resonates appropriately with existing theory.

Such calibrations are routine in qualitative work, even though they are rarely modeled or even stated explicitly. Indeed, from the perspective of conventional quantitative research, it appears that qualitative researchers skew their measurements to fit their preconceptions. In fact, however, the qualitative researcher's goal is simply to interpret "mere indicators" such as "age in years" in the light of knowledge about cases and the interests of the investigator (e.g., whether a state is "early" from the standpoint of state formation theory).

A second essential feature of measurement in qualitative research is that it is more case oriented than measurement in quantitative research. This observation goes well beyond the previous observation that qualitative researchers pay more attention to the details of cases. In case-oriented research, the conceptual focus is on specific *kinds* of cases, for example, "developed countries." In variable-oriented research, by contrast, the focus is on dimensions of variation in a defined sample or population of cases, for example, variation in level of development across currently constituted nation states. The distinction is subtle but important because cases can vary

not only along a given dimension, but also in how well they satisfy the requirements for membership in a category or set. For example, countries vary in how well they satisfy requirements for membership in the set of developed countries--some cases satisfy them fully, some partially, and some not at all. In order to assess how well cases satisfy membership requirements, it is necessary to invoke external standards, for example, what it takes for a country to be considered developed. Thus, in the case-oriented view, the key focus is on sets of cases, the members of which can be identified and studied individually (e.g., the "developed countries"). In the variable-oriented view, by contrast, cases are usually understood simply as sites for taking measurements (that is, they are often seem as mere "observations") which in turn provide the necessary raw material for studying relationships between variables, viewed as cross-case patterns.

It follows that the case-oriented view is more compatible with the idea that measures should be calibrated, for the focus is on the degree to which cases satisfy membership criteria, which in turn are usually externally determined, not inductively derived (e.g., the sample mean). These membership criteria must reflect agreed-upon standards; otherwise, the constitution of a category or set will be contested. In the variable-oriented view, the members of a population simply vary in the degree to which they express a given trait or phenomenon, and there is usually no special motivation for specifying the criteria for membership in a set or for identifying specific cases as instances. Thus, a key difference between the qualitative approach to measurement and the quantitative approach is that in the qualitative approach meaning is attached to or imposed upon specific measurements, for example what constitutes "early" state formation or what it takes to warrant designation as a developed country. In short, measurement in qualitative research is necessarily interpretive.

The sociologist Aaron Cicourel was an early proponent of the understanding of measurement advanced here. In his classic text *Method and Measurement in Sociology*, he (1964:24) argues that it is necessary to consider the three "media" through which social scientists develop categories and link them to observable properties of objects and events: language, cultural meaning, and the properties of measurement systems. In his view, the problem of establishing equivalence classes (like "democracies" or "developed countries") cannot be seen as independent from or separate from problems of language and cultural meaning. He (1964:33) argues:

> Viewing variables as quantitative because available data are expressed in numerical form or because it is considered more "scientific" does not provide a solution to the problems of measurement but avoids them in favor of measurement by fiat. Measurement by fiat is not a substitute for examining and re-examining the structure of our theories so that our observations, descriptions, and measures of the properties of social objects and events have a literal correspondence with what we believe to be the structure of social reality.

In simple terms Cicourel argues that measures and their properties must be evaluated in the context of both theoretical and substantive knowledge. The fact that social scientist may possess a ratio-scale indicator of a theoretical concept does not mean that this aspect of "social reality" has the mathematical properties of this type of scale.

Thus, in qualitative research, the idea that social scientists should use external standards to evaluate and interpret their measures has much greater currency than it does in conventional quantitative research. A key difference with quantitative research, however, is that measurement in qualitative research is often lacking in precision, and the context-sensitive and case-oriented way of measuring that is typical of qualitative research often appears haphazard and unscientific to

quantitative researchers.

Fuzzy Sets: A Bridge Between the Two Approaches

With fuzzy sets it is possible to have the best of both worlds, namely, the precision that is prized by quantitative researchers and the use of substantive knowledge to calibrate measures that is central to qualitative research. With fuzzy sets, precision comes in the form of quantitative assessments of degree of set membership, which can range from a score of 0.0 (full exclusion from a set) to 1.0 (full inclusion). For example, a country might have a membership score of 0.75 in the set of democracies, indicating that it is clearly more in this set than out, but still not fully in. Substantive knowledge provides the external criteria that make it possible to calibrate measures. This knowledge indicates what constitutes full membership, full nonmembership, and the point at which cases are more "in" a given set than "out." This value is known as the "cross-over point," with scores greater than 0.50 indicating that a case is more in than out of a given set (Ragin 2000; Smithson and Verkuilen 2006).

The external criteria that are used to calibrate measures and translate them into set membership scores may reflect standards based on social knowledge (e.g., the fact that twelve years of education constitutes an important educational threshold), collective social scientific knowledge (e.g., about variation in economic development and what it takes to be considered fully in the set of "developed" countries), or the researcher's own accumulated knowledge, derived from the in-depth study of cases. These external criteria must be stated explicitly, and they also must be applied systematically and transparently. This requirement separates the use of fuzzy sets from conventional qualitative work, where the standards often remain implicit.

Fuzzy sets are able to bridge quantitative and qualitative approaches to measurement because they are simultaneously qualitative and quantitative. Full membership and full non-

membership are qualitative states. In between these two qualitative states are varying degrees of membership ranging from "more out" (closer to 0.0) to "more in" (closer to 1.0) and divided by a third qualitative anchor, the cross-over point (0.5). They are also simultaneously qualitative and quantitative because they are both case-oriented and variable-oriented. They are case-oriented in their focus on sets and set membership. In case-oriented work, the identity of cases matters, as does the sets to which a case may belong (e.g., the set of democracies). Fuzzy sets are simultaneously variable-oriented in their allowance for degrees of membership and thus for fine-grained variation across cases.

The use of external criteria to calibrate fuzzy sets is the primary focus of the remainder of this paper. I focus specifically on situations where the researcher has a serviceable interval- or ratio-scale indicator of a concept and seeks to transform it into a well-calibrated fuzzy set.

Transforming Interval-Scale Variables Into Fuzzy Sets

Ideally, the calibration of degree of membership in a set should be based entirely on the researcher's substantive and theoretical knowledge. That is, the collective knowledge base of social scientists should provide the basis for the specification of precise, empirically grounded calibrations. For example, armed with an adequate knowledge of development, social scientists should be able to specify the per capita income level that signals "full membership" in the set of developed countries. Unfortunately, the social sciences are still in their infancy, and this knowledge base does not exist. Furthermore, the dominance of variable-oriented research, with its paramount focus on mean-centered variation and on covariation as the key to assessing relationships between case aspects, undermines interest in substantively based thresholds and benchmarks. While the problem of specifying of thresholds and benchmarks has not attracted the attention it deserves, it is not a daunting task. The primary requirement for useful calibration is

simply sustained attention to the substantive issues at hand (e.g., what constitutes full membership in the set of developed countries).

Despite the imperfections of the existing knowledge base, it is still possible to demonstrate techniques of calibration. All that is lacking is precise "agreed upon standards" for calibrating measures. To the extent possible, the calibrations presented here are based on the existing theoretical and substantive literature. Still, the focus is more on techniques of calibration than on the specific empirical benchmarks used to structure calibration.

The proposed techniques assume that researchers already have at their disposal conventional interval-scale indicators of their concepts, for example, per capita national income as an indicator of development. The techniques also assume that the underlying concept can be structured and labeled in set-theoretic terms, for example, "degree of membership in the set of developed countries." Notice that this labeling requirement moves the investigation in a decidedly case-oriented direction. "The set of developed countries" identifies specific countries, while "level of development" does not. The latter simply identifies a dimension of cross-national variation.

I present two methods of calibration. The "direct method" focuses on the three qualitative anchors that structure fuzzy sets: the threshold for full membership, the threshold for full nonmembership, and the cross-over point. The "indirect method," by contrast, uses regression techniques to estimate degree of set membership based on a six-value coding scheme. Both methods yield precise calibrations of set membership scores based on translations using either qualitative anchors (direct methods) or qualitative groupings (indirect method) and estimates of the log of the odds of full membership in a set. While this translation route--using estimates of the log odds of full membership--may seem roundabout, the value of the approach will become clear as the demonstration proceeds.

For now, consider Table 1 which shows the different metrics that are used in the demonstration that follows. The first column shows various verbal labels that can be attached to differing degrees of set membership, ranging from full nonmembership to full membership. The second column shows the degree of set membership linked to each verbal label. For convenience, degree of membership is rounded to 3 decimal places. The third column shows the odds of full membership that result from the transformation of the set membership scores (column 2) into the odds of full membership, using the following formula:

odds of membership = (degree of membership)/(1 - (degree of membership))

The last column shows the natural log of the odds reported in column 3. In effect, columns 2 through 4 are different representations of the same numerical values, using different metrics. For example, the membership score attached to "threshold of full membership" is 0.953. Converting it to an odds yields 20.09. Calculating the natural log of 20.09 yields a score of 3.0.[7]

Working in the metric of log odds is useful because this metric is completely symmetric around 0.0 (an odds of 50/50) and suffers neither floor nor ceiling effects. Thus, for example, if a calibration technique returns a value in the log of odds that is either a large positive number or large negative number, its translation to degree of membership stays within the 0.0 to 1.0 bounds, which is a core requirement of fuzzy membership scores. Both the direct and the indirect methods of calibration demonstrated in this paper use log odds as an intermediate metric in the translation of interval scales to set membership scores. The essential task of calibration is to transform interval-scale variables into the log odds metric in a way that respects the verbal labels shown in column 1 of Table 1.[8]

It is important to note that the set membership scores that result from these transformations (ranging from 0.0 to 1.0) are *not* probabilities, but instead should be seen simply

as transformations of interval scales into degree of membership in the target set. In essence, a fuzzy membership score attaches a *truth value*, not a probability, to a statement (for example, the statement that a country is in the set of developed countries). The difference between a truth value and a probability is easy to grasp, and it is surprising that so many scholars confuse the two. For example, the *truth value* of the statement "beer is a deadly poison" is perhaps about .05--that is, this statement is almost but not completely out of the set of true statements, and beer is consumed freely, without concern, by millions and millions of people every day. However, these same millions would be quite unlikely to consume a liquid that has a .05 *probability* of being a deadly poison, with death the outcome, on average, in one in twenty trials.

The Direct Method of Measurement Calibration

The starting point of any set calibration is clear specification of the target set. The focus of this demonstration is the set of developed countries, and the goal is to use per capita national income data to calibrate degree of membership in this set. Altogether, 136 countries are include in the demonstration; Table 2 presents data on 24 of these 136 countries, which were chosen to represent a wide range of national income values.

The direct method uses three important qualitative anchors to structure calibration: the threshold for full membership, the threshold for full nonmembership, and the cross-over point (see Ragin 2000). The cross-over point is the value of the interval-scale variable where there is maximum ambiguity as to whether a case is more in or more out of the target set. For the purpose of this demonstration, I use a per capita national income value of $5,000 as the cross-over point. An important step in the direct method of calibration is to calculate the deviations of raw scores (shown in column 1) from the chosen cross-over point ($5,000). These values are shown in column 2 of Table 2. Negative scores indicate that a case is more out than in the target

set, while positive scores signal that a case is more in than out.

For the threshold of full membership in the target set, I use a per capita national income value of $20,000, which is a deviation score of $15,000 (compare columns 1 and 2 of Table 2). This value corresponds to a set membership score of .95 and a log odds of 3.0. Thus, cases with national income per capita of $20,000 or greater (i.e., deviation scores of $15,000 or greater) are considered fully in the target set, with set membership scores ≥ .95 and log odds of membership ≥ 3.0. In the reverse direction, the threshold for full nonmembership in the target set is $2,500, which is a deviation score of -$2,500. This national income value corresponds to a set membership score of .05 and a log odds of -3.0. Thus, cases with national income per capita of $2,500 or lower (i.e., deviation scores of -$2,500 or lower) are considered fully out of the target set, with set membership scores ≤ .05 and log odds of membership ≤ -3.0.

Once these three values (the two thresholds and the cross-over point) have been selected, it is possible to calibrate degree of membership in the target set. The main task at this point is to translate the cross-over centered national income data (column 2) into the metric of log odds, utilizing the external criteria that have been operationalized in the three qualitative anchors. For deviation scores *above* the cross-over point, this translation can be accomplished by multiplying the relevant deviation scores (column 2 of Table 2) by the ratio of the log odds associated with the verbal label for the threshold of full membership (3.0) to the deviation score designated as the threshold of full membership (i.e., 20,000 - 5,000 = 15,000). This ratio is 3/15,000 or .0002. For deviation scores *below* the cross-over point, this translation can be accomplished by multiplying the relevant deviation scores (column 2 of Table 2) by the ratio of the log odds associated with the verbal label for the threshold of full nonmembership (-3.0) to the deviation score designated as the threshold of full nonmembership ($2,500 - $5,000 = -$2,500). This ratio is -3/-2500 or .0012.

These two scalars are shown in column 3, and the products of columns 2 and 3 are shown in column 4.[9] Thus, column 4 shows the translation of income deviation scores into the log odds metric, using the three qualitative anchors to structure the transformation via the two scalars.

The values in column 4, in effect, are per capita national income values that have been rescaled into values reflecting the log odds of membership in the set of developed countries, in a manner that strictly conforms to the values attached to the three qualitative anchors--the threshold of full membership, the threshold of full nonmembership, and the cross-over point. Thus, the values in column 4 are not mere mechanistic rescalings of national income, for they reflect the imposition of external criteria via the three qualitative anchors. The use of such external criteria is the hallmark of measurement calibration.

It is a small step from the log odds reported in column 4 to the degree of membership values reported in column 5. It is necessary simply to apply the standard formula for converting log odds to scores that range from 0.0 to 1.0, namely:

degree of membership = exp(log odds)/(1 + exp(log odds))

where "exp" represents the exponentiation of log odds to simple odds. Note that the membership values reported in the last column of Table 2 strictly conform to the distribution dictated by the three qualitative anchors. That is, the threshold for full membership (0.95) is pegged to an income per capita value of $20,000; the cross-over point (0.50) is pegged to an income of $5,000; and so on. For further illustration of the results of the direct method, consider Figure 1, which shows a plot of degree of membership in the set of developed countries against per capita national income, using data on all 136 countries included in this demonstration. As the plot shows, the line flattens as it approaches 0.0 (full nonmembership) and 1.0 (full membership), consistent with the conceptualization of degree of set membership. What the plot does not reveal is that most of the

world's countries are in the lower-left corner of the plot, with low national income and exclusion from the set of developed countries (i.e., set membership scores ≤ .05).

To illustrate the importance of external criteria to calibration, consider using the same national income data (column 1 of Table 2) to calibrate degree of membership in the set of countries that are "at least moderately developed." Because the definition of the target set has changed, so too must the three qualitative anchors. Appropriate anchors for the set of "at least moderately developed" countries are: a cross-over value of $2,500; a threshold of full membership value of $7,500 and a threshold of full nonmembership value of $1,000. The appropriate scalars in this example are 3/5000 for cases above the cross-over value, and -3/-1500 for cases below the cross-over value. The complete procedure is shown in Table 3, using the same cases as in Table 2.

The key point of contrast between Tables 2 and 3 is shown in the last column, the calibrated membership scores. For example, with a national income per capita of $2,980, Turkey has a membership of .08 in the set of developed countries. Its membership in the set of "at least moderately developed" countries, however, is 0.57, which places it just above the cross-over point. Notice, more generally, that in Table 3 there are many more cases that register set membership scores close to 1.0, consistent with the simple fact that more countries have high membership in the set of countries that are "at least moderately developed" than in the set of countries that are fully "developed." The contrast between Tables 2 and 3 underscores both the knowledge-dependent nature of calibration and the impact of applying different external standards to the same measure (per capita national income). Again, the key to understanding calibration is to grasp the importance of external criteria, which are based, in turn, on the substantive and theoretical knowledge that researchers bring to their research.

17

The Indirect Method of Measurement Calibration

In contrast to the direct method, which relies on precise specification of the numerical values linked to three qualitative anchors, the indirect method relies on the researcher's broad groupings of cases according to their degree of membership in a target set. Consider again the data on per capita national income, this time presented in Table 4. The first and most important step in the indirect method is to categorize cases in a qualitative manner, according to their presumed degree of membership in the target set. These qualitative groupings can be preliminary and open to revision. However, they should be based as much as possible on existing theoretical and substantive knowledge. The six key qualitative categories used in this demonstration are:[10]

(a) in the target set (log odds = 5),

(b) probably in the target set (log odds = 2),

(c) more in than out of the target set (log odds = .5),

(d) more out than in the target set (log odds = -.5),

(e) probably out of the target set log odds = -2), and

(f) out of the target set (log odds = -5).

These categorizations are shown in column 2 of Table 4, using explicit numerical values to reflect log odds of membership, as just presented (see also Table 1). These six numerical values are not arbitrary, of course, but chosen as rough estimates of the log odds of membership specific to each qualitative grouping. The goal of the indirect method is to re-scale an interval-scale indicator to reflect knowledge-based, qualitative groupings of cases, categorized according to degree of set membership. These qualitative interpretations of cases must be grounded in empirical knowledge. The stronger the empirical basis for making qualitative assessments of set membership, the more precise the calibration of the interval-scale indicator as set membership scores.

Note that the qualitative groupings implemented in Table 4 have been structured so that they utilize roughly the same criteria used to structure the calibrations shown in Table 2. That is, countries with national income per capita greater than $20,000 have been coded as fully in the set of developed countries; countries with income per capita greater than $5,000 have been coded as more in than out; and so on. By maintaining fidelity to the qualitative anchors used in Table 2, it is possible to compare the results of the two methods. The direct method utilizes precise specifications of the key benchmarks, while the indirect method requires only a broad classification of cases.

The next step is to use the two series reported in columns 1 and 2 of Table 4 to estimate the predicted qualitative coding of each case, using per capita national income as the independent variable and the qualitative codings as the dependent variable. Usually, some type of curvilinear estimation (e.g., cubic regression) is necessary in order to derive a good fit.[11] The purpose of this regression analysis is simply to rescale the independent variable (per capita national income, in this example) into the metric of the dependent variable (i.e., the numerical values attached to the qualitative codings). The predicted values resulting from a cubic regression analysis are reported in column 3 of Table 1. The reported values are based on a regression analysis using all 136 cases, not the subset of 24 presented in the table. The predicted values, in essence, constitute rough estimates of the log odds of membership in the target set (the set of developed countries) based on per capita national income and the qualitative analysis that produced the codings shown in column 2. For graphic illustration of the procedure consult Figure 2 which shows a plot of the results of the cubic regression. The plotted line shows the predicted qualitative codings using national income per capita as the predictor variable. As the figure illustrates, the use of a simple linear regression would produce a poor fit.

It is a simple matter to transform the predicted values (column 3) into membership scores by first exponentiating the predicted log odds, to turn them into simple odds, and then converting these odds into membership scores (as in Tables 2 and 3). Comparison of the set membership scores in column 5 of Tables 2 (direct method) and column 4 of Table 4 (indirect method) reveals great similarities, but also some important differences. First notice that Table 2 faithfully implements $20,000 as threshold for full membership in the set of developed countries (0.95). In Table 4, however, this threshold value drops below New Zealand's score ($13,680). Second, observe that even though the cross-over point appears to be roughly the same in both tables ($5,000), the membership score of the first case below the cross-over in the (Hungary's) is different. In Table 2, Hungary's membership score is 0.40, while in Table 4, it is 0.19. These differences, which arise despite the use of the same criteria, follow from the indirectness of the second method and its necessary reliance on regression estimation. Still, if researchers lack the precise external criteria demanded by the direct method, the comparison of Tables 2 and 4 confirms that the indirect method produces useful set membership scores.

Using Calibrated Measures

Calibrated measures have many uses. While they are specifically designed for set-theoretic analysis, they also can be utilized in standard multivariate analyses. For example, a researcher might want to calculate the exact degree of membership in the set of developed countries that is linked to the threshold value for full membership in the set of democratic countries (i.e., a predicted membership score of at least 0.95 in the set of democracies). This analysis would involve the regression of degree of membership in the set of democratic countries on degree of membership in the set of developed countries. Is very modest membership in the set of developed countries (e.g., a value of 0.25) linked to democracy, or must it be more substantial (e.g., a value

of 0.75)? Or must a country reach the threshold for full membership in the set of developed

countries (0.95) in order to achieve the same level of membership in the set of democracies?

Using the metric of set membership provides a more nuanced connection to empirical evidence

and thus a more direct connection to theory. By contrast, regression coefficients estimated on the

basis of conventional interval-scale indicators cannot directly provide this same assessment.

Consider the regression coefficient from the regression of a standard measure of democracy, the

polity index (which ranges from -10 to +10) on national income per capita. This coefficient

simply translates dollars of national income into points on an index of democracy, showing the

national income difference linked to a one-point difference in polity scores.

Of course, calibrated measures are even more useful for evaluating theory that is

formulated in terms of set relations. While some social science theory is strictly mathematical, the

vast majority of it is verbal. Verbal theory, in turn, is formulated almost entirely in terms of set

relations (Ragin 2000; 2006). Unfortunately, social scientists have been slow to recognize this

fact. Consider, for example, the statement that "the developed countries are democratic." As in

many statements of this type, the assertion is essentially that instances of the set mentioned first

(developed countries) constitute a *subset* of instances of the set mentioned second (democracies).

 (It is common in English to state the subset first, as in the statement "ravens are black.") Close

examination of most social science theories reveals that they are composed almost entirely of

statements describing set relations, such as the subset relation. These set relations, in turn, may

involve a variety of different types of empirical connections--descriptive, constitutive, or causal,

among others.

The set relation just described (with developed countries as a subset of democratic

countries) is also consistent with a specific type of causal argument, namely, that development is

sufficient but not necessary for democracy. In arguments of this type, if the cause (development) is present, then the outcome (democracy) also should be present. However, instances of the outcome (democracy) without the cause (development) do not count against or undermine the argument that development is sufficient for democracy (even though such cases dramatically undermine the correlation). Rather, these instances of the outcome without the cause in question are due to the existence of alternate routes or recipes for the same outcome (e.g., the imposition of a democratic form of government by a departing colonial power). Thus, in situations where instances of a causal condition constitute a subset of instances of the outcome, a researcher may claim that the cause is sufficient but not necessary for the outcome.[12]

Before the advent of fuzzy sets (Zadeh 1965, 1972, 2002; Lakoff 1973), many social scientists disdained the analysis of set-theoretic relations because such analyses required the use of categorical-scale variables (i.e., conventional binary or "crisp" sets), which in turn often necessitated the dichotomization of interval and ratio scales. For example, using crisp sets, in order to assess a set theoretic statement about developed countries, a researcher might be required to categorize countries into two groups, developed and not developed, using per capita national income. These practices are sometimes suspect because it is possible to manipulate breakpoints when dichotomizing interval- and ratio-scale variables in ways that enhance the consistency of the evidence with the set-theoretic claim in question. However, as demonstrated in this paper (and in Ragin 2000), it is possible to calibrate degree of membership in sets and thereby avoid arbitrary dichotomizations.

With fuzzy sets, the subset relation is established by demonstrating that membership scores in one set are consistently less than or equal to membership scores in another. In other words, if for every case degree of membership in set **X** is less than or equal to degree of membership in set

Y, then set **X** is a subset of set **Y**. Of course, social science data are rarely perfect and some allowance must be made for these imperfections. As Ragin (2006) demonstrates, it is possible to assess the *degree* of consistency of empirical evidence with the subset relation using the simple formula:

$$\text{Consistency } (\mathbf{X_i} \leq \mathbf{Y_i}) = \Sigma(\min(\mathbf{X_i}, \mathbf{Y_i}))/\Sigma(\mathbf{X_i})$$

where $\mathbf{X_i}$ is degree of membership in set **X**; $\mathbf{Y_i}$ is degree of membership in set **Y**; $(\mathbf{X_i} \leq \mathbf{Y_i})$ is the subset relation in question; and "min" dictates selection of the lower of the two scores.

For illustration, consider the consistency of the empirical evidence with the claim that the set of developed countries (as calibrated in Table 2) constitutes a subset of the set of democracies, using data on all 136 countries. For this demonstration, I use the Polity IV democracy/autocracy measure, which ranges from -10 to +10. (This measure is used because of its popularity, despite its many shortcomings. See, e.g., Goertz 2005: chapter 4.) The calibration of membership in the set of democracies, using the direct method, is shown in Table 5. Polity scores for 24 of the 136 countries included in the calibration are presented in column 1 of Table 5. These specific cases were selected in order to provide a range of polity scores. Column 2 shows deviations from the cross-over point (a polity score of 2), and the column 3 shows the scalars used to transform the polity deviation scores into the metric of log odds of membership in the set of democracies. The threshold of full membership in the set of democracies is a polity score of 9, yielding a scalar of 3/7 for cases above the cross-over point; the threshold of full nonmembership in the set of democracies is a polity score of -3, yielding a scalar of -3/-5 for cases below the cross-over point. Column 4 shows the product of the deviation scores and the scalars, while column 5 reports the calibrated membership scores, using the procedures previously described (see the discussion surrounding Table 2).

Applying the formula for set-theoretic consistency just described to all 136 countries, the consistency of the evidence with the argument that the set of developed countries constitute a subset of the set of democracies is 0.99. (1.0 indicates perfect consistency.) Likewise, the consistency of the evidence with the argument that the set of "at least moderately developed" countries (as calibrated in Table 3) constitutes a subset of the set of democratic countries is 0.95. In short, both subset relations are highly consistent, providing ample support for both statements ("developed countries are democratic" and "countries that are at least moderately developed are democratic"). Likewise, both analyses support the argument that development is sufficient but not necessary for democracy. Note, however, that the set of "at least moderately developed" countries is a much more inclusive set, with higher average membership scores than the set of "developed" countries. It thus offers a more demanding test of the underlying argument. The greater the average membership in a causal condition, the more difficult it is to satisfy the inequality indicating the subset relation ($X_i \leq Y_i$).[13] Thus, using set-theoretic methods it is possible to demonstrate that membership in the set of countries with a moderate level of development is sufficient for democracy; membership in the set of fully developed countries is not required.

It is difficult to evaluate set theoretic arguments using correlational methods. There are three main sources of this difficulty:

(1) Set theoretic statements are about kinds of cases, not about relationships between variables. The statement that developed countries are democratic (i.e., that they constitute a subset of democratic countries) invokes cases, not dimensions of cross-national variation. This focus on cases as instances of concepts follows directly from the set theoretic nature of social science theory. The computation of a correlation, by contrast, is premised on an interest in

assessing how well dimensions of variation parallel each other across a sample or population, not on an interest in cases, per se.

(2) Correlational arguments are fully symmetric, while set theoretic arguments are typically asymmetric. The correlation between development and democracy (treating both as conventional variables) is weakened by the fact that there are many less developed countries that are democratic, However, such cases do not challenge the set theoretic claim or weaken its consistency. The theoretical argument in question addresses the qualities of developed countries-- that they are democratic--and does not make specific claims about relative differences between less developed and more developed countries in their degree of democracy.

(3) Correlations are insensitive to the calibrations implemented by researchers. The contrast between Tables 2 and 3 is meaningful from a set theoretic point of view. The set represented in Table 3 is more inclusive and thus provides a more demanding set-theoretic test of the connection between development and democracy. From a correlational perspective, however, there is little difference between the two ways of representing development. Indeed, the Pearson correlation between fuzzy membership in the set of developed countries and fuzzy membership in the set of "at least moderately developed" countries is .911. Thus, from a strictly correlational viewpoint the difference between these two fuzzy sets is trivial. The insensitivity of correlation to calibration follows directly from the fact that correlation provides its own implicit, de facto calibration in its computational reliance on deviations from an inductively derived, sample-specific measure of central tendency--the mean.

Conclusion

Current practices in quantitative social science undercut serious attention to calibration. These difficulties stem from reliance on the "indicator approach" to measurement, which requires

25

only variation across sample points and typically treats all variation as meaningful. The limitations of the indicator approach are compounded and reinforced by correlational methods, which are insensitive to calibrations implemented by researchers. Reliance on deviations from the mean tends to neutralizes the impact of any direct calibration implemented by the researcher. A further difficulty arises when it is acknowledged that almost all social science theory is set theoretic in nature and that correlational methods do a poor job of assessing set theoretic relations.

The set theoretic nature of most social science theory is not generally recognized by social scientists today. In tandem with this recognition, social scientists must also recognize that the assessment of set theoretic arguments and set calibration go hand in hand. Set theoretic analysis without careful calibration of set membership is pointless. It follows that researchers need to be faithful to their theories, clarify the target sets that correspond to the concepts central to their theories, and develop useful external criteria that can be used to guide the calibration of set membership. The research strategy advocated here brings measurement calibration to the social sciences through the combination of greater theoretical clarity and deeper substantive knowledge of cases.

Endnotes

1. Perhaps the greatest calibration efforts have been exerted in the field of poverty research, where the effort to establish external standards has deep policy relevance. Another example of a calibrated measure is the Human Development Index developed by the United Nations and published in its *Human Development Report*. In economics, by contrast, calibration often has a different meaning. Researchers "calibrate" parameters in models by fixing them to particular values, so that the properties and behavior of other parameters in the model can be observed. This type of calibration is very different from the explicit calibration of measures, the central concern of this essay.

2. Actually, there is a world of difference between living in a country with a GNP per capita of $2,000 and living in one with a GNP per capita $1,000; however, there is virtually no difference between living in one with $22,000 and living in one with $21,000. Such fine points are rarely addressed by researchers who use the conventional indicator approach, but they must be confronted directly in research that uses fuzzy sets.

3. While most textbooks assert that ratio scales are the highest form of measurement because they are anchored by a meaningful zero point, it is important to note that fuzzy sets have three numerical anchors: 1.0 (full membership), 0.0 (full nonmembership), and 0.5 (the cross-over point separating "more in" versus "more out" of the set in question. See Ragin (2000).

4. Indeed, in then empirical analysis presented in the second half of this paper, the major oil-exporting countries are excluded from analyses that use per capita national income as an indicator of development.

5. Of course, researchers sometimes transform their variables (e.g., using logs) in order to reduce skew and shift the weight of the variation. However, such adjustments are relatively uncommon and, in any event, are usually understood mechanistically, as a way to improve robustness of a model.

6. Notice also that the idea that variation at either end of a distribution should be de-emphasized or truncated is usually viewed with suspicion because truncating variation tends to attenuate correlations.

7. The values shown for degree of membership in column 2 have been adjusted (e.g., using .993 instead of .99 for full membership) so that they correspond to simple, single-digit entries in column 4.

8. The procedures for calibrating fuzzy membership scores presented in this paper are mathematically incapable of producing set membership scores of exactly 1.0 or 0.0. These two membership scores would correspond to positive and negative infinity, respectively, for the predicted log of the odds. Instead, scores that are 0.95 or greater may be interpreted as full membership in the target set, and scores that are 0.05 or less may be interpreted as full nonmembership.

9. These two scalars constitute the slopes of the two lines extending from the origin (0,0) to the two threshold points (15000,3) and (-2500,-3) in the plot of the deviations of national income from the cross-over point (X axis) against the log odds of full membership in the set of developed countries (Y axis).

10. Of course, other coding schemes are possible, using as few as three qualitative categories. The important point is the scoring of these categories, which should reflect log odds of set membership. These qualitative assessments provide the foundation for set calibration.

11. In SPSS, for example, this procedure is accomplished by clicking analyze, regression, and then curve estimation; specifying the independent and dependent variables; and then clicking cubic, save, and predicted values.

12. As always, claims of this type cannot be based simply on the demonstration of the subset relation. Researchers should martial as much corroborating evidence as possible when making any type of causal claim.

13. The two statements differ substantially in their set theoretic "coverage." Coverage is a gauge of empirical importance or weight (Ragin 2006). It shows the proportion of the outcome membership scores (in this example, the set of democratic countries) that is "covered" by a causal condition. The coverage of "democratic" countries by "developed" countries is 0.35; however, the coverage of "democratic" countries by "at least moderately developed" countries is 0.52. These results indicate that the latter gives a much better account of degree of membership in the set of democratic countries.

References

Becker, Howard S. 1958. "Problems of Inference and Proof in Participant Observation."
American Sociological Review 23:652-60.

Bollen, Kenneth. 1989. *Structural Equations with Latent Variables*. New York: Wiley
Interscience.

Byrne, David. 2002. *Interpreting Quantitative Data*. London: Sage.

Cicourel, Aaron V. 1964. *Method and Measurement in Sociology*. New York: Free Press.

Duncan, Otis Dudley. 1984. *Notes on Social Measurement*. New York: Russell Sage Foundation.

Glaser, Barney and Anslem Strauss. 1967. *The Discovery of Grounded Theory: Strategies for
Qualitative Research*. New York: Weidenfeld and Nicholson.

Goertz, Gary. 2005. *Social Science Concepts: A User's Guide*. Princeton: Princeton University
Press.

Katz, Jack 1982. *Poor People's Laywers in Transition*. New Brunswick: Rutgers University
Press.

Laitin, David. 1992. *Language Repertiores and State Construction in Africa*. New York:
Cambridge University Press.

Lakoff, George. 1973. "Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts."
Journal of Philosophical Logic 2:458-508.

Nunnally, Jum and Ira Bernstein. 1994. *Psychometric Theory*. New York: McGraw Hill.

Pawson, Ray. 1989. *A Measure for Measures: A Manifesto for Empirical Sociology*. New York:
Routledge.

Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

Ragin, Charles C. 1994. *Constructing Social Research*. Thousand Oaks, CA: Pine Forge.

Ragin, Charles C. 2006. "Set Relations in Social Research: Evaluating Their Consistency and
Coverage." *Political Analysis* 14(3):291-310.

Rokkan, Stein. 1975. "Dimensions of State Formation and Nation Building: A Possible Paradigm
for Research on Variations Within Europe." In Tilly, Charles (ed.), *The Formation of
Nation States in Western Europe*. Princeton: Princeton University Press.

Smithson, Michael and Jay Verkuilen. 2006. *Fuzzy Set Theory*. Thousand Oaks, CA: Sage.

Zadeh, Lotfi. 1965. "Fuzzy Sets." *Information and Control* 8:338-353.

Zadeh, Lotfi. 1972. "A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges." *Journal of
Cybernetics*. 2(3): 4-34.

Zadeh, Lotfi. 2002. "From Computing with Numbers to Computing with Words". *Applied
Mathematics and Computer Science*. 12(3):307-32

Table 1: Mathematical Translations of Verbal Labels

1. Verbal label	2. Degree of membership	3. Associated odds	4. Log odds of full membership
Full membership	0.993	148.41	5.0
Threshold of full membership	0.953	20.09	3.0
Probably in	0.881	7.39	2.0
More in than out	0.622	1.65	0.5
Cross-over point	0.500	1.00	0.0
More out than in	0.378	0.61	-0.5
Probably out	0.119	0.14	-2.0
Threshold of full nonmembership	0.047	0.05	-3.0
Full nonmembership	0.007	0.01	-5.0

Table 2: Calibrating Degree of Membership in the Set of Developed Countries: Direct Method

Country	1. National income	2. Deviations from cross-over	3. Scalars	4. Product of 2 x 3	5. Degree of membership
Switzerland	40110	35110.00	.0002	7.02	1.00
United States	34400	29400.00	.0002	5.88	1.00
Netherlands	25200	20200.00	.0002	4.04	.98
Finland	24920	19920.00	.0002	3.98	.98
Australia	20060	15060.00	.0002	3.01	.95
Israel	17090	12090.00	.0002	2.42	.92
Spain	15320	10320.00	.0002	2.06	.89
New Zealand	13680	8680.00	.0002	1.74	.85
Cyprus	11720	6720.00	.0002	1.34	.79
Greece	11290	6290.00	.0002	1.26	.78
Portugal	10940	5940.00	.0002	1.19	.77
Korea, Rep	9800	4800.00	.0002	.96	.72
Argentina	7470	2470.00	.0002	.49	.62
Hungary	4670	-330.00	.0012	-.40	.40
Venezuela	4100	-900.00	.0012	-1.08	.25
Estonia	4070	-930.00	.0012	-1.12	.25
Panama	3740	-1260.00	.0012	-1.51	.18
Mauritius	3690	-1310.00	.0012	-1.57	.17
Brazil	3590	-1410.00	.0012	-1.69	.16
Turkey	2980	-2020.00	.0012	-2.42	.08
Bolivia	1000	-4000.00	.0012	-4.80	.01
Cote d'Ivoire	650	-4350.00	.0012	-5.22	.01
Senegal	450	-4550.00	.0012	-5.46	.00
Burundi	110	-4890.00	.0012	-5.87	.00

Table 3: Calibrating Degree of Membership in the Set of "Moderately" Developed Countries: Direct Method

Country	1. National income	2. Deviations from cross-over	3. Scalars	4. Product of 2 x 3	5. Degree of membership
Switzerland	40110	37610	.0006	22.57	1.00
United States	34400	31900	.0006	19.14	1.00
Netherlands	25200	22700	.0006	13.62	1.00
Finland	24920	22420	.0006	13.45	1.00
Australia	20060	17560	.0006	10.54	1.00
Israel	17090	14590	.0006	8.75	1.00
Spain	15320	12820	.0006	7.69	1.00
New Zealand	13680	11180	.0006	6.71	1.00
Cyprus	11720	9220	.0006	5.53	1.00
Greece	11290	8790	.0006	5.27	.99
Portugal	10940	8440	.0006	5.06	.99
Korea, Rep	9800	7300	.0006	4.38	.99
Argentina	7470	4970	.0006	2.98	.95
Hungary	4670	2170	.0006	1.30	.79
Venezuela	4100	1600	.0006	.96	.72
Estonia	4070	1570	.0006	.94	.72
Panama	3740	1240	.0006	.74	.68
Mauritius	3690	1190	.0006	.71	.67
Brazil	3590	1090	.0006	.65	.66
Turkey	2980	480	.0006	.29	.57
Bolivia	1000	-1500	.0020	-3.00	.05
Cote d'Ivoire	650	-1850	.0020	-3.70	.02
Senegal	450	-2050	.0020	-4.10	.02
Burundi	110	-2390	.0020	-4.78	.01

Table 4: Calibrating Degree of Membership in the Set of Developed Countries: Indirect Method

Country	1. National income	2. Qualitative coding	3. Predicted Coding	4. Degree of membership
Switzerland	40110	5.00	5.29756	1.00
United States	34400	5.00	4.95689	.99
Netherlands	25200	5.00	4.85464	.99
Finland	24920	5.00	4.84512	.99
Australia	20060	5.00	4.48582	.99
Israel	17090	2.00	4.01521	.98
Spain	15320	2.00	3.61410	.97
New Zealand	13680	2.00	3.14872	.96
Cyprus	11720	2.00	2.46013	.92
Greece	11290	2.00	2.28823	.91
Portugal	10940	2.00	2.14250	.89
Korea, Rep	9800	.50	1.63045	.84
Argentina	7470	.50	.39440	.60
Hungary	4670	-.50	-1.46123	.19
Venezuela	4100	-.50	-1.89258	.13
Estonia	4070	-.50	-1.91581	.13
Panama	3740	-2.00	-2.17482	.10
Mauritius	3690	-2.00	-2.21463	.10
Brazil	3590	-2.00	-2.29469	.09
Turkey	2980	-2.00	-2.79614	.06
Bolivia	1000	-5.00	-4.58348	.01
Cote d'Ivoire	650	-5.00	-4.92571	.01
Senegal	450	-5.00	-5.12492	.01
Burundi	110	-5.00	-5.46974	.00

[*]5.0 = fully in the target set; 2.0 = probably in the target set; 0.5 = more in than out of the target set; -0.5 more out than in the target set; -2.0 probably out of the target set; -5.0 = fully out of the target set.

Table 5: Calibrating Degree of Membership in the Set of Democratic Countries: Direct Method

Country	1. Polity score	2. Deviations from cross-over	3. Scalars	4. Product of 2 x 3	5. Degree of Membership
Norway	10	8.00	0.43	3.43	0.97
United States	10	8.00	0.43	3.43	0.97
France	9	7.00	0.43	3.00	0.95
Korea, Rep	8	6.00	0.43	2.57	0.93
Colombia	7	5.00	0.43	2.14	0.89
Croatia	7	5.00	0.43	2.14	0.89
Bangladesh	6	4.00	0.43	1.71	0.85
Ecuador	6	4.00	0.43	1.71	0.85
Albania	5	3.00	0.43	1.29	0.78
Armenia	5	3.00	0.43	1.29	0.78
Nigeria	4	2.00	0.43	0.86	0.70
Malaysia	3	1.00	0.43	0.43	0.61
Cambodia	2	0.00	0.60	0.00	0.50
Tanzania	2	0.00	0.60	0.00	0.50
Zambia	1	-1.00	0.60	-0.60	0.35
Liberia	0	-2.00	0.60	-1.20	0.23
Tajikistan	-1	-3.00	0.60	-1.80	0.14
Jordan	-2	-4.00	0.60	-2.40	0.08
Algeria	-3	-5.00	0.60	-3.00	0.05
Rwanda	-4	-6.00	0.60	-3.60	0.03
Gambia	-5	-7.00	0.60	-4.20	0.01
Egypt	-6	-8.00	0.60	-4.80	0.01
Azerbaijan	-7	-9.00	0.60	-5.40	0.00
Bhutan	-8	-10.00	0.60	-6.00	0.00

Figure 1: Plot of Degree of Membership in the Set of Developed Countries Against National Income Per Capita: Direct Method
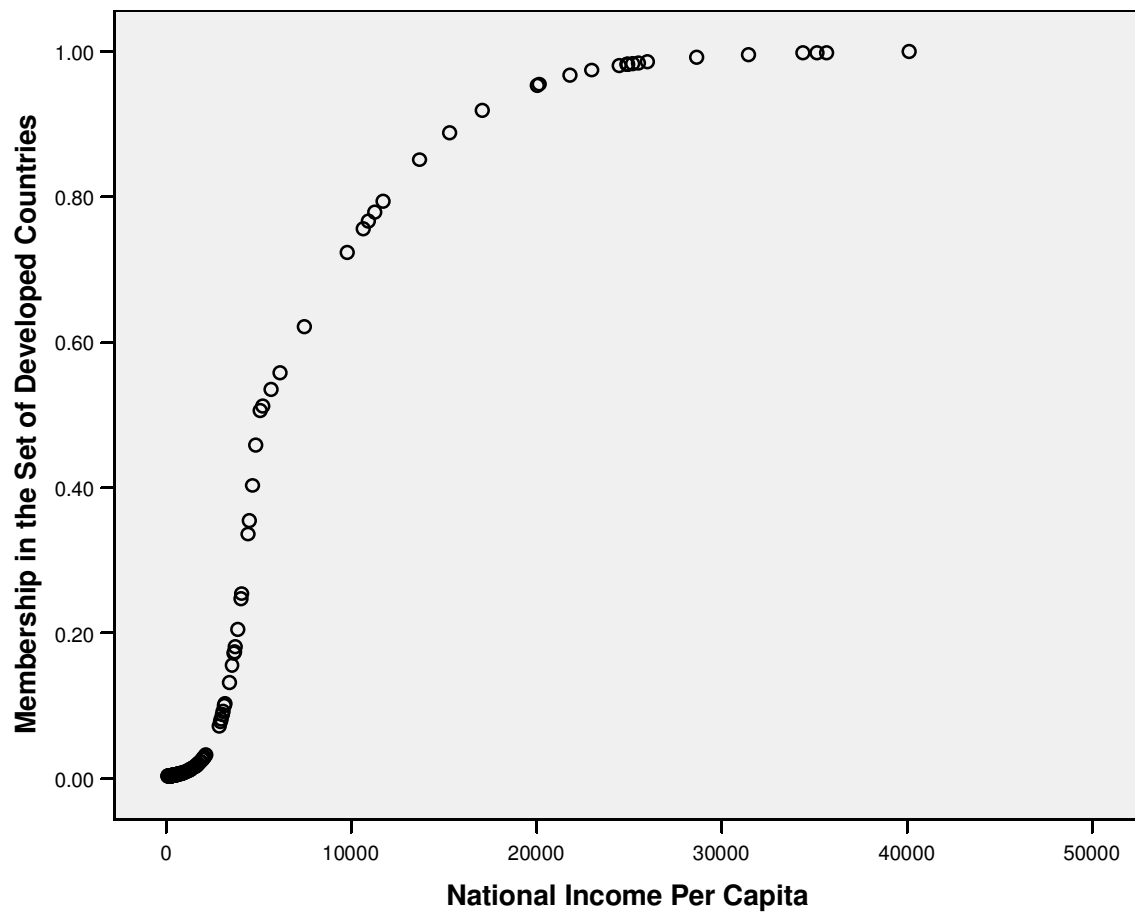
Figure 2: Using Cubic Regression Analysis to Predict Qualitative Codings of Membership in the Set of Developed Countries: Indirect Method