

Levels of Measurement and Political Research: An Optimistic View

Author(s): William G. Jacoby

Source: *American Journal of Political Science*, Vol. 43, No. 1 (Jan., 1999), pp. 271-301

Published by: Midwest Political Science Association

Stable URL: <http://www.jstor.org/stable/2991794>

Accessed: 23-01-2018 16:16 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

Midwest Political Science Association is collaborating with JSTOR to digitize, preserve and extend access to *American Journal of Political Science*

Levels of Measurement and Political Research: An Optimistic View

William G. Jacoby, *University of South Carolina*

This paper argues that all measurement is theory testing. Therefore, measurement always constitutes a tentative statement about the nature of reality. Some general implications of this perspective are discussed. Alternating Least Squares, Optimal Scaling (ALSOS) is then presented as a general strategy for obtaining *empirical* information about the measurement properties that exist within a given data set. The great advantage of this approach is that characteristics such as the levels of measurement associated with particular variables are viewed as testable hypotheses rather than a priori assumptions. An ALSOS regression analysis is performed on data from the CPS 1992 National Election Study. The results show that several variables which are usually interpreted as “pseudo-interval” measures actually only provide ordinal-level information. More generally, levels of measurement are important because they affect the degree of ambiguity in researchers’ interpretations of variability in empirical data.

1. INTRODUCTION

Science and measurement are closely intertwined with each other. In any discipline, scientific progress is strictly limited by the capacity to measure relevant concepts. Indeed, some investigators contend that the vitality of a scientific field can be judged by examining the quality of its measurement (Kuhn 1977). From this perspective, political science may seem to be in a very troublesome position, since we often bemoan the poor quality of the data that we are forced to use for our analyses (e.g., Wilson 1971). In this essay, I will show that the situation may not be as hopeless as it initially appears. In fact, some of the apparent problems in social scientific measurement stem more from rigid adherence to a set of oversimplified, incomplete measurement principles than from any fundamental deficiencies in our measures themselves. A much more optimistic conclusion can be achieved when we place the nature and characteristics of measurement within the overall context of attempting to model social and political phenomena.

Generally speaking, political scientists tend to think of measurement as the starting point of an empirical analysis. That is, measurement characteristics constitute a fixed and immutable property of the data. An alternative

Special thanks go to Sandra K. Schneider for her advice, assistance, and encouragement on this project. The paper could not have been completed without her help. I would also like to thank Kenneth Meier, William Mishler, Harvey Starr, Mark Tompkins, and the anonymous *AJPS* referees for their comments on earlier drafts of the manuscript.

American Journal of Political Science, Vol. 43, No. 1, January 1999, Pp. 271–301 ©1999 by the Board of Regents of the University of Wisconsin System

perspective—the one that I advocate in the following discussion—holds that measurement is always a theory about one's observations. In other words, measurement constitutes a proposition about the ways that numerical scores reflect substantively interesting properties of data. The great advantage of this approach is that characteristics such as the levels of measurement associated with particular variables are viewed as testable hypotheses about one's data, rather than a priori assumptions about the observations under investigation. I contend that this is a constructive and useful view of measurement and its role in modern empirical analyses. It encourages greater flexibility in our approaches to research and thereby may facilitate new findings and insights about interesting political phenomena.

2. MEASUREMENT AS THEORY TESTING

Measurement always involves constructing a formal model of empirical data. The components of the formal model are the real number set and some or all of the properties of the real number system. Measurement occurs when the researcher applies this abstract model to an attribute or property associated with the set of empirical objects (Lingoes 1979). The general process of comparing abstract models against the empirical world is the central focus of scientific theory construction (Kaplan 1964). This is the case with measurement, just as it is with any other kind of modeling endeavor. Therefore, *all measurement is theory testing*.

The preceding interpretation is fully consistent with the more common view that measurement is the process of assigning numbers to objects in meaningful ways (Stevens 1951). Measurement always begins by placing a set of empirical objects into mutually exclusive and jointly exhaustive categories, based upon some discernible attribute of the objects. This classification task becomes measurement when numbers are assigned to the categories, and by implication, to the objects within the categories. The differences between the numbers assigned to the objects must reflect, in some specified way, the differences between the objects with respect to the attribute. In this sense, measurement invariably entails examining the goodness of fit between an abstract model and a property of a set of empirical objects (Coombs, Raiffa, and Thrall 1954).

As an example, consider the well known index of party identification (e.g., Campbell et al. 1960). This variable plays an extremely prominent role in the study of political behavior. It is obtained by sorting individual survey respondents into seven categories, based upon their answers to a series of questions. The categories are: strong Democrat, Democrat, independent leaning toward the Democrats, independent, independent leaning toward the Republicans, Republican, and strong Republican. Measurement takes place when the categories are assigned numeric values—for example, integers

from zero through six. Differences in the resultant numeric scores presumably reflect differences in the psychological attachments that people feel toward one or the other major political party.

It is important to emphasize that this numerical scoring scheme implies a specific set of assertions about the nature of the differences among the partisan categories—in other words, a theory. If the analyst interprets the variability in the numbers assigned to the categories fairly literally (as is often done in the research literature) then the theory maintains the following: Democratic and Republican identifications are, in some sense, opposites of each other; independence is a psychological state that exists “in between” the two partisan directions; non-strong partisanship is still more partisan than independence leaning toward a party; and the amount of difference in psychological feelings between adjacent categories is constant for each pair of categories across the range of the scale. Each time the party identification index is employed in a substantive analysis, it constitutes a test of this theory; an opportunity to see whether the successive-integer scoring scheme truly reflects the differences that can be observed across the individual survey respondents placed within the seven different categories.

The interpretation that measurement is theory testing has been recognized by philosophers (Carnap 1966), mathematicians (Kline 1985), statisticians (Velleman and Wilkinson 1993), sociologists (Carter 1971; Duncan 1984), and particularly psychologists (Torgerson 1958; Young 1987; Cliff 1993). However, it does not seem to have surfaced within political science, even in works that have been explicitly concerned with measurement problems (e.g., Achen 1983; King, Keohane, and Verba 1994). Nevertheless, measurement must be recognized as a positive step on the part of the researcher (Porter 1995). No measurement whatsoever is “natural,” “pre-ordained,” or exists prior to/apart from human interpretation. Just like any other theory, measurement is always a tentative and potentially falsifiable statement about the nature of reality. Falsification occurs whenever the specified properties of the real number system do not correspond to the empirical property under investigation.

In everyday practice, researchers rarely test their measurement theories. Data values are simply taken as given quantities. This perspective has detrimental consequences for testing *substantive* theories. Consider the following scenario: A researcher develops a model to account for some form of political behavior and generates testable hypotheses that can be compared against empirical data. Upon carrying out the tests, the model-based predictions are not supported. The usual conclusion drawn in this situation is that some aspect of the structural model is inappropriate. However, it is important to remember that the data values used in the empirical tests are themselves models of the original observations. And, these measurements

could easily be inconsistent with their respective “segments” of the observable world.

Once again, party identification provides a ready illustration of this problem. A few years ago, several studies reported unexpected empirical results, such as “intransitive” relationships with other criterion variables. In other words, independent leaners appeared to behave in more clearly partisan ways than nonstrong party identifiers (Petrocik 1974). This led to fairly widespread questioning about the meaning of the party identification concept itself. Specifically, independence was viewed as a separate psychological orientation, distinct from individual ties to partisan reference groups. This in turn fostered multidimensional conceptualizations of party identification (Weisberg 1980; Jacoby 1982). More recent evidence suggests that the apparently anomalous findings which led to these “revisionist” conceptualizations were actually due to measurement errors in the seven-point index of party identification rather than to fundamental deficiencies in our understanding about the nature of psychological partisan attachments (e.g., Green 1988; Keith et al. 1992). Thus, measurement effects and structural effects are so closely intertwined as to be nearly indistinguishable—a point that is recognized by virtually every analyst who considers carefully the role of measurement in scientific research (Cliff 1993).

3. LEVELS OF MEASUREMENT

Measurement is a broad topic, encompassing several more specific characteristics of empirical data. But one measurement characteristic has clearly received a great deal of attention (and sometimes caused high levels of consternation) among political scientists: This is the *levels of measurement* (or alternatively, *scales of measurement*) associated with the variables we employ in empirical models. Levels of measurement represent variations in the ways that numbers are assigned to objects. Potentially, there are many different measurement levels. Only a handful, however, seem to receive much attention (Coombs, Raiffa, and Thrall 1954; Young 1987). These are the four familiar levels of measurement—nominal, ordinal, interval, and ratio—that were originally identified by S. S. Stevens in a series of influential articles published from the 1940s through the 1960s (Stevens 1946, 1951, 1968). Measurement levels are usually regarded as fixed characteristics of data, but this is an excessively restrictive interpretation. From the perspective of measurement as theory testing, it is more useful to regard levels of measurement as hypotheses to be supported or refuted on the basis of empirical evidence.

3.1 Measurement Levels as Functions

For present purposes, it is convenient to view measurement levels in terms of a function mapping empirical observations into numerical values

(Young 1987). We will denote this function as f^q , where the superscript is used to designate a specific measurement level. Assume that S is a set of objects, with members s_1, s_2, \dots, s_n . Within S , objects are placed into categories, based upon their observed equivalences or differences with respect to some observed attribute or property. M is a subset of the real numbers, the members of which are placed in a one-to-one relation with the categories containing the members of S ; by implication, the numbers can be assigned to the individual members of S that are contained within the categories. So, $M(s_1)$ is the number assigned to the category containing s_1 , $M(s_2)$ is the number assigned to the category containing s_2 , and so on, for the entire set of n objects. The function f^q maps the observational categories and, therefore, the elements of S , into M . The nature of f^q determines the level of measurement.

At the nominal level, the function simply preserves the identity of the categories in the sense that all objects within a given category are assigned the same number, while those placed into different categories receive different numbers:

$$f^n: \quad s_i \approx s_j \Rightarrow M(s_i) = M(s_j) \quad (1a)$$

$$s_i \neq s_j \Rightarrow M(s_i) \neq M(s_j) \quad (1b)$$

Note that the symbol “ \approx ” on the left-hand side of Equation 1a indicates empirical equivalence in terms of the measured attribute. Therefore, the relation $s_i \approx s_j$ means that these two objects are indistinguishable with respect to the attribute under investigation. Thus they are placed into the same observational category. But, these two objects are not, in any numerical sense, “equal” to each other. Similarly, the “ \neq ” on the left-hand side of Equation 1b means that the objects differ and are placed into different observational categories. In contrast, the “=” and “ \neq ” symbols on the right-hand sides do indicate the mathematical relationships of equality and inequality, because it is meaningful to compare the values of two numbers— $M(s_i)$ and $M(s_j)$ in this case—to each other. Beyond the fairly simple restrictions implied by Equations 1a and 1b, the function f^n places no limitations on the way that objects are mapped into the real number system.

The ordinal level of measurement retains the equality condition for equivalent objects that was used at the nominal level. But among objects that differ with respect to the attribute under examination, there is also an order restriction placed on the elements of M . Accordingly, the function for an ordinal measure is defined as follows:

$$f^o: \quad s_i \approx s_j \Rightarrow M(s_i) = M(s_j) \quad (2a)$$

$$s_i \succ s_j \Rightarrow M(s_i) \geq M(s_j) \quad (2b)$$

In this case, the symbol “ \succ ” on the left-hand side of Equation 2b refers to empirical asymmetry. The latter relation is assumed to exist among any pair of objects in S that differ with respect to the observational attribute. Thus, $s_i \succ s_j$ indicates that objects s_i and s_j have been placed into categories that are asymmetric with respect to the specified attribute. It is important to emphasize that empirical asymmetry need not involve any sort of quantity, magnitude, or ordering among the categories into which the observations have been placed. In this context, the existence of asymmetry merely implies that if $s_i \succ s_j$ then it cannot be the case that $s_j \succ s_i$ for any specified pair of objects in S . Again, however, the symbols “ $=$ ” and “ \geq ” on the right-hand sides of Equations 2a and 2b do correspond to the mathematical relations of equality and weak inequality (i.e., “greater than or equal to”), respectively. So, f^0 does reflect the perceived differences among the observational categories in the identities and ordering of the numbers that are assigned to those categories.

Interval measurement implies that there is a specific type of numeric function mapping S into M . One straightforward example specifies a linear relationship between the observational attribute and the numeric values:

$$f^i: \quad M(s_i) = \beta_0 + \beta_1(s_i) \quad (3)$$

This particular functional form is convenient for interpretation, because differences in the numbers assigned to two objects will be proportional to the amount of difference that exists between those objects with respect to the observational attribute. However, interval measurement occurs whenever a specific type of numeric function—which could be nonlinear, as well as linear—is used to represent the correspondence between properties of empirical objects and real numbers.¹

Finally, the ratio level of measurement is similar to the interval level in the sense that it requires a specific numeric function mapping from S into M . However, it also includes an additional restriction as follows:

$$f^r: \quad \beta_0 = 0 \quad (4)$$

Equation 4 unambiguously equates a particular observational category of the empirical objects to the numeric value of zero. Stated differently, ratio-level measurement implies that the location of the numeric origin is meaningful in substantive terms.

¹For example, nonlinear measurement functions occur frequently in psychophysics, where *perceived* magnitudes of physical stimuli are related to *actual* stimulus magnitudes through a power function which is specific to the type of stimulus under investigation (e.g., Stevens 1975).

In summary, all levels of measurement can be represented quite succinctly as functions.² This interpretation is quite different from the traditional approach to measurement and operationalization, which focuses on maximizing epistemic correlations (Blalock 1982). The analytic methods for addressing this concern, however, are almost always based upon an implicit *linear* model of the relationship between unobserved constructs and empirical variables (e.g., Bollen 1989). This is perfectly appropriate for dealing with the problem of measurement *error*.³ For measurement *level*, however, it is important to focus more directly on the *functional form* of the relationship between concept and empirical indicator. Doing so provides a more complete and accurate picture of the information that is obtained when a substantive attribute is expressed as a subset of the real number system—that is, when it is *measured*.

3.2 Measurement Levels are Determined by the Analyst

The preceding discussion about levels of measurement begs a fairly fundamental question: Who or what determines the measurement level associated with any particular variable? This is definitely not a trivial question, but the answer is almost disappointingly simple: The researcher always decides upon the level of measurement for each of the variables that he/she employs in any particular empirical analysis (Young 1987; Cliff 1993; Velleman and Wilkinson 1993).

The nature of the function, f^q , mapping from empirical objects into the real numbers is usually unknown. It depends entirely upon the researcher's interpretation of the differences among the observational categories into which the empirical objects are divided. If the analyst decides that it is only possible to make qualitative distinctions among the objects, then the measurement level is nominal. Or, if he/she thinks the observational categories can be ranked with respect to some interesting property of the objects, then the level is ordinal, and so on. *The researcher* makes this determination and not anyone or anything else.

²Note that this discussion has not addressed the distinction between discrete and continuous variables. According to Young (1981) this is completely separate from a variable's measurement level. Discrete measurement occurs when all of the objects within a given observational category must be assigned exactly the same numerical value. With continuous measurement, objects within a given category can be assigned different values, so long as these values do not overlap with values assigned to other observational categories.

³Measurement errors are conceptualized as imperfections in the functional translation from empirical object attribute to assigned numerical value; analytically, they can be represented by including disturbance terms in conditions 1a through 4. Note that the presence of measurement error does not compromise the existence of a particular measurement level. The latter is determined entirely by the nature of the function from object set S into number set M . Measurement errors merely imply that this function will not be deterministic in nature.

This idea has at least three important, but rather subtle, implications. First, two separate analysts might attribute different measurement levels to exactly the same variable. Neither one is “correct” or “incorrect,” so long as they can both justify their measurement assumptions in terms of their respective research contexts and objectives. This situation occurs very frequently in political science. As an example, consider once again the familiar seven-point index of party identification. What is its level of measurement? A perusal of the literature reveals that this variable has been treated as an ordinal (e.g., Fiorina 1981), interval (e.g., Markus and Converse 1979), and even nominal (e.g., Keith et al. 1992; Valentine and Van Wingen 1980) measure by different researchers. Each of these treatments was fully justifiable within the context of the respective studies; the party identification index helped generate interesting, useful, and substantively interpretable results in each case.

The second important implication is that the level of measurement is not in any way determined by the nature of the variable itself. Instead, it is based strictly upon the researcher’s own theory about the differences among the observational categories. This point is often overlooked, frequently with detrimental consequences. Consider an example originally discussed by Carter (1971): The variable “years of formal schooling” is commonly employed as an indicator of social status. Superficially, this may appear to be an interval-level variable, since there is a standard increment of one year between adjacent values on this variable. But if education is used as a surrogate for status, then it probably only provides ordinal information: After all, the difference between two and three years of school attendance is far less than, say, that between eleven and twelve years (i.e., high school graduation) or fifteen and sixteen years (i.e., college graduation) *with respect to social status* even though the numerical difference is identical in each case. Thus, the empirical measure is inconsistent with the theory that leads to the use of that variable in the first place; and, one could argue that years of education do not comprise a valid, *interval-level* indicator of social status, despite their intrinsically numeric nature.

The third implication is that the traditional fourfold typology does not, in any way, exhaust all possible variations in the ways that numbers can be assigned to objects. Nor does familiarity and ubiquity in the literature confer any special status on the nominal-ordinal-interval-ratio hierarchy. Instead, it is often helpful to consider the measurement characteristics of one’s observations without the limitations that stem from the usual constraints. As an example, consider the variable “religious affiliation.” This is almost always treated as a nominal-level, categorical variable. However, one could easily rank order at least some specific denominations in terms of their degree of dogmatism (e.g., atheist, followed by Jewish, mainstream Protestant, Roman

Catholic, and then fundamentalist Protestants), thereby generating an ordinal-level measure. If the resultant variable is correlated with other interesting phenomena (e.g., attitudes on social issues such as abortion), then the ordering is justifiable in terms of the research objective. The important point is that the level of measurement, like measurement itself, is entirely a creation of the researcher. It depends solely upon the way that the analyst chooses to interpret the numbers that have been assigned to the objects under investigation.

The subjectivity involved in specifying the levels of measurement provides researchers with opportunities to try creative interpretations in order to discern structures within data that might otherwise be hidden. The traditional levels of measurement, however, provide no means for doing so. The analyst is then left with a difficult dilemma: Either assuming the higher level of measurement and running the risk of generating meaningless conclusions or assuming the lower level of measurement and thereby producing results that are more limited than they need to be. A more flexible view of measurement levels allows for much finer distinctions. This, in turn, reduces the magnitude of the dilemma confronting the researcher, even if it does not eliminate it entirely.

3.3 Measurement Levels and Empirical Transformations

The definition of measurement level as a function relating observational categories to real numbers is of limited practical utility on its own, since the "true" nature of f^q is usually unknown. However, the measurement function for a variable has empirical implications, as shown by the following theorem:

A given measure, M_1 , can be arbitrarily transformed into any other measure, M_2 , so long as the function relating M_1 to M_2 is the same type of function (identity-preserving, monotone, or specific numeric) as that which is used to map the elements of S into M_1 .⁴

⁴The proof is quite simple. Consider a set of objects measured at the nominal level. According to expressions 1a and 1b, this is defined as $s_i \approx s_j \Rightarrow M_1(s_i) = M_1(s_j)$ and $s_i \neq s_j \Rightarrow M_1(s_i) \neq M_1(s_j)$. Now, define another function, f^{n*} , such that the following conditions hold:

$$f^{n*}: M_1(s_i) \approx M_1(s_j) \Rightarrow M_2\{M_1(s_i)\} = M_2\{M_1(s_j)\} \quad (1a^*)$$

$$M_1(s_i) \neq M_1(s_j) \Rightarrow M_2\{M_1(s_i)\} \neq M_2\{M_1(s_j)\} \quad (1b^*)$$

By the composition of a function and by transitivity, expressions 1a and 1a* can be combined to produce the following: $s_i \approx s_j \Rightarrow M_2(s_i) = M_2(s_j)$. Similarly, expressions 1b and 1b* can be combined to produce: $s_i \neq s_j \Rightarrow M_2(s_i) \neq M_2(s_j)$. And as is easily seen, these latter two expressions are equivalent to f^n , the function that defines nominal measurement. Similar arguments can be made for the ordinal, interval, and ratio levels of measurement.

For present purposes, the most important implication of this theorem is the fact that *empirical* transformations of a variable's values can be used to gain insights about the appropriate level of measurement for that variable. Specifically, the measurement level is determined by the least restrictive transformation that can be applied to a set of data values without violating the empirical representation of the object attribute being measured.

If a variable can only be transformed linearly, then it can be viewed as an interval-level representation. But, if the values can be subjected to an arbitrary monotonic transformation without altering the interpretation of the information contained in the variable, then the latter can only be regarded as an ordinal measure. Similarly, if a variable's information is invariant under an arbitrary nonmonotone transformation, then it cannot be measured at anything more than the nominal level.

Thus, measurement level issues can be embedded within the topic of data transformation which pertains to the functional form of an empirical model. But, the question remains, how can one determine the nature of the appropriate empirical transformation in the first place? The answer depends upon the interaction between measured values and the statistical models that are applied to them.

4. MEASUREMENT CHARACTERISTICS AND STATISTICAL MODELS

Statistical models provide formal representations of structure among observations *given* a particular set of data values (Belsley, Kuh, and Welsch 1980). Traditional approaches to statistical analysis take the numbers assigned to the observations as a fixed set of assumptions that cannot be altered during the course of the analysis, but this rigid interpretation of the numeric values ignores the fact that measurement is an abstract model for a data set. Therefore, the numeric values assigned to the respective observations can also be regarded as parameters to be estimated during the course of the analysis.

4.1 The Optimal Scoring Approach to Measurement

The preceding idea is embodied in a measurement philosophy that can be called "optimal scoring." According to this perspective, the measurement *characteristics* of the data set are specified prior to an empirical analysis. However, the numerical *values* that comprise the actual measurements can be obtained as part of the output from the analysis. The optimal scoring approach assumes implicitly that all measurement has solely instrumental, rather than intrinsic, utility for the researcher. That is, measured values are really nothing more than the raw material for a model—usually, statistical in nature—that is to be fitted to the data. Therefore, the numbers that are assigned to the empirical objects under investigation should facilitate this model-fitting process to the greatest extent possible.

Optimal scoring assigns numeric values to the observations in a way that simultaneously fulfills two conditions: (1) The assigned scores strictly maintain the specified measurement characteristics for the data, and (2) they fit the statistical model as well as possible. This optimal scaling strategy provides the *best* set of numerical assignments for the data, where “best” is defined in terms of goodness of fit between an analytic model and a set of empirical observations (Young 1987).

Consider the situation that occurs when a researcher wants to assess the relationship between two variables, both of which are assumed to be measured at the ordinal level. Common practice in the social sciences would be to assign numeric values (probably, successive integers) to the ordered categories on each variable and then simply calculate the product-moment correlation between them. In contrast, the optimal scoring approach would not only calculate the correlation coefficient—it would also seek the numeric scores that produce the highest possible correlation between the two variables. Thus, the values assigned to the categories of the two variables are estimated as part of the empirical analysis.

There is absolutely nothing wrong in doing this, as long as the assigned scores maintain the original orderings of the categories on the respective variables. The optimal scoring approach simply takes advantage of transformations that are permitted under a particular level of measurement and the resultant indeterminacy in the specific measurement values that are assigned to particular observations. Rather than using some arbitrary scheme or employing numerical values that have been selected for mere convenience (e.g., they were assigned by the organization that coded the original data), this strategy selects the set of numeric scores that optimize the statistical model (i.e., produce the largest value for the correlation coefficient) while still conforming to the researcher’s assumptions about the measurement characteristics of the data (i.e., the ordinal nature of the two variables, in this example).

The differences between standard statistical methods and the optimal scoring approach are most evident in the ways they deal with variables that are measured at the nominal and ordinal levels. Traditionally, these “lower” levels of measurement have been regarded as problematic impediments to the use of powerful statistical techniques. The optimal scoring strategy takes exactly the opposite view: Lower measurement levels provide greater flexibility, which can be exploited in order to maximize the fit between a model and a set of empirical observations.

4.2 The ALSOS Approach to Data Analysis

The Alternating Least Squares, Optimal Scaling (ALSOS) approach to data analysis provides a straightforward means of operationalizing the scoring strategy discussed in the previous section (Young 1981). ALSOS is a well-known estimation strategy in psychometrics. However, it seems to be

virtually unknown in political science. There are no ALSOS applications in the research literature, and it is not even mentioned in a recent review of political methodology (Brady and Bartels 1993). This is a very serious gap in our array of methodological tools, because ALSOS is potentially useful for many of the measurement problems that arise in political science research contexts.

Specifically, ALSOS produces optimal scores for the variables. These scores are transformations of the original coded values assigned to the variables. The nature of the transformations from the initial values to the optimal scores provides useful information about the measurement levels of the respective variables.

The ALSOS approach holds that empirical analyses always involve two different models of the observations under investigation. The first represents the structural relationships among the variables, and the second involves the measurement characteristics of the variables. The parameters of *both* models are estimated in an optimal way during the course of an ALSOS analysis.

In this discussion, we will focus on the ALSOS version of multiple regression, in which a single dependent variable is expressed as a linear function of several independent variables (Young, de Leeuw, and Takane, 1976).⁵ However, it is important to keep in mind that the ALSOS approach could be used with any other statistical procedure (e.g., ANOVA, principal components, discriminant analysis, etc.). In fact, one of the major strengths of ALSOS is its generality and broad applicability under a wide variety of situations (Young 1981).

The ALSOS regression algorithm is similar to the usual OLS approach in that it provides the “best-fitting” parameter estimates for a given data set. The ALSOS strategy differs in that it estimates *two* distinct sets of parameters. First, there are the *model* parameters—the regression coefficients. As usual, these are calculated in order to minimize the sum of squared residuals (or maximize the R^2).

At this point, ALSOS and traditional procedures diverge. Most statistical approaches view the data values as fixed, so the estimation would be completed after the first set of parameters. However, ALSOS proceeds to estimate a second set of *measurement* parameters. Here, the routine seeks a set of specific data values that simultaneously maximize the R^2 and retain the pre-specified measurement characteristics for each variable. Interval- and ratio-level variables are generally left unchanged because the only permissible changes to their values (i.e., linear transformations) would have no effect on the R^2 for the statistical model. Nominal and ordinal variables, how-

⁵The specific procedure used to carry out ALSOS regression is sometimes called the MOR-ALS algorithm, for *M*ultiple *O*ptimal *R*egression through *A*lternating *L*east *S*quares.

ever, would be assigned the values that result in the highest possible R^2 and still maintain either the categories (for nominal variables) or the ordering (for ordinal variables) of the original observation categories. In this sense, the measurement parameter estimates are the *optimally scaled* values for the observations. The end result of an ALSOS analysis is an estimated statistical model that provides the best fit to the data in terms of *both* the coefficients and the data values.

The full ALSOS regression algorithm is a bit more complex, but not by much. It begins with two matrices of empirical observations which, together, comprise the initial data:

Y = An n by 1 vector of dependent variable values

X = An n by $k+1$ matrix of independent variable values.

X can be viewed as a partitioned matrix of column vectors:

$[x_0 \ x_1 \ x_2 \ \dots \ x_j \ \dots \ x_k]$. The x_0 vector is composed entirely of 1's, and it is included to give the model an intercept.

Note that the measurement characteristics of the data must also be specified at the beginning of the analysis. These characteristics can differ across the variables. Some of the Y and x_j might be interval-level, others ordinal, and still others could be nominal categories.

The ALSOS regression algorithm estimates the parameters in the following equation:

$$Y^* = X^* \beta^* + E^* \quad (5)$$

Where

$$Y^* = f_y^q(Y) \quad (6)$$

$$X^* = [f_0^q(x_0) \ f_1^q(x_1) \ f_2^q(x_2) \ \dots \ f_k^q(x_k)] \quad (7)$$

$$\beta^* = (X^{*'} X^*)^{-1} X^{*'} Y^* = [b_0 \ b_1 \ \dots \ b_k]' \quad (8)$$

$$\hat{Y}^* = X^* \beta^* \quad (9)$$

$$E^* = Y^* - \hat{Y}^* \quad (10)$$

Y^* and X^* are matrices of quantitative (that is, at least interval-level), optimally-scaled data values. They are functions of the original Y and X , respectively. As indicated in Equations 6 and 7, the exact nature of this function can differ from one variable to the next; it always depends upon the pre-specified measurement characteristics of the variables. Thus, if x_j is assumed to be an ordinal-level measure, then f_j^q is an order-preserving function—that is, it conforms to conditions (2a) and (2b)—and so on. The functions are

used to provide the optimally scaled values for the respective variables. β^* is the vector of coefficients from the regression of Y^* on X^* . \hat{Y}^* is the vector of predicted values for the optimally scaled dependent variable, and E^* is the residual, also calculated from the optimally scaled values.

All of the starred matrices of data values, as well as the coefficient vector, are estimated through an iterative procedure. On each iteration, there are two phases. Each phase minimizes the current value of $E^{*'}E^*$ (the sum of squared residuals from the predicted values of the transformed dependent variable), subject to appropriate constraints. First, there is the “model estimation phase” in which β^* is estimated, holding the current values of X^* and Y^* constant. As was obvious from Equation 8, β^* is simply the OLS estimator applied to the optimally scaled data. Since the Y^* and X^* are all measured at the interval level (regardless of the assumed measurement characteristics of the original Y and X), the interpretation of β^* is straightforward and requires no further discussion here.

Second, new values of Y^* and X^* are estimated while holding the current values of β^* constant. This is the “optimal scaling phase” for the current iteration. In this phase, the previous data values (i.e., those used to calculate the current β^*) are transformed to maximize the fit of the model to the data (i.e., minimize the value of $E^{*'}E^*$) while still conforming to the measurement characteristics that the analyst specified prior to the estimation. The optimal scaling is carried out separately for each variable.

If a variable is measured at the nominal or ordinal levels, the process is carried out as follows: Start the optimal scaling step by generating a set of model-based predicted values. For the dependent variable, this is simply \hat{Y}^* from Equation 9. For an independent variable, say x_k , the predicted values would be calculated as:

$$\hat{x}_k^* = [Y^* - \sum_{j \neq k} x_j^* b_j^*] / b_k^* \quad (11)$$

Now, take the means of the predicted values *within* the categories of the original variable. In other words, if there are C different categories of x_k (or Y), then there will be C different mean predicted values ($\bar{\hat{x}}_{k,c}$ with $c = 1, 2, \dots C$).

At this point, it is necessary to distinguish between nominal and ordinal variables. If x_k is a set of nominal categories, then the $\bar{\hat{x}}_{k,c}$ are used as the optimally scaled values;⁶ they merely replace the previous values assigned

⁶From the ALSOS perspective, nominal measurement implies a numeric variable in which the relative positions of the observational categories (i.e., the ordering and the intervals between categories) are unknown.

to x_k^* . The $\hat{x}_{k,c}$ are the “best-fitting” values, because they are as close as possible (in the least squares sense) to the model-based predicted values, \hat{x}_k , while still maintaining the nominal-level conditions given back in conditions 1a and 1b. This procedure is identical to Fisher’s (1954) appropriate scoring strategy.

If the variable is measured at the ordinal level, the ordering of the x_k^* values must be identical to the ordering of the original C categories that can be discerned among the empirical values of x_k . Kruskal’s (1964) monotone regression algorithm is used to find the appropriate set of values. Whenever the $\hat{x}_{k,c}$ values are monotonically related to the x_k values, they are used, as is, for the x_k^* . But, if monotonicity is violated in the mean predicted values (that is, $\bar{\hat{x}}_{k,1} < \bar{\hat{x}}_{k,2}$ while $x_{k,1} > x_{k,2}$), the nonmonotonic $\hat{x}_{k,c}$ values are averaged until the resulting values are weakly monotonic with respect to the order of the original categories of x_k . These latter values are used for the values of x_k^* . Once again, they are consistent with the necessary measurement level restrictions given by conditions 2a and 2b, and they provide the best least-squares fit to the model-based predicted values.

Finally, if x_k is defined as an interval- or ratio-level variable, then x_k^* must be a specific function of the original values (e.g., linear). This could be handled by regressing the $\hat{x}_{k,c}$ on x_k and using the predicted values from this regression as the new set of x_k^* values. However, this step is not really necessary, since the fit of the regression model to the data would be unaffected by a linear transformation applied to one of the variables.

Regardless of the assumptions about the measurement characteristics, the optimally scaled values replace the previous values in the appropriate data matrix (Y^* or X^*). Also, after each variable is scaled, it must be normalized in order to avoid trivial, degenerate solutions for the coefficients and the variable values (e.g., setting them all to zero). The optimal scaling phase is completed when all variables have had their values “updated” in this manner.

The entire ALSOS regression algorithm proceeds as follows: Given observation matrices Y and X , the analyst first specifies the measurement characteristics for all of the variables and chooses initial values for Y^* and X^* . The easiest way to do this is to simply set $X^* = Y$ and $X^* = X$. Next, the iteration process begins. Each iteration is composed of (1) a model estimation phase, where the β^* vector is calculated from the current Y^* and X^* and (2) an optimal scaling phase, where the appropriate transformations are used to obtain new, better-fitting values for Y^* and X^* subject to the measurement constraints imposed on the respective variables. Within each iteration, the procedure alternates between the two phases. Each phase gives a least-squares estimate of the appropriate parameters (model or measurement) while holding the other set of parameters constant at their current values.

This is obviously the source of the term *Alternating Least Squares*, *Optimal Scaling*. At the end of each iteration, the latest value of $E^* ' E^*$ is calculated; after the first iteration, the current $E^* ' E^*$ is compared to those calculated on the previous iteration(s). The procedure stops when the values of $E^* ' E^*$ converge (i.e., they do not change from one iteration to the next), indicating that the best-fitting model has been found for the given set of observations.

ALSOS provides a set of “conditionally best fitting” parameter estimates. This rather clumsy term refers to the fact that the final ALSOS results are always contingent upon the initial data values. But given some initialization of the variables, the ALSOS routine does minimize the overall discrepancy between the statistical model and the measurement model for that particular set of observations. A more extensive discussion of the ALSOS strategy can be found in Young (1981). A variety of closely related methods are also covered in Gifi (1990) and Van de Geer (1993).⁷

4.3 Using ALSOS to Test Measurement Levels

The ALSOS strategy provides a diagnostic tool for considering hypotheses about *measurement* characteristics (particularly, measurement levels), apart from hypotheses about the structural coefficients in a statistical model. Specifically, an ALSOS analysis can be repeated several times, using the same observations (i.e., initial data values) and statistical model, but altering the specification of measurement characteristics in the data. Any differences in results would have to be due to the changes in the measurement characteristics. This, in turn, provides researchers with some empirical basis for their measurement assumptions (Young 1987).

For example, assume that a regression model is fitted to a set of data in which all variables are assumed to be interval-level measures. Then, repeat the analysis, but assume that the variables are only measured at the ordinal level. In other words, permit the ALSOS routine to estimate the relative sizes of the intervals between adjacent categories on each variable. If the fit im-

⁷The general data analysis strategy underlying ALSOS has been reinvented (apparently independently) several different times by separate sets of researchers. For example, Lyons' (1971) method of effect-proportional recalibration, Bonacich and Kirby's (1976) procedure for establishing conditional metrics, and Hensler and Stipak's (1979) interval scale estimates from criterion variables are all precursors of the more general ALSOS approach proposed by Young and his colleagues. Tenenhaus and Young (1985) have shown that the ALSOS approach is actually a special case of the more general multiple correspondence analysis (also see Gifi 1990). The alternating conditional expectations or “ACE” algorithm is also closely related to the ALSOS regression routine (Breiman and Friedman 1985; Hastie and Tibshirani 1990). ACE is slightly different from ALSOS in that it generates smoothed versions of the data values rather than measurement-based transformations which are explicitly intended to optimize the fit of the statistical model. Still, ACE does tend to produce very similar results to ALSOS when they are applied to the same data, with identical assumptions about the variables.

proves markedly with the second regression, then the original, interval-level measurement assumption must be questioned because that is the only thing that was varied from the first analysis to the second.

This line of reasoning works in the opposite direction as well: If a variable that is assumed to be an ordinal-level measure actually behaves like an interval level measure in an ALSOS analysis (i.e., the optimally scaled values are a linear function of the original data values), then it may as well be treated like an interval-level variable until the evidence suggests otherwise. In this manner, the ALSOS approach encourages researchers to aggressively exploit the characteristics of their data in an effort to discern as clearly as possible the structure underlying their empirical observations. The general point is that the ALSOS approach allows for empirical inspection of things that are typically left as untested assumptions in statistical analyses—namely, the measurement levels of the variables.

It is important to emphasize that ALSOS can never recover *the* measurement levels or even the *best* set of measurement values for the variables included in an analysis. Indeed, when stripped to its barest essentials, ALSOS only provides a set of optimal transformations for the original data values. The resultant “new” data values may provide measurement information about the variables, but they could also reflect functional form in the statistical model that is being fitted to the data. How can the analyst distinguish between these two possibilities? Unfortunately, there is no general answer to this question. Some writers on this topic contend that the distinction is itself a false one (e.g., Velleman and Wilkinson 1993; Cliff 1993). Their basic argument is that nonrandom structure in the data provides useful substantive information regardless of its source; attempts to differentiate the relative impacts of measurement and functional form are unnecessary and generally nonproductive.

Nevertheless, there are some basic guidelines for interpreting ALSOS-based transformations. First, theory should always prevail: If substantive considerations suggest a type of empirical relationship that is confirmed by the ALSOS analysis, then the results should be attributed to functional form rather than measurement. Second, if the ALSOS transformations conform to relatively straightforward functions (e.g., logarithmic, logistic, exponential, etc.) then it is probably reasonable to treat the effects as a component of the statistical model; doing so would not introduce excessive complications into the structural representation of the relationships between the empirical variables. Third, if the optimally-scaled data values are complex transformations of the original variable (i.e., they cannot be expressed as a simple function), then it is probably more useful to treat the phenomenon as a measurement characteristic specific to that variable. And fourth, if repeated ALSOS analyses tend to produce similar transformations for a variable across many

different statistical models (i.e., relating that variable to a variety of other variables), then the results are more likely due to the measurement characteristics of that variable rather than the functional relationships within which it is embedded. Once again, these rules of thumb merely provide some guidance for interpreting ALSOS results. In reality, there is just no clear way to distinguish unambiguously between measurement characteristics and other structural aspects of an empirical statistical model. In any actual research context, the final interpretation will always depend upon theoretical parsimony, analytic convenience, and even the investigator's personal biases.

This kind of indeterminacy may seem to be troublesome. Indeed, it is, but only if one adheres to the overly strict view that measurement levels are fixed and unalterable characteristics of the data. I contend that this latter perspective is unrealistic and contrary to the principles of good scientific research. Measurement is a model of a set of observations. Whenever possible, measurement models (like any other kind of model) should be tested in order to assess their consistency with empirical information. Traditional views and the prevailing practices in the social sciences seldom do so. Hence, it is difficult to establish the utility of any particular measurement scheme for the achievement of scientific objectives.

In contrast, the approach suggested here directly follows the logic of the scientific method: Measurement represents a hypothesized structure for a set of data. It is impossible to prove that this structure is the "true" representation of the observations; there are an infinite number of alternative structures that would work just as well. But the hypothesized structure can be *rejected*, in light of empirical evidence to the contrary. The ALSOS strategy operationalizes exactly this kind of reasoning, with respect to levels of measurement in regression models.

It is important to emphasize that ALSOS is a theory-driven research strategy. This approach recognizes explicitly the two types of theories that researchers bring to bear on any empirical analyses. On the one hand, substantive theory suggests the appropriate variables for investigation as well as the nature of the statistical relationships between them. On the other hand, measurement theory indicates how the analyst interprets the values of the respective variables. Both of these theories are always present whenever a researcher attempts to discern structure within a set of empirical observations. ALSOS merely provides a convenient method for jointly estimating the parameters of the two resultant models which are derived from these theories. In the statistical model, these parameters comprise the coefficients in the regression equation. In the measurement model, these parameters consist of the numerical values assigned to the observations on each of the variables.

5. A SUBSTANTIVE EXAMPLE: VOTING CHOICE IN 1992

The ALSOS approach is potentially useful in all substantive fields of the social sciences. It should be particularly relevant for the kinds of research questions that arise in the study of mass political attitudes and behavior. Most of the empirical work in this area is based upon data collected from public opinion surveys. The variables obtained in such contexts are precisely the type where measurement characteristics are ambiguous. For example, many of the questions routinely included in the CPS National Election Studies interview schedules employ response formats that are intended to provide "pseudo-interval" measurement of respondents' underlying political orientations. These include feeling thermometer ratings, the seven-point party identification index, the liberal-conservative self-placement item, and a variety of other Likert-type rating scales. While it is usually assumed that the differences in the scores assigned to adjacent categories reflect roughly equal differences in the magnitudes of the respective beliefs or attitudes, this assumption is almost always left untested. The arbitrary nature of this assumption is particularly troublesome because several recent studies have raised serious questions about the measurement characteristics of these variables (e.g., Lodge and Tursky 1979; Alwin 1997).

In order to show how the ALSOS approach can be used to test hypotheses about levels of measurement, let us consider a specific example: the determinants of citizens' 1992 presidential candidate choices. The data are taken from the 1992 CPS National Election Study, and the variables are operationalized in the standard ways.

The dependent variable is the difference in feeling thermometer ratings for George Bush and Bill Clinton. The resultant score ranges from -100 to +100. Zero is the neutral point, and larger values indicate a stronger preference for Bush over Clinton. This variable is often used as a nearly-continuous, interval-level surrogate for individual voting choice (e.g., Markus and Converse 1979) although this assumption has never really been tested.

There are three independent variables: the seven-point party identification and ideological self-placement indices, along with a five-category rating scale gauging respondents' feelings about the nation's economy over the preceding four years. These variables are all very prominent in the voting behavior literature. They represent the effects of longstanding partisan loyalties (e.g., Campbell et al. 1960), liberal-conservative symbols (e.g., Levitin and Miller 1979), and retrospective sociotropic judgments (Kiewiet 1983), respectively. All of these variables employ successive integer-scoring schemes, taken straight from the 1992 NES codebook. Table 1 shows the categories, scale values, and frequency distributions for each of the independent variables.

Table 1. Category Definitions, Initial Measurement Values, and Frequency Distributions for Independent Variables in ALSOS Analysis of 1992 Candidate Choice

| Category Definitions | Initial Data Values | Frequency | Percent |
|---|---------------------|-----------|---------|
| <i>Party Identification:</i> | | | |
| Strong Democrat | 0 | 300 | 18.1 |
| Democrat | 1 | 265 | 16.0 |
| Independent, leaning Democratic | 2 | 230 | 13.9 |
| Independent | 3 | 169 | 10.2 |
| Independent, leaning Republican | 4 | 218 | 13.2 |
| Republican | 5 | 241 | 14.6 |
| Strong Republican | 6 | 230 | 13.9 |
| <i>Ideological Self-Placement:</i> | | | |
| Extremely liberal | 1 | 43 | 2.6 |
| Liberal | 2 | 199 | 12.0 |
| Slightly liberal | 3 | 220 | 13.3 |
| Moderate, middle of the road | 4 | 522 | 31.6 |
| Slightly conservative | 5 | 327 | 19.8 |
| Conservative | 6 | 286 | 17.3 |
| Extremely conservative | 7 | 56 | 3.4 |
| <i>Did nation's economy become better or worse over past 4 years?</i> | | | |
| Much better | 1 | 11 | 0.7 |
| Better | 2 | 73 | 4.4 |
| Stayed the same | 3 | 186 | 11.3 |
| Worse | 4 | 532 | 32.2 |
| Much Worse | 5 | 851 | 51.5 |

Note: These data are obtained from the 1992 CPS National Election Study. The number of observations is 1653, and the measurement values are taken directly from the codes assigned in the NES Codebook.

For the first step in the analysis, we simply assume that all of the variables are measured at the interval level (or higher). Ordinary least squares is used to estimate the model coefficients and the goodness of fit value. The results are shown in the leftmost column of Table 2. The R^2 for this equation is quite good, at 0.541. All three independent variables exhibit coefficients that are much larger than their standard errors and they are all appropriately signed. This means that preferences for George Bush are most likely to occur among Republican identifiers, self-styled conservatives, and people who thought that the nation's economy had improved since 1988. Preferences for Clinton, of course, are most prominent among people who display the opposite characteristics.

Table 2. OLS and ALSOS Estimates of Regression Model Predicting 1992 Candidate Choices as a Function of Party Identification, Ideological Self-Placement, and Retrospective Judgments about the Nation's Economy

| | OLS Coefficient Estimates | ALSO Coefficient Estimates |
|-------------------------------------|---------------------------------|----------------------------------|
| Party Identification | 11.873 (0.398) | 12.845 (0.496) |
| Ideological Self-Placement | 6.394 (0.577) | 5.590 (0.611) |
| Judgments about Nation's Economy | -7.731 (0.860) | -7.852 (0.901) |
| Intercept | -30.115 | -28.979 |
| Standard deviation of residuals | 29.795 | 27.996 |
| R ² | 0.541 | 0.595 |

Note: Entries in the left column of the table are OLS coefficients estimated for the original data values, as coded in the 1992 NES Codebook. Entries in the right column of the table are OLS coefficients estimated on the monotonically-transformed values of the variables, obtained from the ALSOS routine. The number of observations is 1653. The figures within parentheses are the standard errors. In the left column, these are the usual standard errors, obtained through least-squares theory and the assumption of Gaussian residuals. The right column shows bootstrap standard errors, since ALSOS estimates usually cannot be used for traditional statistical inference tests (Kuhfeld 1990). Specifically, the figures are the standard deviations of the bootstrapped ALSOS coefficient estimates obtained across 100 bootstrap samples, drawn with replacement from the original data. Inspection of normal quantile plots for the bootstrap replications suggests that the ALSOS coefficient estimates are, themselves, normally distributed. If this is the case, then the bootstrap standard errors can be used in the usual ways to perform hypothesis tests and construct confidence intervals for the coefficients. Gifi (1990) discusses bootstrapping in the context of ALSOS estimation. Complete information about the bootstrapping procedure used here and a SAS macro for performing the bootstrap resampling of the ALSOS estimates are available from the author upon request.

The next step is to perform an ALSOS regression analysis on the same data, making the assumption that the variables are only measured at the ordinal level. These results are presented in the rightmost column of Table 2. If the variables truly are interval-level measures, then the ALSOS estimates should be identical to those from the OLS estimation. However, this is not the case. Notice that the R^2 in the right column is higher, at 0.595: an increase of 0.054 over the value obtained with the OLS estimates. This means that the linear statistical model fits the empirical data more closely when the sizes of the inter-category intervals are permitted to vary in each of the variables—in other words, when the variables are interpreted as providing ordinal-, but not interval-level information about their respective attributes.

Figure 1 plots the optimally-scaled values (hereafter, “OS values”) from the ordinal analysis against the assumed interval values (i.e., the original coding schemes) for each of the variables. These graphs are useful for determining exactly where and how the interval-level measurement assumption breaks down in the data. The reasoning is as follows: If the variables truly are interval-level measures, then the ordinal OS values should remain a linear function of the original scores, even though they are not constrained to do so by the ALSOS algorithm. If, however, the ordinal OS values are a non-linear monotone function of their interval-level counterparts, then the shape of the plotted point array should reveal where the assumption of equal intervals between categories is violated.

Inspection of Figure 1 reveals that the equal-interval assumption is questionable for three of the variables. Consider the graph for the dependent variable shown in Figure 1A. Here, the plotted points conform to an “S-shaped” pattern. This indicates that the interval sizes in the differential feeling thermometer are smaller at the two extremes than they are in the center of the variable’s range. For example, the OS values show that two people with scores of +90 and +100 on the original thermometer differential do not seem to differ very much in terms of their relative preference for Bush over Clinton. In contrast, the difference in strength of preference is apparently much larger for two people with scores of 0 and +10. This pattern violates systematically the interval-level assumption that provides the primary justification for using the net thermometer rating variable in empirical analyses of voting behavior in the first place. The sigmoid function, however, does make a great deal of sense from the perspective of decision-making theories (Luce and Suppes 1965), and it may provide separate validation for the binary choice models (e.g., logit and probit) that are now commonly used to study voting behavior. With the data used here, the problem can be handled very easily by respecifying the functional form of the model—either by logging the separate thermometer ratings for Bush and Clinton or by performing a logistic transformation on the thermometer difference variable. In either case, it appears that the structure in the data can be accommodated by adjusting the statistical model, rather than modifying the measurement assumptions about the dependent variable.

Turning to the independent variables, the graph for party identification (Figure 1B) shows that the differences between independent leaners and weak partisans are relatively small, compared to the intervals separating the other categories on this variable. This finding is quite consistent with results obtained in many other studies of the party identification variable (e.g., Lodge and Tursky 1979; Jacoby 1982). The fact that very similar patterns show up across a large variety of empirical relationships involving the party identification index suggests that the interval-level measurement properties often attributed to this variable are at least partially inappropriate. A more

Figure 1. Monotonic Transformations of Original Data Values, Obtained from ALSOS Regression Analysis of the Candidate Choice Model

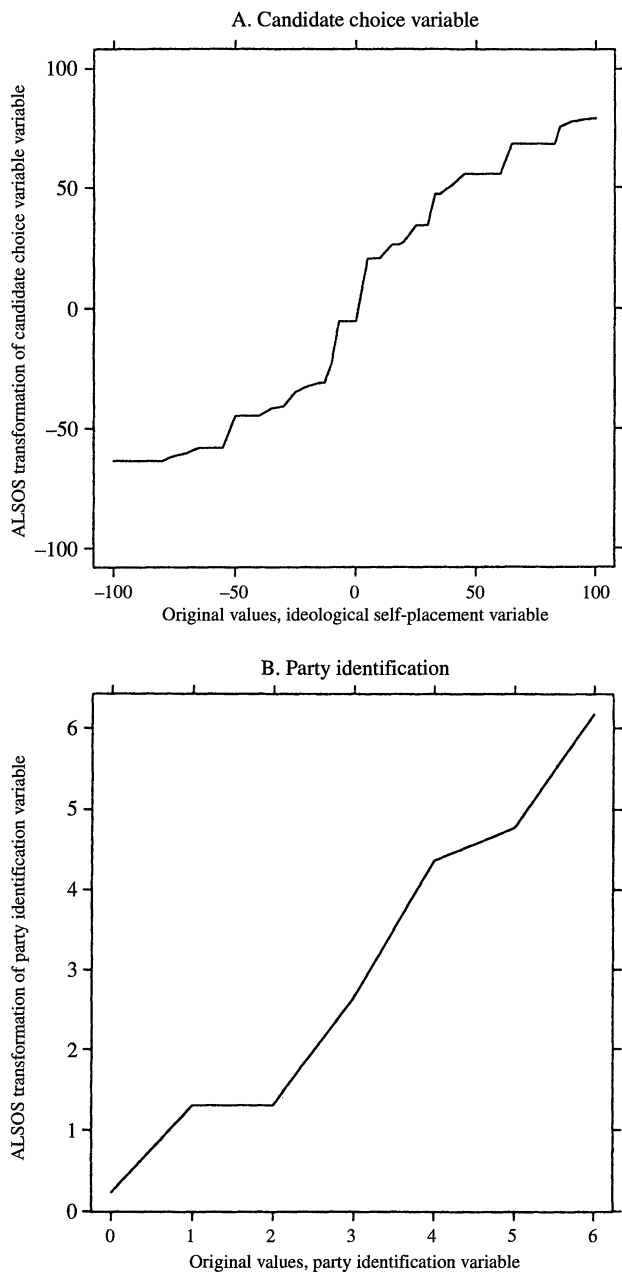
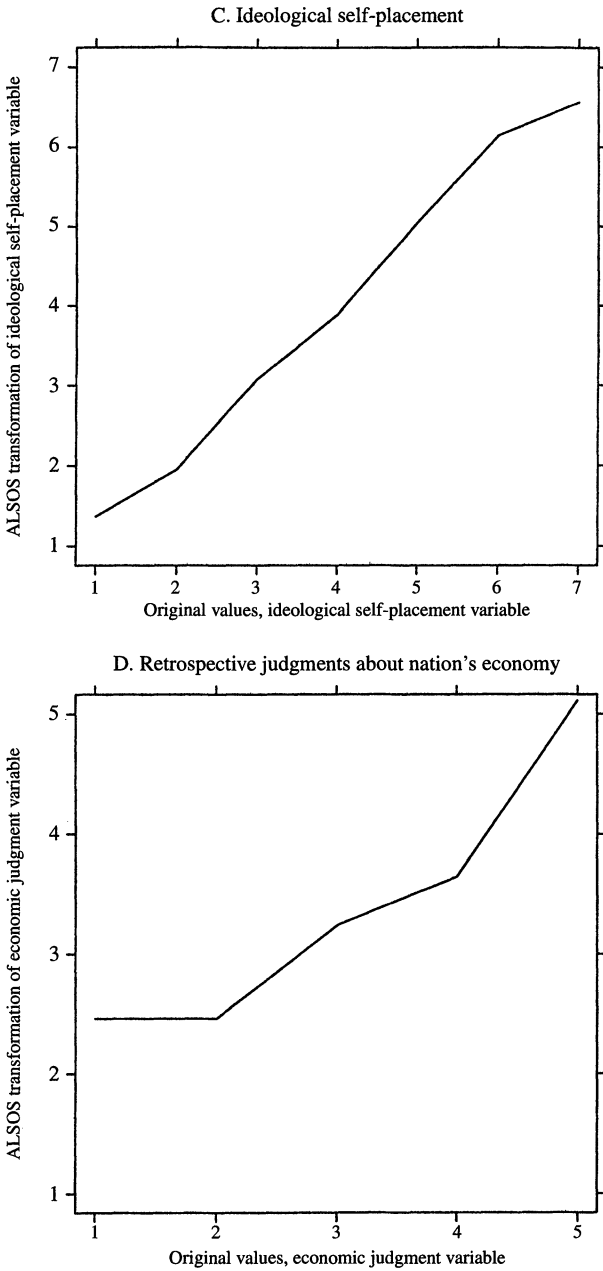


Figure 1. Monotonic Transformations of Original Data Values, Obtained from ALSOS Regression Analysis of the Candidate Choice Model (continued)



accurately-specified version would adjust the category widths accordingly (Weisberg 1980). In this case, it appears to be the measurement characteristics of the party identification variable, rather than the statistical models within which it is embedded, that are at fault.

Figure 1C shows that the interval-level measurement assumption does receive empirical support in the case of ideological self-placements. The OS values approximate closely a linear functional relationship with respect to the original, successive-integer scores assigned to this variable. Of course, the pattern is not perfect; for example, the interval between people who are “slightly liberal” and those who are “liberal” is a bit larger than the two intervals on either side of it. At the other end of the continuum, the interval between the OS values for “conservatives” and “extreme conservatives” is relatively small. Nevertheless, these departures from linearity within the array of points simply do not appear to be serious enough to call the assumed measurement level of the original variable into question.

The interval-level measurement assumption seems to be most problematic for retrospective economic judgments. Figure 1D shows that the OS estimates of the intercategory differences vary quite a bit across the range of the original, equal-interval scale. There appears to be no difference between people who said the U.S. economy was “Much better” and those who said it was only “Somewhat better.” In contrast, the largest difference occurs between respondents who said the economy was “Much worse” and those who said it was only “Somewhat worse.” This result is fully consistent with theories about “negativity effects” in political evaluations (Lau 1985). It shows that incumbents suffer from a bad economy, but do not receive additional approval for a good economy. From the perspective of measurement, the OS values suggest that the original five-category variable is over-parameterized. In other words, there are more distinct data values than are sustained by the psychological characteristic that is being measured.

The discussion so far has emphasized a rather pessimistic interpretation: i.e., the results in the left column of Table 2 are questionable because the measurement assumptions required for the variables are not met in the data. But it is possible to take a more optimistic view of the same evidence. The R^2 from the ALSOS estimation (0.595) is not that much larger than the value obtained in the original OLS analysis (0.541); thus, the fit is not *that much* better under the ordinal measurement assumptions.⁸ Note, too, that the optimally scaled variables are normalized to the same means and standard

⁸This small increase in the R^2 value is fairly typical for ALSOS analyses with ordinal data. It occurs because ordinal data with many observations and relatively few distinct measurement categories actually impose a sizable number of metric constraints across the observations (e.g., Abelson and Tukey 1970). Thus, ALSOS cannot be used (as some skeptics may fear) to arbitrarily inflate the magnitudes of the relationships between the variables.

deviations as the original variables; therefore, it is possible to compare the results for the respective variables across the two columns of Table 2. Doing so reveals that the regression coefficients are very similar in every case. The relative influences of the separate independent variables (and hence, the substantive interpretations of their effects) appear to be virtually identical in both the OLS and the ALSOS results.

We are thus left in a somewhat ambiguous situation. From one perspective, the measurement assumptions employed in a typical regression analysis seem to be invalid. But from another perspective, we do not cause any great violations to the structure in the data by making these strong measurement assumptions. The final results are basically unaffected when the assumptions are relaxed. Accordingly, there is probably no reason to abandon the traditional measures of candidate preference, party identification, ideology, and retrospective economic judgments. However, it is probably useful to maintain a somewhat skeptical view of the accuracy with which these equal-interval variables capture gradations in the underlying psychological characteristics that they seek to measure.

For now, the important point is that we have obtained *empirical* information about the measurement levels appropriate for these variables. This provides some sense of the implications involved in taking either an optimistic or pessimistic view of the data's measurement properties. From the optimistic perspective, the assumption of interval-level data results in a fairly simple regression equation, but this ignores the structure in the data that is implied by the unequal intercategory intervals revealed in the ALSOS analysis. Alternatively, the more pessimistic view of the variables' measurement levels creates problems for the regression model—what are the appropriate data values that should be input to the equation? There is more information (again, from the unequal intercategory intervals) about variability in citizens' political attitudes and preferences, however. This kind of information would have been completely invisible in any analysis that relied strictly on the traditional fourfold distinction between nominal, ordinal, interval, and ratio variables.

6. CONCLUSIONS

Before closing this discussion, it is useful to consider why characteristics like the levels of measurement are important in the first place. Most of the previous work on this topic has focused on the relationship between measurement levels and statistical techniques. Some authors have supported the position, originally articulated by Stevens (1951), that analytic strategies are constrained by the measurement scales of the variables under investigation (e.g., Stine 1989). Others have argued that measurement levels are largely irrelevant to statistical modeling efforts (e.g., Borgatta and Bohrnstedt 1980; Gaito 1980) and that blind adherence to a set of simplistic measurement prin-

ciples can be very misleading (Duncan 1984; Velleman and Wilkinson 1993). Still other researchers have taken a variety of intermediate positions. They suggest that, at least under certain circumstances, classical statistical methods can be applied to variables with “low” levels of measurement in order to draw somewhat limited inferences about the structures underlying empirical data (e.g. Labovitz 1970; Davison and Sharma 1990).

Disagreements about the relationship between measurement levels and statistics continue, with no sign of any authoritative resolution in the near future (Michell 1986). Therefore, I prefer an entirely different view about the importance of measurement levels, based upon the general objectives of scientific research. Scientists try to explain differences among empirical objects. Because of this, they are more concerned with variability in a measure, rather than the actual numeric values that are assigned to observations on a given variable. But observed variability can arise from at least two sources: (1) the underlying substantive property of the objects—that is, the attribute the researcher is trying to measure; (2) the ways that numbers are assigned to the objects—that is, the nature of the function mapping the observations into the real number system.

The levels of measurement differ among themselves with respect to the second source of variability. Higher measurement levels are more restrictive in the kinds of functions that can be used to assign numbers to objects. Therefore, a higher proportion of the observed variance is going to be attributable to the substantive differences among the objects, rather than to the “slippage” that invariably occurs among the numerical assignments. This alone justifies the use of higher measurement levels wherever possible. Within the limits imposed by measurement accuracy (i.e., the degree of measurement error), variables measured at higher levels convey more information about substantive differences between objects than do variables measured at lower levels. Thus, measurement levels are critically important for the achievement of scientific research objectives.

To say that measurement levels are important, however, does not provide any guidance for understanding their implications in empirical research contexts. Thus many researchers believe that they are forced to deal with poorly-measured variables simply because of the substantive problems that they want to address (e.g., “public opinion surveys only produce ordinal-level rating scales, at best”). In this paper, I have taken a rather iconoclastic perspective on this problem. The prevailing view in political science is that measurement characteristics are fixed and unalterable properties of one’s data. I argue that this is *not* a realistic or constructive view about the nature of measurement.

A more useful interpretation centers around the idea that *all* measurement is theory testing. The specific variable values used in any empirical analysis are never immutable characteristics of the observations. Instead,

they are simply the components of an abstract model. My position is that the parameters of any measurement model (i.e., the measurement level assumed, along with the specific numeric values assigned) should be subjected to empirical testing, rather than left as a priori, unchangeable assertions. This is particularly important in the social sciences, where critics often charge that the available analytic techniques far exceed the capacities of the data to which they are applied. By explicitly testing our measurement characteristics, we should be able to formulate responses to these critics. And in so doing, researchers will be able to exploit the information contained within their data to the greatest possible extent.

Manuscript submitted November 10, 1997.

Final manuscript received June 1, 1998.

APPENDIX

Software for Performing ALSOS Analyses

The ALSOS approach has already been programmed into software that is widely available within the political science research community. Young (1981) provides sources for a variety of ALSOS programs. The need for these special-purpose FORTRAN routines, however, has been largely eliminated because powerful procedures are now included within two widely-available statistical packages. In SAS, PROC TRANSREG can be used to carry out precisely the kind of ALSOS regression analysis discussed above, while PROC PRINQUAL performs an ALSOS version of principal components analysis and PROC MDS carries out a related form of multidimensional scaling. In addition, PROC IML, the matrix programming language within SAS, contains a function called OPSCAL. This provides an optimally-scaled version of the values in a data vector, at either the nominal or ordinal measurement level. With the OPSCAL function, it is extremely easy to program one's own ALSOS algorithms.

The CATEGORIES module in SPSS for Windows performs several kinds of ALSOS analyses. However, the terminology used for the commands and documentation are drawn a bit more directly from Gifi (1990) than from Young's (1981) nomenclature. SPSS also contains ALSCAL, a multidimensional scaling routine that is based directly upon ALSOS principles.

Another alternative is Young's *ViSta* program (Young 1996), which includes a routine for regression with an optimally-scaled monotonic transformation of the dependent variable. Hence, it could be viewed as a "partial" implementation of the ALSOS approach. *ViSta* is available as freeware on the Worldwide web, at the following address: <http://forrest.psych.unc.edu/research/ViSta.html>.

There is also a function in S-Plus called "ACE," which carries out the alternating conditional expectations algorithm. In this function, the user can specify any combination of identity-preserving or monotonic transformations, so it can be employed in a manner very similar to ALSOS in order to test measurement hypotheses.

REFERENCES

- Abelson, Robert P., and John W. Tukey. 1970. "Efficient Conversion of Nonmetric Information to Metric Information." In *The Quantitative Analysis of Social Problems*, ed. Edward R. Tuft. Reading, Mass.: Addison-Wesley.
- Achen, Christopher H. 1983. "Toward Theories of Data: The State of Political Methodology." In *Political Science: The State of the Discipline*, ed. Ada W. Finifter. Washington, D.C.: American Political Science Association.
- Alwin, Duane F. 1997. "Feeling Thermometers Versus Seven-Point Scales: Which are Better?" *Sociological Methods and Research* 25:318–340.
- Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley.
- Blalock, Hubert M. 1982. *Conceptualization and Measurement in the Social Sciences*. Beverly Hills, CA: Sage Publications.
- Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: John Wiley.
- Bonacich, Philip, and Douglas Kirby. 1976. "Using Assumptions of Linearity to Establish a Metric." In *Sociological Methodology 1976*, ed. David R. Heise. San Francisco: Jossey-Bass.
- Borgatta, Edgar F., and George W. Bohrnstedt. 1980. "Levels of Measurement: Once Over Again." *Sociological Methods and Research* 9:147–160.
- Brady, Henry E., and Larry M. Bartels. 1993. "The State of Quantitative Political Methodology. In *Political Science: The State of the Discipline II*, ed. Ada W. Finifter. Washington, D.C.: American Political Science Association.
- Breiman, L., and J. H. Friedman. 1985. "Estimating Optimal Transformations for Multiple Regression and Correlation." *Journal of the American Statistical Association* 80:580–597.
- Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. Chicago: University of Chicago Press.
- Carnap, Rudolf. 1966. *An Introduction to the Philosophy of Science*. New York: Basic Books.
- Carter, Lewis F. 1971. "Inadvertent Sociological Theory." *Social Forces* 50:12–25.
- Cliff, Norman. 1993. "What Is and Isn't Measurement." In *A Handbook for Data Analysis in the Behavioral Sciences*, ed. Gideon Keren and Gordon Lewis. Hillsdale, N.J.: Lawrence Erlbaum.
- Coombs, C. H., H. Raiffa, and R. M. Thrall. 1954. "Some Views on Mathematical Models and Measurement Theory." *Psychological Review* 61:132–144.
- Davison, Mark L., and Anu R. Sharma. 1990. "Parametric Statistics and Levels of Measurement: Factorial Designs and Multiple Regression." *Psychological Bulletin* 107:394–400.
- Duncan, Otis Dudley. 1984. *Notes on Social Measurement: Historical and Critical*. New York: Russell Sage Foundation.
- Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven: Yale University Press.
- Fisher, R. A. 1954. *Statistical Methods for Research Workers (12th Edition)*. Edinburgh, Scotland: Oliver and Boyd.
- Gaito, John. 1980. "Measurement Scales and Statistics: Resurgence of an Old Misconception." *Psychological Bulletin* 87:564–567.
- Gifi, Albert. 1990. *Nonlinear Multivariate Analysis*. New York: Wiley.
- Green, Donald Philip. 1988. "On the Dimensionality of Sentiment toward Partisan and Ideological Groups." *American Journal of Political Science* 32:758–780.
- Hastie, T. R., and R. J. Tibshirani. 1990. *Generalized Additive Models*. London: Chapman and Hall.
- Hensler, Carl, and Brian Stipak. 1979. "Estimating Interval Scale Values for Survey Item Response Categories." *American Journal of Political Science* 23:627–649.
- Jacoby, William G. 1982. "Unfolding the Party Identification Scale: Improving the Measurement of an Important Concept." *Political Methodology* 8:33–60.

- Kaplan, Abraham. 1964. *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler Publishing.
- Keith, B. E., D. B. Magleby, C. J. Nelson, E. Orr, M. C. Westlye, and R. E. Wolfinger. 1992. *The Myth of the Independent Voter*. Berkeley: University of California Press.
- Kiewiet, D. Roderick. 1983. *Macroeconomics and Micropolitics*. Chicago: University of Chicago Press.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Kline, Morris. 1985. *Mathematics and the Search for Knowledge*. New York: Oxford University Press.
- Kruskal, Joseph B. 1964. "Nonmetric Multidimensional Scaling: A Numerical Method." *Psychometrika* 29:115–129.
- Kuhfeld, Warren F. 1990. *SAS Technical Report R-108: Algorithms for the PRINQUAL and TRANSREG Procedures*. Cary, NC: SAS Institute, Inc.
- Kuhn, Thomas S. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.
- Labovitz, Sanford. 1970. "The Assignment of Numbers to Rank Order Categories." *American Sociological Review* 35:515–524.
- Lau, Richard R. 1985. "Two Explanations for Negativity Effects in Political Behavior." *American Journal of Political Science* 29:119–138.
- Levitin, Teresa R., and Warren E. Miller. 1979. "Ideological Interpretations of Presidential Elections." *American Political Science Review* 70:751–771.
- Lingoes, James C. 1979. "Measurement." In *Geometric Representations of Relational Data*, ed. James C. Lingoes et al. Ann Arbor, Mich.: Mathesis Press.
- Lodge, Milton, and Bernard Tursky. 1979. "Comparisons Between Category and Magnitude Scaling of Political Opinion Employing SRC/CPS Items." *American Political Science Review* 73:50–66.
- Luce, R. Duncan, and Patrick Suppes. 1965. "Preference, Utility, and Subjective Probability." In *Handbook of Mathematical Psychology, Volume III*, ed. R. D. Luce et al. New York: John Wiley.
- Lyons, Morgan. 1971. "Techniques for Using Ordinal Measures in Regression and Path Analysis." In *Sociological Methodology 1971*, ed. Herbert L. Costner. San Francisco: Jossey-Bass.
- Markus, Gregory, and Philip E. Converse. 1979. "A Dynamic Simultaneous Equation Model of Electoral Choice." *American Political Science Review* 73:1055–1070.
- Michell, Joel. 1986. "Measurement Scales and Statistics: A Clash of Paradigms." *Psychological Bulletin* 102:398–407.
- Petrocik, John R. 1974. "An Analysis of Intransitivities in the Index of Party Identification." *Political Methodology* 1:31–48.
- Porter, Theodore M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.
- Stevens, S. S. 1946. "On the Theory of Scales of Measurement." *Science* 103:677–680.
- Stevens, S. S. 1951. "Mathematics, Measurement, and Psychophysics." In *Handbook of Experimental Psychology*, ed. S. S. Stevens. New York: John Wiley.
- Stevens, S. S. 1968. "Measurement, Statistics, and the Schemapiric View." *Science* 161:849–856.
- Stevens, S. S. 1975. *Psychophysics*. New York: John Wiley.
- Stine, William Wren. 1989. "Meaningful Inference: The Role of Measurement in Statistics." *Psychological Bulletin* 105:147–155.
- Tenenhaus, M., and F. W. Young. 1985. "An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis, and Other Methods for Quantifying Categorical, Multivariate Data." *Psychometrika* 50:91–119.

- Torgerson, Warren S. 1958. *Theory and Methods of Scaling*. New York: John Wiley.
- Valentine, David C., and John R. Van Wingen. 1980. "Partisanship, Independence, and the Partisan Identification Question." *American Politics Quarterly* 8:165–186.
- Van de Geer, John P. 1993. *Multivariate Analysis of Categorical Data: Theory*. Newbury Park, CA: Sage Publications.
- Velleman, Paul F., and Leland Wilkinson. 1993. "Nominal, Ordinal, Interval, and Ratio Typologies are Misleading." *The American Statistician* 47:65–72.
- Weisberg, Herbert F. 1980. "A Multidimensional Conceptualization of Party Identification." *Political Behavior* 2:33–60.
- Wilson, Thomas P. 1971. "Critique of Ordinal Variables." *Social Forces* 49:432–444.
- Young, Forrest W. 1981. "The Quantitative Analysis of Qualitative Data." *Psychometrika* 46:357–388.
- Young, Forrest W. 1987. "Theory." In *Multidimensional Scaling: History, Theory, and Applications*, ed. Forrest W. Young and Robert M. Hamer. Hillsdale, N.J.: Lawrence Erlbaum.
- Young, Forrest W. 1996. *ViSta: The Visual Statistics System*. Chapel Hill, NC: L. L. Thurstone Psychometrics Laboratory, University of North Carolina.
- Young, Forrest W., Jan De Leeuw, and Yoshio Takane. 1976. "Regression with Qualitative and Quantitative Variables: An Alternating Least Squares Method with Optimal Scaling Features." *Psychometrika* 41:505–529.