Using democracy in empirical work requires accurate measurement. Yet, most policy and academic research presupposes the accuracy of available measures. This article explores judge-specific measurement errors in cross-national indicators of liberal democracy. The authors evaluate the magnitude of these errors in widely used measures of democracy and determine whether their results replicate during a 17-year period (1972 to 1988). Then, they examine the nature of these systematic errors, hypothesizing that three different processes—(a) the information available for rating, (b) the judges' processing of this information, and (c) the method by which a judge's processing decisions are translated into a rating—could create error. The authors find that for the 17-year period from 1972 to 1988, there is unambiguous evidence of judge-specific measurement errors, which are related to traits of the countries. In the conclusion, the authors discuss the implications for democracy research and for other subjective measures.

SUBJECTIVE MEASURES OF LIBERAL DEMOCRACY

KENNETH A. BOLLEN
University of North Carolina at Chapel Hill

PAMELA PAXTON
Ohio State University

The global democratization trend of the last decade has been one of the most consequential transitions in the contemporary world. Scholars and policy makers share an interest in monitoring this trend to determine its extent and whether it has peaked, plateaued, or is starting to decline. At the same time, social scientists seek to understand the determinants and consequences of liberal democracy (e.g., Burkhart & Lewis-Beck, 1994; Knack & Keefer, 1997; Lipset, 1994). Both this policy and academic work presupposes the accuracy of measures of democracy. Random and nonrandom measurement error can bias coefficient estimators of effects and distort comparisons across countries, undermining the empirical results that ignore it.

AUTHORS' NOTE: *Our thanks to Barbara Entwisle, HyoJoung Kim, Arthur Banks, Leonard Sussman, and the referees for their comments on an earlier draft of this article. We gratefully acknowledge the support of the Sociology Program and the Division of Graduate Education of the National Science Foundation (SES-9121564 and GER-9554569).*

COMPARATIVE POLITICAL STUDIES, Vol. 33 No. 1, February 2000 58-86

The majority of current books and articles in comparative research give minimal attention to measurement problems (Bollen, Entwisle, & Alderson, 1993, pp. 321-351). In this article, we address measurement issues in one area of comparative research—studies of liberal democracy. Currently, anecdotal evidence of systematic errors in democracy indicators exists (Hartman & Hsiao, 1988; Nagle, 1985, p. 95; Scoble & Wiseberg, 1981, p. 160), and two recent quantitative empirical analyses (Bollen, 1993; Bollen & Paxton, 1998) document the presence of judge-specific method effects for eight democracy indicators circa 1980.

We extend these prior studies in several major ways. First, we replicate the finding of method effects in liberal democracy measures during a 17-year period, 1972 to 1988. The replication demonstrates that these systematic errors are not unique to a particular time point. A second extension lies in our development of theoretical arguments to explain the origin of judge-specific method factors in these ratings. Previous work suggested several possible sources of the judge-specific effects. Here, we elaborate the most likely source—the perceptual biases of the judges making the subjective ratings. As a third extension, we empirically test the various theoretical arguments for the determinants of judge-specific method factors in the ratings. This reveals the association between these systematic errors and characteristics of countries, and we alert researchers to substantive areas where the bias in the ratings could matter for empirical research. Furthermore, we provide suggestions for how to correct for the errors in empirical work and forewarn researchers who are devising subjective ratings of liberal democracy (or similar concepts) to likely sources of error in their judgments.

In the next section of the article we briefly discuss the definition and measurement of liberal democracy. Then, a theoretical section outlines several processes by which a judge can introduce systematic errors, in the form of judge-specific method factors, into subjective indicators of democracy. Next, we replicate a model of method factors in measures of liberal democracy during a 17-year period. Finally, we examine possible determinants of the judge-specific method factors based on the hypotheses outlined in the theoretical section. The conclusion reviews our findings and discusses their implications for democracy research.

THE DEFINITION AND MEASUREMENT OF LIBERAL DEMOCRACY

Our working definition of liberal democracy is the degree to which a political system allows *democratic rule* and *political liberties*. The first

dimension, democratic rule, exists to the extent that the national government is accountable to the general population, and each individual is entitled to participate in the government directly or through representatives (Bollen, 1986, pp. 568-569; Bollen, 1993). Political liberties, the second dimension, exist when the people of a country have the freedom to express a variety of political opinions in any media and the freedom to form and to participate in any political group. The first dimension, democratic rule, is similar to the "procedural" definition of democracy of Schumpeter (1950) and the definitions of Lipset (1959), Huntington (1991, pp. 5-13), and the many others influenced by Schumpeter's (1950) definition.[1] These theorists give primary emphasis to electoral accountability of elites, with much less attention to the political-liberties dimension. But Dahl (1971) employs eight "institutional guarantees" as part of his definition of democracy ("polyarchy") that we can classify as either political liberties or democratic rule (see Bollen 1990, pp. 10-11).[2] So our use of a two-dimensional concept of democracy, incorporating both democratic rule and political liberties, is not identical to, but has clear ties to, other conceptualizations of democracy.

Efforts to measure liberal democracy have a long history.[3] Reviewing these efforts reveals two broad traditions. One attempts to use objective measures such as voter turnout statistics (e.g., Lerner, 1958), the composition of legislative bodies (e.g., Cutright, 1963; Vanhanen, 1990), or the franchise in a political system (e.g., Cutright & Wiley, 1969) to determine democracy. The other tradition relies on expert judges who rate countries' standings on aspects of liberal democracy such as the fairness of elections, the freedom of the media, or the liberties of political groups. The subjective indicators dominate practical and scholarly usage and are our primary concern.

In our investigation of subjective indicators, we consider three judges of liberal democracy who produce subjective measures. These judges' subjective ratings have three valuable characteristics: (a) They correspond to the theoretical definition of liberal democracy, (b) they are available for multiple years, and (c) there are multiple indicators available from each judge. The desirability of the first two characteristics is obvious. The third is necessary

1. Schumpeter (1950) defines democracy as "that institutional arrangement for arriving at political decisions in which individuals acquire the power to decide by means of a competitive struggle for the people's vote" (p. 269). Schumpeter only acknowledged that some freedom was necessary for a democracy (pp. 272-273).

2. For instance, Dahl's (1971) "freedom of expression" and "freedom to form and join organizations" are instances of political liberties, but his "right to vote" and "free and fair elections" are instances of democratic rule.

3. See Vanhanen (1990), Arat (1991), and Bollen (1980, 1995) for summaries of these efforts.

because we are unable to detect judge-specific errors if we have just a single measure.

Fortunately, some of the most widely used indicators of liberal democracy satisfy the above criteria. To begin, we consider Raymond Gastil's Freedom House scales of political rights and civil liberties (e.g., Gastil, 1988). Freedom House has reported annual ratings from 1972 to the present. Also, the Freedom House ratings are very widely used in empirical research (e.g., Burkhart & Lewis-Beck, 1994; Muller & Seligson, 1994; Poe & Tate, 1994) as well as used by a broad policy community. Of all subjective measures, the Freedom House ratings are the most conceptually similar to the definition of democracy. Gastil's (1988) political rights and civil liberties are very similar to the dimensions of democratic rule and political liberties. The second judge we consider, Sussman (1982), provides lesser known measures of political liberties from Freedom House. Sussman's measures are the freedom of print and broadcast media. Finally, we consider Arthur Banks's (1971, 1979) Cross-Polity Time-Series Archive (freedom of group opposition, competitiveness of nomination process, chief executive elected, and effectiveness of elected legislative body). Banks's (1971, 1979) data are widely incorporated into and similar to other measures of democracy (e.g., Arat, 1991; Gurr, Jaggers, & Moore, 1990, footnote 12; Przeworski, Alvarez, & Limongi, 1996, footnote 1). Like Gastil (1988), Banks's (1971, 1979) data are also widely used in empirical research (e.g., Arat, 1988; Bueno de Mesquita, Siverson, & Woller 1992; Gonick & Rosh, 1988). Other recent subjective measures of democracy do not meet the necessary criteria for inclusion.[4]

4. For instance, Gasiorowski (1995, 1996) reports a three-category measure of democracy (democracy/semidemocracy/nondemocracy) for developing countries. However, Gasiorowski's focus is on regime change rather than multiple indicators of degrees of democracy. Also, his rating is limited to a more restrictive sample than Freedom House and Banks. Furthermore, Gasiorowski provides only one measure, although we need at least two measures from the same source to develop an estimate of judge-specific systematic error. Przeworski, Alvarez, and Limongi (1996) also develop a dichotomous measure of democratic regime change. We do not use this measure for similar reasons. Also, because the Przeworski et al. measure (p. 53, footnote 1) incorporates information from Banks (1971, 1979), our findings for Banks's measures should have relevance to the Przeworski et al. (1996) measures. Gurr's and colleagues' (e.g., Jaggers & Gurr, 1995) Polity III is another recent data set with an index of democracy based on subjective ratings of several component variables. The problem with the component measures for our purposes is that they do not have a monotonic relationship to democracy. Some are nominal categories. For example, "openness of executive recruitment" has four categories: election, dual (hereditary/election), dual (hereditary/designation), and closed. The ordering of these categories with respect to democracy is ambiguous. In addition, the categories of each indicator are assigned weights before going into the composite democracy score so it is difficult to determine the values that are appropriate if one were to use each component as a distinct indicator. Also, using the democracy

In sum, we will use eight indicators of political liberties and democratic rule that come from three judges: Gastil, Sussman, and Banks. We hypothesize that each judge brings his own "method factor," or judge-specific component, to his judgment of democracy, and this method factor systematically influences the ratings he produces.[5] So, although a subjective measure might be reliable ("measuring something consistently and dependably" [Singleton, Straits, Straits, & McAllister, 1988, p. 111]), the reliability could be partly due to the judge-specific method effect. That is, the reliability of a measure is probably partly due to validity—as an accurate reflection of democracy, but it could also indicate the presence of other consistent factors like biases. The judge-specific method effect can produce a reliable (but not valid) contribution to the rating of democracy.

In the next section we explain our reasoning further and describe the processes that might lead to judge-specific method factors in subjective measures of democracy. Our goal is to better understand how the characteristics of judges might produce judge-specific error.

JUDGE-SPECIFIC ERRORS IN SUBJECTIVE MEASURES

To understand the origins of the judge-specific method factors, consider first the following definition of judgment: A judgment or an evaluation is the act of putting a value on some object (Rossi & Anderson, 1982). It is useful to break the rating process into three parts: (a) obtaining information for rating, (b) the judge's processing of this information, and (c) the method by which a judge's processing decisions are translated into a rating. Consider the three parts sequentially. First, available information about the object must be

index alone would give us only one measure from Gurr and not allow us to estimate the presence of systematic error, if any. Their "autocracy" measure is developed from the same component variables, although it is supposed to measure a concept distinct from democracy. Given these complexities it is not clear how to model these variables. Fortunately, Banks's, Jaggers, and Moore (1998) indicators will indirectly provide information relevant to the Gurr et al. measures because they have similarities in the component indicators. For example, although not identical, there is considerable overlap in Banks's "party legitimacy" variable and Gurr's "competitiveness of political participation" variable. Similarly, we use Banks's "legislative effectiveness" indicator and it is analogous to Gurr's "constraints on chief executive" measure. Indeed, Gurr et al. (1990, p. 103, footnote 12) recommend Banks's archive data as a way to check the variable codings in the polity data set.

5. We use *judge-specific method factor* and *judge-specific method effect* interchangeably. Each term refers to systematic error that is present in a measure and that is specifically associated with the judge who creates the rating.

gathered, then a judge must make an evaluation of the object on the basis of the information, and finally that judge must translate his or her evaluation into an actual value or rating for the object. Judge-specific method effects can enter ratings at any point in the three stages.

In democracy ratings, method factors could first enter ratings through the systematic distortion of information as it makes it way from the country of origin to the publications, largely Western, that are accessible to the expert judges. Because this "filtering of information" is biased, the judges are likely to rely on nonrepresentative information when forming their ratings of a country. Method factors could be created through differences in the amount and type of information judges receive (Bollen, 1986, 1993).

Once a judge has information on a country, he or she then evaluates it to form ratings. In this step, the way in which one judge's *use* of available information differs from another judge's could cause a judge-specific method factor. In the processing of information, the judge must discriminate between events relevant and those irrelevant to the democracy rating (valid and invalid pieces of information). In the determination of relevant and irrelevant events, we see two interesting possible outcomes: A judge can *incorporate only relevant events* into his or her rating of democracy, or he or she can *incorporate both relevant and irrelevant* pieces of information into the rating.

Consider first the incorporation of relevant information into democracy ratings. Even if an event corresponds to the conceptual definition of liberal democracy and a judge incorporates it, method effects might enter a rating through how judges weight that piece of information. For example, a country's political system might exclude from participation "extremist groups" that have a small following. Two judges could agree that this fact is relevant to liberal democracy. However, one judge's rating might give this little weight because of the small number of people affected, whereas the second judge might take any restriction on political organizations as a serious violation of the liberties that should be present in a liberal democracy. If a judge uses a consistent pattern of weighting that differs from other judges, it could create a judge-specific method factor.

Another way that judges can create systematic error in the evaluation process is through the incorporation of external traits that are irrelevant to the construct of democracy in the rating (incorporating irrelevant information). Here, we refer to traits of a country that do not directly fall under the concept but that influence a judge's rating. For instance, suppose the age of a country or its dominant religion shades the rating of liberal democracy of a country for a judge. Although age and religion could affect the level of democracy, such traits are not part of the theoretical definition and are therefore external to the concept of democracy. A judge who implicitly considers these or

related traits would therefore rely on extraneous information in forming the ratings and create a judge-specific method factor.

In the final step of the rating process, judges can differ in their methods of measurement or the way they translate their evaluations into actual ratings of democracy. For example, judges can make three specific decisions when creating ratings: The judge can rate countries individually or with a panel, the judge can rate countries each year at the same time or every few years, and the judge can either compare countries side by side or use an abstract scale as a standard. Judges who make different decisions on these measurement issues could cause their own method factors to appear in their ratings of democracy.

To summarize, judge-specific method factors could appear for three reasons. First, differential use of sources of information, combined with the filtering of information across the world, could lead to specific judge-centered method factors. Second, judges can process the information available to them in such a way as to differentially weight relevant events or to include irrelevant factors. Finally, the methods of constructing a measure might introduce method effects.

In the next section we present a model that allows a test of whether judge-specific method factors are present in indicators of democracy during a 17-year period. After that, we will test these hypotheses about the origins of the method factors during the same time period.

DATA AND MODEL OF DEMOCRACY

Figure 1 presents the hypothesized relation between the substantive latent variables of political liberties and democratic rule; the judge-specific method factors for Sussman, Gastil, and Banks; the observed ratings; and the random errors of measurement.[6] Generally, each indicator is composed of three components: one due to the substantive variable it measures (political liberties or democratic rule), a judge-specific method factor (Sussman, Gastil, or Banks method), and a random error of measurement. Four of the observed measures tap political liberties and four gauge democratic rule. The judge-specific method factor enters the measures for the reasons we described in the last section. Gastil's and Banks's measures are available for every year from 1972 to 1988, whereas Sussman's variables are available for 1979 to 1981 and

6. The diagram conforms to path analysis conventions, where ellipses represent latent variables, boxes enclose observed variables, and disturbances (random errors) are not enclosed. Single-headed straight arrows stand for the influence of a variable at the base of the arrow on the variable at its head. Curved two-headed arrows show unanalyzed associations between the variables that are connected.

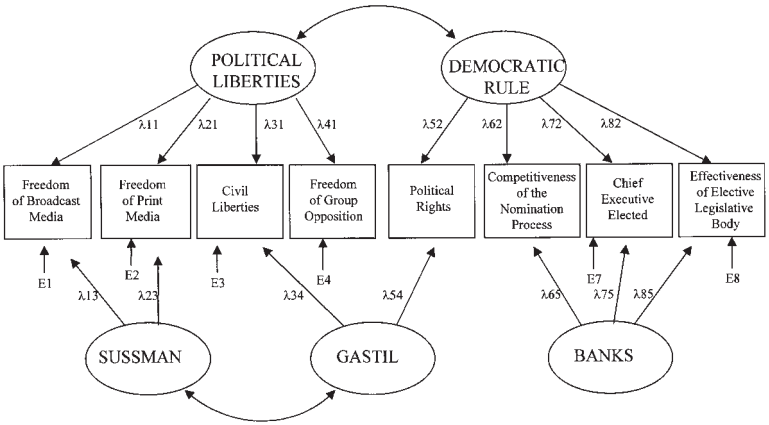

Figure 1. **Model with method factors.**
Source: Sussman (1982); Gastil (1988); Banks (1971, 1979).

1983 to 1987. Each of the rating scales is rescored to range from 0 to 10, with higher values signifying greater democracy.

The typical equation for an observed variable from the model is

$$x_i = \lambda_{ij}L_j + \lambda_{ik}M_k + \delta_i,$$

where i = 1, 2, . . . , 8; j = 1, 2; and k = 3, 4, 5. The x_is are the eight measures of liberal democracy. The L_j is the latent variable for either political liberties or democratic rule; M_k is the latent variable for either the Sussman, Gastil, or Banks method factor. The λ_{ij} and λ_{ik} are coefficient parameters that give the effects of L_j and M_k on x_i. Random measurement error is represented by δ_i. We assume that δ_i is uncorrelated with L_j and M_k and that $E(\delta_i)$ is zero.[7]

7. Previous research that considered this model from 1979 to 1981 found that Banks's freedom of group opposition was relatively free of the judge-specific method factor for Banks, that Gastil's political-rights variable and Banks's competitiveness of the nomination process had negligible random measurement error, and that only the Gastil and Sussman judge-specific method factors had a substantial correlation (Bollen, 1993). The correlation is consistent with Sussman's and Gastil's opportunity to communicate about their ratings because they were at Freedom House during the same time. These findings are incorporated into the model as represented in Figure 1. We used algebraic means to demonstrate that the model is identified; each free parameter was written as a function of the variances and covariances of the observed variables. For the years without Sussman's data, the model looks similar to Figure 1, with Sussman's method factor and indicators removed. In that model, there are only two measures for political rights, four for democratic rule, and two method factors, Gastil's and Banks's. To reduce confusion, in discussions of the models without the Sussman variables, we make the variable names explicit. The model is algebraically identified.

A previous analysis found the significant presence of method factors in the years 1979, 1980, and 1981 (Bollen, 1993). In the next section, we consider whether the method factors are consistent during a much broader time period.

REPLICATION ANALYSES

A major limitation of previous work is that it has only considered 1979 to 1981 data. In this section, we determine whether the method factors present in circa 1980 replicate throughout the period 1972 to 1988. Replication is a rather general term and is a matter of degree. Our replication analyses address the following questions: (a) Do the method factors appear in the ratings during this 17-year period? (b) Are the parameter values from a circa 1980 analysis stable during the whole period? If the parameter values are stable during the entire time period, this is an indication that the judge-specific component of the ratings has a consistent effect over time—it is a stable influence on the ratings. Condition (b) is obviously a stronger condition of replicability than is (a). If we answer "no" to (a), there is little point in continuing.

To address Question (a), we estimate two models. The first one is the whole model as represented in Figure 1. We also estimate a second model that eliminates the three judge-specific method factors so that each variable is a function only of the substantive variable (political liberties or democratic rule) and a random error of measurement. If there are no judge-specific method factors in these measures of liberal democracy, then the fit of the second, simpler model should be essentially the same as that of the first model. Alternatively, a nonnegligible improvement in fit with the inclusion of method factors is evidence favoring their presence.

We use confirmatory factor analysis to fit both models (see Bollen, 1989; Jöreskog & Sörbom, 1993). We applied the maximum likelihood fitting function to estimate the models for each year from 1972 to 1988.[8] Fit statistics

8. The maximum likelihood estimator assumes that the observed variables come from a distribution that has multivariate kurtosis that is the same as a multinormal variable. We found evidence of multivariate kurtosis for these data (Bollen, 1989, pp. 423-425). Although the estimator remains consistent, the presence of excessive kurtosis raises the possibility that the chi-square tests of overall model fit and the other significance tests may have inaccuracies. However, recent work suggests that the test statistics from the maximum likelihood estimator may be robust under some conditions of multivariate kurtosis (see Bollen, 1989, chap. 9; Satorra, 1990). Establishing whether the robustness conditions hold is an unresolved issue in structural equation modeling. As a precaution and as a check on our tests of model fit, we applied bootstrapping techniques (Bollen & Stine, 1993). Furthermore, parts of our assessments of fit are based on fit indices that are not based on significance testing.

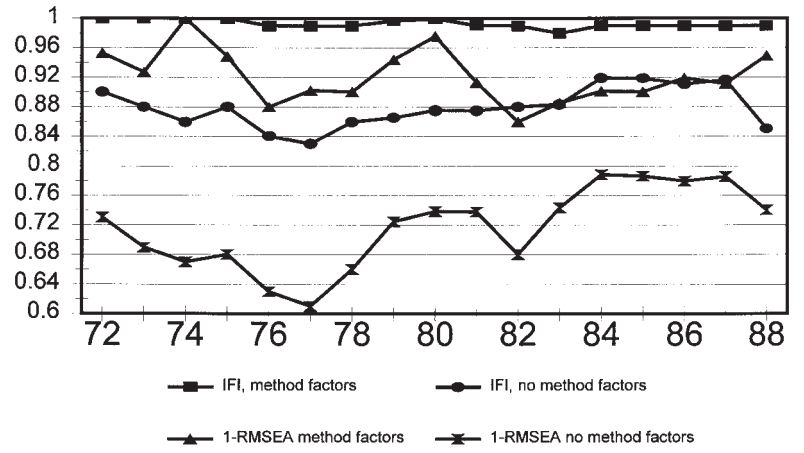

Figure 2. **Method versus no method, Incremental Fit Index (IFI), and 1-RMSEA (root mean square error of approximation).**
Note: Higher values indicate better fit.

provide an estimate of the fit between our data and our models or, specifically, between S, the sample covariance matrix, and $\Sigma(\theta)$, the implied covariance matrix at the final estimates, $\hat{\theta}$. A comparison of all fit measures for the model in Figure 1 and the model without method factors consistently shows the superior fit of the model with method factors. Details of the fits of each of these models for the 17-year period are given in the appendix. As an illustration, Figure 2 plots two measures of fit for the two models across the 17-year period. The Incremental Fit Index (IFI) (also called Δ_2; Bollen, 1989) and 1-RMSEA (root mean square error of approximation) (Steiger & Lind, 1980) are both scored so that values close to 1 signify superior fit, and for all years, the judge-specific method factor model has a value that is better than the simpler two-factor model. It is easy to see from the graph that there is an improvement in fit gained by introducing the method factors for all the years, 1972-1988. Overall, this provides very strong evidence in support of the continual presence of judge-specific method factors in these ratings of political democracy.

Not only is there evidence for the continual presence of method factors across the years; the method factors have a nonnegligible impact on the ratings. For each year we can estimate the percentage of the variance in an indicator that is due to the substantive variable (political liberties or democratic

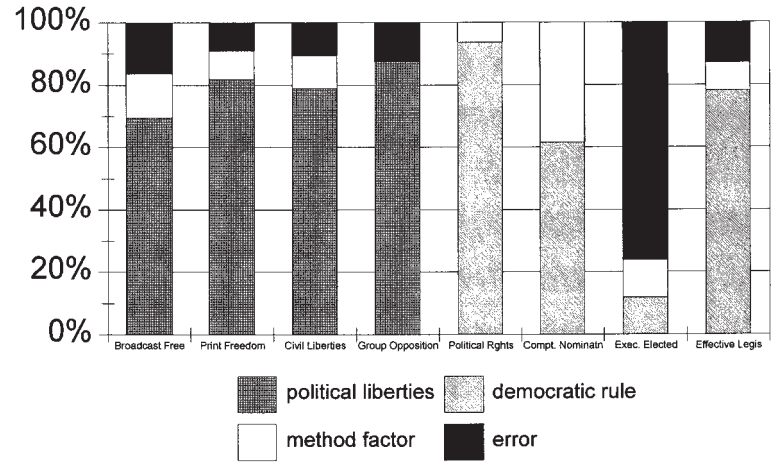

Figure 3. **Breakdown of variance in indicators.**

rule), the judge-specific method factor (Sussman, Gastil, or Banks), and random error of measurement. Figure 3 displays the mean percentage of variance due to each component for each variable during the period 1972 to 1988 (excluding the base years 1979 to 1981). All but one of the indicators have substantial percentages of variance due to the judge-specific method factor. The most extreme is Banks's competitiveness of the nomination process, which has nearly 40% of its variance due to the method factor. Several of the other measures (print media freedom, civil liberties, chief executive selection, and effectiveness of the legislative body) have roughly 10% due to the method factors. At the same time, some of the variables have impressive percentages of variances due to the substantive variables they were supposed to measure. Banks's freedom of group opposition has about 88%, whereas Gastil's political rights is even higher at 94%. Overall, we find that judge-specific method factors have nonnegligible effects, yet some indicators are simultaneously linked closely to their substantive latent variable. Put another way, the indicators do have an element of validity with the concept of democracy but also incorporate nonvalid elements over the years. The next part of the replication will determine the stability of these components over time.

Our second question about replicability asks whether the specific parameter values for these models are stable during the 17-year period. To address this issue, we took the average (mean) of each parameter estimate during

three base years, 1979, 1980, and 1981. The averaging provides base values for the factor loadings, λ (lambda), the covariance matrix of the latent variables, Φ (phi), and the variances of the error variables, Θ_δ (theta-delta).[9]

To test the replicability of the average values, we form a hierarchy of models where, for each year, we estimate four models. These are the models with no restrictions on the values of the free parameters (FORM), with λ's (factor loadings) restricted to equal the average λ (LAMBDA), with λ's and ϕ's (variances and covariances of latent variables) restricted to equal the average of their 1979-1981 counterparts (LAMBDA-PHI), and finally with λ's, ϕ's, and Θ_δ's (variances and covariances of errors of measurement) forced to be equal to the average values (LAM-PHI-DELTA).

The more restrictions that we can introduce without suffering a severe decline in fit, the greater is the degree of replication. This procedure is analogous to the testing of parameters in multiple group analysis (see, e.g., Jöreskog & Sörbom, 1993, chap. 9), except that here we are restricting parameter matrices to specific values. The models resulting from these restrictions are nested, so we can construct chi-square difference tests to assess the statistical significance of the decline in fit when moving from one model to a more restricted one. We can form tests for 14 years once we exclude 1979-1981 as the base years. The test for a significant difference in the fit when the factor loadings are the same as the circa 1980 values is statistically significant at a *p* value less than .05 for 5 of the 14 years; the test of the model with restricted factor loadings and the restricted covariance matrix of the latent variables being the same relative to just the restricted factor loadings is statistically significant for only 1 of the 14 years; finally, the model that restricts all parameters relative to the one with the factor loadings and the covariance matrix of latent variables restricted is significant for 8 of the 14 years.[10] These results suggest some evidence in favor of stable values for the factor loadings and the covariance matrix of the latent variables for most years. The evidence for the additional constraints on the covariance matrix of the errors is more mixed. Given the small to moderate sample size and the nonnormality of the data, however, we need to consider other ways of comparing the parameters over time.

Another perspective on the relative fit comes from the graph of the IFI for these models over time (see Figure 4). Note that with the exception of the

9. These correspond to the years analyzed in Bollen (1993). The average values were λ11, .83; λ21, .90; λ31, .69; λ41, 1; λ52, 1.08; λ62, .91; λ72, .40; λ82, 1; λ13, 1; λ23, 1.22; λ34, 1; λ54, .87; λ65, 1; λ75, .71; λ85, .48; Φ11, 17.48; Φ21, 13.01; Φ22, 10.18; Φ33, 2.58; Φ43, 1.64; Φ44, 1.67; Φ55, 5.34; Θ_δ11, 3.10; Θ_δ22, 1.33; Θ_δ33, .76; Θ_δ44, 1.06; Θ_δ77, 8.75; Θ_δ88, 1.54.

10. See appendix for detailed information on the fit statistics across the constrained models.

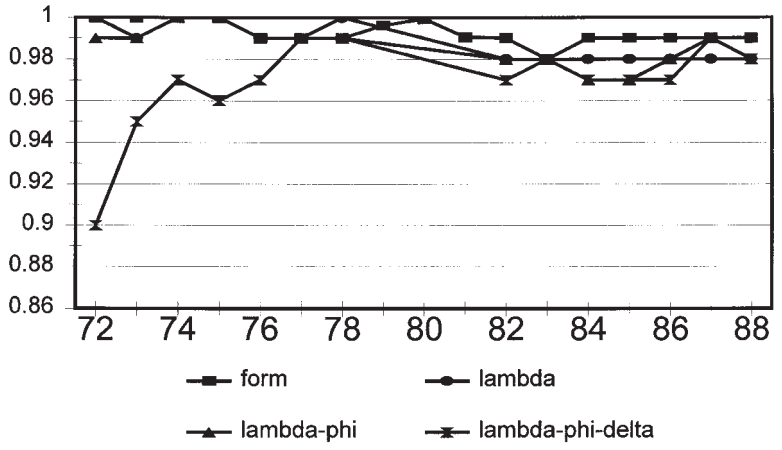

Figure 4. **Incremental Fit Index (IFI) for restricted models.**

1972 and 1973 models that impose all parameters to equal the circa 1980 constants, all models have IFI values of .96 or better. By conventional standards, these are relatively high IFI values. (The results using 1-RMSEA are similar, except that the models with constrained parameters often fit better than models that only impose the form of the structure.) These results suggest that any decline in fit that results from imposing the same parameter values is modest for most years.[11]

11. Considering the variation in the graphs, the degree to which we can impose restrictions does not appear to be constant across the years. The years 1972 to 1974 have the greatest discrepancies in goodness of fit. By considering these specific years more carefully, we can gain a greater sense of the degree of the *practical* difference in the models with and without restrictions on the parameters. First, we consider the mean difference between the observed correlation and the model-predicted correlation of the observed variables. This "correlation residual" (Bollen, 1989, p. 258) differs only by .024 or less between the models with all versus no parameter equality restrictions for these 3 years. Second, we consider the parameter estimates. The largest difference between an unrestricted lambda and a restricted lambda occurs for the path from BANKS to executive elected (λ_{54}) and is .31, a 44% difference. Other lambdas are a maximum of .13 apart. Similarly, the biggest absolute difference between the restricted and unrestricted ϕ's is 3.41 for ϕ_{11}, although larger percentage differences occur for the variance of the method factor for Gastil (1988) (ϕ_{33}). The largest difference in the theta delta (Θ_δ) matrix is 4.14 for VAR(δ_5) (executive elected). The largest percentage differences occur in the variances of the errors of measurement. Overall, these numbers generally indicate only small deviations in the values of the factor loadings and covariance matrix of the factors, even in the years with the worst fit.

In summary, we find three things: (a) There is unambiguous evidence to support the model that includes judge-specific method factors during the expanded time period 1972-1988. (b) The parameter estimates for the model with the method factors are fairly consistent across the period, with the greatest discrepancies appearing in the error variances. (c) Even the years with the worst fit in the parameter replication show little practical difference in their parameters. Points (b) and (c) mean that the effect of the judge-specific method factors remains constant in the ratings over time (as well as the effect of the valid substantive component). The judges are creating consistent, systematic error in the ratings of democracy.

DETERMINANTS OF METHOD FACTORS

The replication section shows that method factors are continually present in measures of democracy and have consistent effects. The next step is to understand which of the earlier hypothesized reasons for judge-specific method factors (information, evaluation, measurement construction) best explains their presence. Given the similarity in the methods of constructing measures across the judges, we are less concerned with this element of judge-specific error in democracy ratings (although it could affect other cross-national measures). Therefore, we examine only two of the factors that might lead to systematic method factors: the amount of information available on a country and the judge's processing of this information (which might include extraneous factors). To compare these two hypotheses, we estimate the judge-specific method factors of the democracy ratings and regress those values on traits of countries related to each hypothesis. Significant coefficients will indicate that factors outside of the construct of democracy are entering into the ratings through the method factors.

We begin by estimating the method factors associated with each judge by using the regression method of estimating factors:

$$\hat{L} = \hat{\Sigma}_{LL} \hat{\Lambda} \hat{\Sigma}_{xx}^{-1} x,$$

Where $\hat{L}$ is the matrix of estimated latent variables (or factor scores). The variables to the right of the equal sign are the estimated covariance matrix for the latent variables, the estimated factor loadings, the estimated covariance matrix of the observed variables, and the deviation form of the observed variables.[12]

12. The values of the method factor estimates come from the factor analysis in which all parameters are restricted to equal the average value of the 1979 to 1980 solutions. To check

This procedure led to three estimated judge-specific method factors, one for each judge. We divided the estimated latent variables by their sample standard deviations in each year to create standardized variables in standard deviation units. A positive standardized score indicates a judge's tendency to overestimate the degree of political liberties or democratic rule in a country, whereas a negative value suggests the opposite.

VARIABLES

Our model introduces variables from the two relevant hypotheses: information filtering and extraneous factors. To test the influence of information availability, we included several variables that should be related to the accessibility of information about a country. The variables included the number of times that a country appeared in articles in the *New York Times Index*, *Facts on File*, and the *CBS* or *Vanderbilt Television News Index* (1980, 1984, 1988); log of population size; log of land area; and log of radios/TVs per 1,000 people (all were logarithmically transformed to lessen their skewness). Population size, land area, and radios/TVs were obtained from Taylor & Jodice, 1983. These variables will test whether the extent of information and visibility of a country were related to the method factors.

The second hypothesis—that extraneous, irrelevant events enter into a judge's evaluation process—requires the elaboration of two subhypotheses: (a) that judges will be influenced by how situationally and personally similar a country is to them, and (b) that judges will be influenced by the "seriousness" of events. One potential source of extraneous information in a judgment is how situationally or personally similar the subjects are to the judge (Burger, 1981). This source of bias hinges on responsibility; if persons attribute more responsibility to an object, they tend to assign it more "blame." Objects that are closer to a judge tend to receive less responsibility for events and therefore more favorable judgments. In rating democracy, personal philosophy or ideology could enter the ratings through this mechanism. If a judge were ideologically leftist, then that judge could feel situationally closer to Marxist-Leninist countries and rate them more favorably. Or, judges in Western industrial societies such as the United States could feel situationally closer to Western industrialized nations and rate them accordingly.

For variables related to this type of extraneous factor, we consider those that represent a left-to-right distinction among countries. These variables are

---

whether this restriction had any influence, we also estimated the latent variables from the solutions that kept the model parameters unrestricted for each year. The results were essentially the same as those reported in the text.

intended to measure the situational closeness of certain types of countries to the judges. The ideologically closer that a judge feels to a country, the more favorable we expect him or her to be in ratings of democracy. In addition to left/right variables, we include variables to capture the similarity of a country to Western industrial countries. Again, the judges are expected to feel closer to such countries. Collectively, the variables capture aspects of stability, westernization, development, and traditionalism—factors that could feed the judge-specific method biases in the ratings.

The situational-closeness portion of the model consists of seven variables. Four of these are dummy variables that code whether (a) a country is Marxist-Leninist, (b) predominately Catholic, (c) predominately Protestant, or (d) a monarchy. The coding of Marxist-Leninist is based on Lenski and Lenski's (1982) classification but is changed for some countries in some years based on historical accounts. Protestant and Catholic are coded from religious data in *The World Almanac and Encyclopedia 1990*. Monarchy comes from information in *Spotlight on World Political Systems* (Derbyshire & Derbyshire, 1991) and *Elections Since 1945* (1989). Three nondichotomous variables are also included in the model: (e) The natural logarithm of energy consumption per capita from *The UN Statistical Yearbook* (United Nations, 1991) gauges the level of industrialization, (f) the natural logarithm of the number of years that a country was independent, measures the "age" of a country (Jackman, 1993), and (g) the number of coups d'état for the year of interest and four previous years from Ted Gurr's (1990) Polity II data set.

The second source of extraneous information involves the idea of defensive attribution. Defensive attribution is also tied to responsibility in that judges tend to assign more responsibility to actions with serious consequences (Shaver, 1970). This attribution error has implications for democracy ratings because some actions of countries can have stronger human rights connotations than others. Thus, a state that takes violent action against a small group of voters might be rated much lower in democracy than one that benignly denies a large proportion of its population suffrage. Or, judges might rate countries involved in wars or military actions more harshly than others.

In our model we include a number variables related to defensive attribution. We expect that countries involved in wars or experiencing political strikes, protests, or riots would be rated differently by the judges because those actions have harsh connotations. Political strikes, riots, and protests were taken from Muller (1988) and logged, whereas interstate and civil wars were taken from Small and Singer (1982).

RESULTS

All the variables mentioned in the last section were available for 1980 so we began the analysis with that year. We regressed the estimated factor scores on the three sets of variables (situational closeness, defensive attribution, and information) to determine the impact of various traits of countries on the judge's method factors in the democracy ratings of 1980. The results appear in Table 1 and show that the variables related to situational closeness hold the most promise for explaining the method factors. Outside of the effect of population on Banks, all of the significant parameters were associated with the situational-closeness hypothesis. Because our sample size was modest (N = 81), we screened for outliers. The general finding that the variables related to situational closeness were the most important was not sensitive to these outliers.[13]

Because there was a slight indication that a few of the information and defensive-attribution variables might be important in determining the method factors, we performed a sensitivity analysis on our results. We considered variables for the three hypotheses in a few more years (the information variables in 1984 and 1988 and the defensive-attribution variables in 1975). None of the information or defensive-attribution variables show consistent significance in those years. The situational-closeness variables did show a consistent pattern in those years, however. These results indicate that of the competing hypotheses (information availability vs. the inclusion of extraneous factors), extraneous factors seem to be the explanation for the method factors. And, of the possible extraneous factors, those associated with the situational closeness of a country to a judge were more important than those associated with defensive attribution.

Because the situational-closeness variables affect the judge-specific method factors of Gastil, Sussman, and Banks, we present their effects in a number of years across the time period—1972, 1975, 1980, 1984, and 1988. Fortunately, these variables are available for a larger number of countries during these time periods, so that our sample size is substantially increased

13. In the Gastil regression, if we take out Kenya, Uganda, Honduras, Bangladesh, Guatemala, Mongolia, and Sudan, only the coefficients for Marx/Lenin and coverage change more than their standard error and neither of those change in significance. For the Sussman regression, if Afghanistan, Thailand, Uganda, Kenya, Bangladesh, Honduras, and Mongolia are removed, several coefficients change but only Marx/Lenin loses its (marginal) level of significance. For the Banks regression, if Uganda, Kenya, Bangladesh, Nicaragua, Yugoslavia, Romania, Turkey, and Ethiopia are removed, the coefficients for Protestant and Catholic decline and lose their significance, the coefficient for log of protests decreases and becomes significant, whereas the log of riots increases and becomes significant, and the coefficient for the log of population decreases and loses significance.

Table 1
*Testing Three Hypotheses in 1980 (*N = *81)*

	Gastil	Sussman	Banks
Intercept	–0.817	–0.494	1.922***
	(.646)	(.614)	(.632)
Situational closeness			
Marx/Lenin	–0.976***	–0.504*	1.366**
	(.294)	(0.3)	(.53)
Log energy per capita	–0.122	–0.063	0.124
	(.126)	(.123)	(.177)
Protestant	0.357	0.459	0.670**
	(.324)	(.295)	(.264)
Roman Catholic	0.835***	1.239****	0.676**
	(.312)	(.331)	(.286)
Log of years since independence	0.152	0.084	–0.515***
	(.153)	(.153)	(.16)
Monarchy	0.343	0.131	–1.441****
	(.263)	(.282)	(.268)
Coups	0.349	0.380	–0.615***
	(.268)	(.287)	(.202)
Defensive attribution			
Wars	–0.360	0.022	–0.234
	(.302)	(.328)	(.438)
Log of protests	0.098	0.068	–0.174
	(.121)	(.139)	(.109)
Log of political strikes	–0.074	–0.080	0.211
	(.137)	(.158)	(.135)
Log of riots	–0.200	–0.023	0.075
	(.142)	(.148)	(.148)
Information			
Coverage	0.019	0.002	–0.034
	(.036)	(.037)	(.032)
Log of population	0.155	0.029	0.279**
	(.117)	(.139)	(.122)
Log of land area	–0.050	–0.025	–0.024
	(.058)	(.061)	(.076)
Log of radios and TVs	0.067	0.022	–0.105
	(.075)	(.067)	(.102)
Adjusted R^2	.24	.26	.29

Note: Numbers in table indicate coefficient estimates, with standard errors in parenthesis. All standard errors are heteroskedastic consistent.
* $p < .10$. ** $p < .05$. *** $p < .01$. **** $p < .001$.

Table 2
Coefficient Estimates and Standard Errors for Factor Score Regression

	Gastil					Sussman		Banks				
	1972	1975	1980	1984	1988	1980	1984	1972	1975	1980	1984	1988
Intercept	–0.51	–0.48	–1.19****	–0.75	–1.03***	–0.99***	–1.19****	–0.39	–1.25****	0.82**	0.92	1.04**
	0.33	0.30	0.29	0.42	–0.32	0.29	0.34	0.37	0.36	0.38	0.48	0.47
Marx/Lenin	–1.16****	–0.87****	–0.88****	–0.43***	–0.38	–0.72****	–0.67****	0.37	0.38	0.82***	0.84***	0.99****
	0.21	0.18	0.18	0.16	0.21	0.18	0.15	0.27	0.26	0.26	0.29	0.26
Log energy per capita	0.06	0.03	0.09	–0.03	0.01	0.04	0.01	0.10**	0.17***	–0.04	–0.08	–0.06
	0.06	0.04	0.05	0.04	0.04	0.05	0.04	0.05	0.06	0.06	0.07	0.06
Protestant	0.24	0.51**	0.24	0.51**	0.65****	0.15	0.57***	–0.21	–0.29	0.28	0.23	–0.01
	0.26	0.22	0.24	0.22	0.18	0.20	0.18	0.21	0.22	0.23	0.27	0.24
Catholic	0.40	0.73****	0.54***	0.67***	0.58****	0.67***	0.59***	–0.01	–0.22	0.16	0.21	0.16
	0.23	0.19	0.21	0.23	0.18	0.24	0.22	0.25	0.23	0.18	0.21	0.18
Log years independence	0.03	0.01	0.12	0.18	0.17**	0.14**	0.22***	–0.05	0.08	–0.15**	–0.14	–0.17**
	0.06	0.05	0.06	0.11	0.08	0.07	0.08	0.06	0.07	0.07	0.10	0.08
Monarchy	0.08	0.50	1.13****	0.75**	1.23****	0.39	1.34****	0.04	–0.34	–0.89***	–0.47	–0.51
	0.25	0.31	0.27	0.32	0.27	0.25	0.33	0.26	0.41	0.28	0.38	0.33
Coups	0.02	0.01	0.39	0.12	0.10	0.46	–0.03	–0.58***	–0.30	–0.54***	–0.61**	–1.14****
	0.15	0.19	0.21	0.27	0.24	0.24	0.23	0.18	0.20	0.19	0.24	0.28
N	129	129	146	130	144	146	130	129	129	146	130	144
Adjusted R^2	.14	.23	.26	.16	.24	.22	.30	.12	.12	.15	.11	.18

Note: All standard errors are heteroskedastic consistent.
**$p < .05$. ***$p < .01$. ****$p < .001$.

by using only these variables (from 80 countries to 130-150 countries). Table 2 presents the regressions.

Consider the estimates for the Gastil method factor first. Marxist-Leninist countries are consistently negatively related to the estimated method factor. The downrating seems most severe in the early years and declines some over time, but even in 1988, a Marxist-Leninist country has an expected .38 standard deviation lower rating on the method factor than does a non-Marxist-Leninist country net of the other variables. Also evident is the effect of religion. Christian countries have a positive impact on the Gastil method factor. Although somewhat variable, it is usual for a Protestant or Catholic country to receive a .5 or more standard deviation boost in the method factor. The strongest effect of the age of a country appears for the more recent years and is positive. Traditional monarchies also fare well on the Gastil method factor, sometimes increasing the method factor by more than a standard deviation. Although generally positive, the effects for coups and industrialization are considerably less than those of the other variables. Overall, these results give the impression that the Gastil method factor tends to favor countries that are non-Marxist-Leninist, Christian, monarchies, and older.[14] Sussman's data were only available for 2 of the 5 years. The pattern of effects on the Sussman method factor are quite similar to those we described for the Gastil method factor.[15]

Banks's estimated method factor has a different profile. The strongest effect is a positive relation between Marxist-Leninist countries and Banks's method factor. This result would be consistent with the general finding that political scientists are more favorable to leftist politics than is the general population. In the 1980 and later regressions, such countries receive a .8 standard deviation or greater boost in the Banks method factor. Coups also have a substantial effect. However, the impact on the method factor is negative. Although less strong, there is some tendency for older countries and monarchies to be down rated, as is reflected in their mostly negative coefficient estimates.[16]

14. Because these regressions have new cases and a new set of variables compared with the previous ones, we need to recheck our regression results for outliers that might affect our findings (see Bollen & Jackman, 1990, for procedures). We found only a few changes when outliers were removed. In 1984, removing Nepal shifts the log of years since independence and monarchy's coefficients down, making the log of years since independence not significant. In 1980, if you exclude South Africa and Uganda, Protestant becomes significant.

15. Outlier diagnostics and checks did not reveal any influential countries.

16. Outlier diagnostics revealed that the removal of Yugoslavia, Cuba/Yugoslavia, Cuba, and Romania led the Marxist-Leninist variable to be significant or nearly so for 1972 and 1975. In 1975, removal of Portugal and Cyprus significantly increases the absolute magnitude of the coups variable. The significance of the log-of-years-since-independence variable is sensitive to the presence or absence of several countries in 1980. In 1988, the monarchy variable increases in significance, and independence decreases when a few outliers are excluded.

CONCLUSIONS

The goal of this analysis was to learn more about subjective measures of liberal democracy. What we found is not comforting. For the 17-year period from 1972 to 1988, judge-specific method factors account for a nonnegligible portion of the variance in most measures. Furthermore, the effects are essentially the same during the 17-year period. Although we cannot be sure how greatly these biases affect empirical analyses, our results reveal the potential for problems. To the extent that extraneous traits, such as age of the country and religion, are associated with the method factors, they are part of our measures of democracy. Use of such measures might distort our findings. For example, the association between coups and the method factors is disturbing because the cross-national analyses relating democracy and coups could be influenced by a democracy measure that has a built-in association with coups. Similar confounding could occur for other democracy measures. Alternatively, it is possible that the biases are insufficient to fundamentally shift conclusions, but we do not have evidence to support that claim.

A partial solution to the problem is to take account of the extraneous variables in democracy measures by including them as explanatory variables in a regression equation. Although this is an improvement over ignoring the problem, the approach has limitations. If democracy is the dependent variable and one of the determinants we wish to study is an extraneous variable confounding the democracy measure, we will not be able to determine whether an estimated effect is due to measurement confounding or is a real substantive effect. For instance, if Protestantism predicts democracy, we will be unsure whether this is an artifact or real if Protestantism is tied to the method factor. Another limitation occurs when democracy is an explanatory variable and the dependent variable is one of the extraneous variables tied to the method factor. Here we cannot include this extraneous variable in a regression equation to predict itself. A final limitation of controlling for the extraneous variables is that it is doubtful that we have identified the full list of confounding variables. So, those variables that escaped our detection will remain part of the democracy measures and potentially bias results.

Another strategy is to develop indices of democracy that have a weaker relation to the method factors than is true of the individual variables that enter the composite. This was done for 1980 in other research (Bollen, 1993), and we are investigating this possibility for all the years. One limitation is that ultimately, the composite would still be made up of subjective measures. We

would hope to find that the part of the measures that is common across judges is close to our construct of liberal democracy. Yet, that fact is difficult to establish because we do not have a "gold standard" of democracy with which to compare the subjective composite.[17]

Those researchers who wish to pursue the creation of subjective measures should be aware of the potential for bias in their ratings. This analysis has shown that, in democracy ratings, the most likely source of bias comes from the inclusion of extraneous factors (ratings of other constructs could have different sources of bias). It is possible that using a panel of judges for ratings, especially judges with diverse orientations or experiences, could reduce the amount of bias in grouped panel ratings.

An alternative line of research is to return to "objective" measures of liberal democracy. This tradition, launched by Lerner (1958) and Cutright (1963), continues today (Vanhanen, 1990). The two main challenges to this effort are finding objective variables that correspond to the definitions of liberal democracy and determining how to weight such objective variables. An example of the first problem is the use of voter turnout as an objective measure. Although turnout is largely objective, it taps many things that are unrelated to democracy (Bollen, 1980, 1990). A final promising route to improving measurement is to look for ways to combine the information from subjective and objective measures of democracy. All of the above solutions suggest that there are alternatives to using unadjusted subjective measures of liberal democracy.

17. The issue involves the difference between consensus and accuracy (e.g., Kenny, 1991). Consensus refers to the agreement among judges, whereas accuracy (validity) refers to the closeness to the phenomenon being measured. Without further information, we cannot assess the accuracy of the judges.

APPENDIX
Goodness-of-Fit Statistics 1972-1988

Year	Model	χ^2	*df*	*p* Value	Bootstrap *p*	1-RMSEA	PCLOSEFIT	GFI	AGFI	NFI	NNFI	IFI
1972	NOMETHOD	92.48	8	0	.001	0.73	0.00000028	0.80	0.48	0.89	0.80	0.90
1972	FORM	6.54	5	.26	.996	0.953	0.44	0.99	0.94	0.99	0.99	1.00
1972	LAMBDA	15.32	12	.22	.725	0.956	0.5	0.96	0.94	0.98	0.99	1.00
1972	LAMBDA-PHI	24.02	17	.12	.444	0.946	0.4	0.94	0.93	0.97	0.99	0.99
1972	LAM-PHI-DELTA	101.98	21	0	.001	0.83	0.00000037	0.82	0.82	0.88	0.93	0.90
1973	NOMETHOD	116.12	8	0	.001	0.69	3.0E-07	0.78	0.42	0.87	0.77	0.88
1973	FORM	8.77	5	.12	.803	0.927	0.26	0.98	0.92	0.99	0.99	1.00
1973	LAMBDA	16.78	12	.16	.572	0.947	0.42	0.96	0.94	0.98	0.99	0.99
1973	LAMBDA-PHI	23.08	17	.15	.404	0.95	0.45	0.95	0.93	0.97	0.99	0.99
1973	LAM-PHI-DELTA	61.01	21	9E-06	.009	0.88	0.0011	0.88	0.88	0.93	0.97	0.95
1974	NOMETHOD	130.39	8	0	.001	0.67	0.00000035	0.76	0.37	0.86	0.74	0.86
1974	FORM	3.96	5	.56	1	1	0.72	0.99	0.96	1.00	1.00	1.00
1974	LAMBDA	9.76	12	.64	.98	1	0.85	0.98	0.96	0.99	1.00	1.00
1974	LAMBDA-PHI	15.50	17	.56	.879	1	0.84	0.97	0.96	0.98	1.00	1.00
1974	LAM-PHI-DELTA	51.59	21	.00022	.019	0.9	0.01	0.89	0.89	0.94	0.98	0.97
1975	NOMETHOD	127.49	8	0	.001	0.68	0.00000064	0.76	0.38	0.87	0.77	0.88
1975	FORM	7.00	5	.22	.976	0.948	0.41	0.98	0.94	0.99	0.99	1.00
1975	LAMBDA	14.21	12	.29	.754	0.965	0.59	0.97	0.95	0.99	1.00	1.00
1975	LAMBDA-PHI	21.36	17	.21	.517	0.959	0.56	0.95	0.94	0.98	1.00	1.00
1975	LAM-PHI-DELTA	58.20	21	2.4E-05	.007	0.89	0.0026	0.88	0.88	0.94	0.97	0.96

1976	NOMETHOD	174.57	8	0	.001	0.63	0.00000084	0.73	0.29	0.84	0.71	0.84
1976	FORM	16.06	5	.0067	.108	0.88	0.035	0.97	0.86	0.99	0.97	0.99
1976	LAMBDA	24.53	12	.017	.131	0.917	0.12	0.95	0.92	0.98	0.99	0.99
1976	LAMBDA-PHI	30.03	17	.026	.134	0.929	0.19	0.94	0.93	0.97	0.99	0.99
1976	LAM-PHI-DELTA	54.13	21	9.4E-05	.011	0.9	0.0071	0.89	0.89	0.95	0.98	0.97
1977	NOMETHOD	199.04	8	0	.001	0.61	9.0E-07	0.70	0.21	0.83	0.68	0.83
1977	FORM	12.37	5	.03	.354	0.902	0.11	0.98	0.90	0.99	0.98	0.99
1977	LAMBDA	19.01	12	.088	.4	0.938	0.32	0.96	0.94	0.98	0.99	0.99
1977	LAMBDA-PHI	25.25	17	.089	.296	0.944	0.37	0.95	0.94	0.98	0.99	0.99
1977	LAM-PHI-DELTA	31.62	21	.064	.335	0.943	0.35	0.94	0.94	0.97	0.99	0.99
1978	NOMETHOD	155.86	8	0	.001	0.66	0.00000011	0.76	0.37	0.86	0.74	0.86
1978	FORM	13.52	5	.019	.308	0.9	0.078	0.97	0.89	0.99	0.98	0.99
1978	LAMBDA	16.76	12	.16	.614	0.95	0.45	0.97	0.94	0.98	0.99	1.00
1978	LAMBDA-PHI	25.98	17	.075	.352	0.942	0.35	0.95	0.94	0.98	0.99	0.99
1978	LAM-PHI-DELTA	32.39	21	.053	.447	0.941	0.33	0.94	0.94	0.97	0.99	0.99
1979	NOMETHOD	237.95	19	0	.001	0.7247	2.8E-06	0.72	0.47	0.86	0.80	0.87
1979	FORM	19.17	13	.1179	.722	0.94411	0.3829	0.97	0.92	0.99	0.99	1.00
1980	NOMETHOD	216.56	19	0	.001	0.7385	2.8E-06	0.72	0.48	0.86	0.82	0.88
1980	FORM	14.16	13	.3626	.948	0.97576	0.6832	0.98	0.94	0.99	1.00	1.00
1981	NOMETHOD	218.38	19	0	.001	0.7381	3.1E-06	0.73	0.48	0.86	0.81	0.87
1981	FORM	28.14	13	.00864	.201	0.91274	0.0796	0.96	0.88	0.98	0.98	0.99

(continued)

APPENDIX Continued

Year	Model	χ^2	*df*	*p* Value	Bootstrap *p*	1-RMSEA	PCLOSEFIT	GFI	AGFI	NFI	NNFI	IFI
1982	NOMETHOD	137.85	8	0	.001	.68	0.00000016	0.78	0.43	0.88	0.78	0.88
1982	FORM	21.32	5	.00071	.032	0.86	0.0072	0.96	0.83	0.98	0.96	0.99
1982	LAMBDA	35.16	12	.00044	.018	0.89	0.011	0.93	0.88	0.97	0.97	0.98
1982	LAMBDA-PHI	36.59	17	.0038	.055	0.916	0.067	0.93	0.91	0.97	0.98	0.98
1982	LAM-PHI-DELTA	52.82	21	.00015	.058	0.904	0.011	0.90	0.90	0.95	0.98	0.97
1983	NOMETHOD	209.35	19	0	.001	0.7433	2.8E-06	0.73	0.49	0.87	0.84	0.88
1983	FORM	39.00	13	.0002	.02	0.8853	0.006178	0.94	0.83	0.98	0.97	0.98
1983	LAMBDA	48.14	23	.0016	.073	0.915	0.047	0.93	0.88	0.97	0.98	0.98
1983	LAMBDA-PHI	57.43	30	.0019	.076	0.922	0.07	0.91	0.90	0.97	0.98	0.98
1983	LAM-PHI-DELTA	62.56	36	.004	.146	0.93	0.13	0.91	0.91	0.96	0.99	0.98
1984	NOMETHOD	148.72	19	0	.001	0.7881	2.8E-06	0.79	0.60	0.91	0.88	0.92
1984	FORM	32.36	13	.0021	.104	0.901	0.031	0.95	0.86	0.98	0.97	0.99
1984	LAMBDA	57.84	23	7.8E-05	.009	0.9002	0.007	0.91	0.86	0.96	0.97	0.98
1984	LAMBDA-PHI	75.26	30	9.3E-06	.005	0.9004	0.0028	0.89	0.87	0.95	0.97	0.97
1984	LAM-PHI-DELTA	84.34	36	9.4E-06	.008	0.906	0.004	0.88	0.88	0.95	0.98	0.97
1985	NOMETHOD	150.75	19	0	.001	0.7864	2.8E-06	0.79	0.61	0.91	0.88	0.92
1985	FORM	34.36	13	.0011	.103	0.9	0.02	0.95	0.86	0.98	0.97	0.99
1985	LAMBDA	60.56	23	3.2E-05	.003	0.9	0.0039	0.91	0.86	0.96	0.97	0.98
1985	LAMBDA-PHI	73.93	30	1.4E-05	.005	0.902	0.0037	0.90	0.88	0.95	0.97	0.97
1985	LAM-PHI-DELTA	86.32	36	5.1E-06	.002	0.904	0.0028	0.88	0.88	0.95	0.98	0.97

1986	NOMETHOD	159.13	19	0	.001	0.7797	2.8E-06	0.78	0.59	0.90	0.87	0.91
1986	FORM	25.83	13	.018	.363	0.919	0.13	0.96	0.89	0.98	0.98	0.99
1986	LAMBDA	62.18	23	1.8E-05	.008	0.894	0.0027	0.90	0.85	0.96	0.97	0.98
1986	LAMBDA-PHI	68.32	30	8.1E-05	.017	0.908	0.011	0.90	0.88	0.96	0.98	0.98
1986	LAM-PHI-DELTA	81.13	36	2.5E-05	.017	0.909	0.0073	0.88	0.88	0.95	0.98	0.97
1987	NOMETHOD	151.68	19	0	.001	0.7857	5.3E-05	0.79	0.61	0.91	0.88	0.92
1987	FORM	28.64	13	.0074	.217	0.911	0.071	0.95	0.87	0.98	0.98	0.99
1987	LAMBDA	47.10	23	.0022	.082	0.917	0.056	0.93	0.89	0.97	0.98	0.98
1987	LAMBDA-PHI	49.89	30	.013	.236	0.934	0.196	0.93	0.91	0.97	0.99	0.99
1987	LAM-PHI-DELTA	57.46	36	.013	.237	0.937	0.23	0.92	0.92	0.96	0.99	0.99
1988	NOMETHOD	93.16	8	0	.001	0.74	0.00000015	0.83	0.55	0.91	0.85	0.85
1988	FORM	7.10	5	.21	.965	0.949	0.42	0.99	0.94	0.99	0.99	0.99
1988	LAMBDA	25.44	12	.013	.083	0.917	0.1	0.95	0.92	0.98	0.98	0.98
1988	LAMBDA-PHI	30.63	17	.022	.125	0.93	0.19	0.95	0.93	0.97	0.99	0.99
1988	LAM-PHI-DELTA	51.35	21	.00024	.029	0.906	0.015	0.91	0.91	0.95	0.98	0.98

Note: χ^2 = the chi-square test (see Bollen, 1989, p. 263); *df* = degrees of freedom for the chi-square test; *p* value = *p* value for the chi-square test; bootstrap *p* = bootstrapped *p* value (Bollen & Stine, 1993); RMSEA = root mean square error of approximation (Steiger & Lind, 1980); PCLOSEFIT = *p* value for the test of close fit (RMSEA < .05); GFI = Goodness of Fit Index (Jöreskog & Sörbom, 1986); AGFI = Adjusted Goodness of Fit Index (Jöreskog & Sörbom, 1986); NFI = Normed Fit Index (Bentler & Bonett, 1980); NNFI = Non-Normed Fit Index (Tucker & Lewis, 1973); IFI = Incremental Fit Index (Bollen, 1989).

REFERENCES

Arat, Zehra F. (1988). Democracy and economic development: Modernization theory revisited. *Comparative Politics, 21*, 21-36.

Arat, Zehra F. (1991). *Democracy and human rights in developing countries*. Boulder, CO: Lynne Rienner.

Banks, Arthur S. (1971). *Cross-polity time-series data*. Cambridge: MIT Press.

Banks, Arthur S. (1979). *Cross-national time-series data archive user's manual*. Binghamton: State University of New York at Binghamton.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.

Bollen, Kenneth A. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review, 45*, 370-390.

Bollen, Kenneth A. (1986). Political rights and political liberties in nations: An evaluation of human rights measures, 1950 to 1984. *Human Rights Quarterly, 8*, 567-591.

Bollen, Kenneth A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Bollen, Kenneth A. (1990). Political democracy: Conceptual and measurement traps. *Studies in Comparative International Development, 25*, 6-24.

Bollen, Kenneth A. (1993). Liberal democracy: Validity and method factors in cross-national measures. *American Journal of Political Science, 37*, 1207-1230.

Bollen, Kenneth A. (1995). Democracy, measures of. In S. M. Lipset (Ed.), *Encyclopedia of democracy*, vol. III (pp. 817-821). Washington, DC: Congressional Quarterly.

Bollen, Kenneth A., Entwisle, Barbara, & Anderson, Arthur S. (1993). Macrocomparative research methods. *Annual Review of Sociology, 19*, 321.

Bollen, Kenneth A., & Jackman, Robert. (1990). Regression diagnostics: An expository treatment of outliers and influential cases. In John Fox & J. Scott Long (Eds.), *Modern methods of data analysis* (pp. 257-291). Newbury Park, CA: Sage.

Bollen, Kenneth A., & Paxton, Pamela. (1998). Detection and determinants of bias in subjective measures. *American Sociological Review, 63*, 465-478.

Bollen, Kenneth A. & Stine, Robert A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In Kenneth A. Bollen & J. Scott Long (Eds.), *Testing structural equation models* (pp. 111-135). Newbury Park, CA: Sage.

Bueno de Mesquita, Siverson, Randolph M., & Woller, Gary. (1992). War and the fate of regimes: A comparative analysis. *American Political Science Review, 86*, 638-646.

Burger, J. M. (1981). Motivational biases in the attribution of responsibility for an accident: A meta analysis of the defensive-attribution hypothesis. *Psychological Bulletin, 90*, 496-512.

Burkhart, Ross E., & Lewis-Beck, Michael S. (1994). Comparative democracy: The economic development hypothesis. *American Political Science Review, 88*, 903-910.

Cutright, Phillips. (1963). National political development: Measurement and analysis. *American Sociological Review, 28*, 253-264.

Cutright, Phillips, & Wiley, James A. (1969). Modernization and political representation: 1927-1966. *Studies in Comparative International Development, 5*, 23-41.

Dahl, Robert A. (1971). *Polyarchy: Participation and opposition*. New Haven, CT: Yale University Press.

Derbyshire and Derbyshire. (1991). *Spotlight on world political systems*. New York: Chambers.

Elections since 1945: A worldwide reference compendium. (1989). Chicago: St. James.

Facts on file. (1980). New York: Facts on File.

Facts on file. (1984). New York: Facts on File.

Facts on file. (1988). New York: Facts on File.
Gasiorowski, Mark J. (1995). Economic crisis and political regime change: An event history analysis. *American Political Science Review, 89*, 882-897.
Gasiorowski, Mark J. (1996). An overview of the political regime change dataset. *Comparative Political Studies, 29*, 469-483.
Gastil, Raymond D. (1988). *Freedom in the world: Political rights and civil liberties 1987-1988*. New York: Freedom House.
Gonick, Lev S., & Rosh, Robert M. (1988). The structural constraints of the world-economy on national political development. *Comparative Political Studies, 21*, 171-199.
Gurr, Ted. (1990). *Polity II: Political structure and regime change 1800-1986*. Ann Arbor, MI: Interuniversity Consortium for Political and Social Research.
Gurr, Ted, Jaggers, Keith, & Moore, Will H. (1990). The transformation of the Western state: The growth of democracy, autocracy, and state power since 1800. In Alex Inkeles (Ed.), *On measuring democracy* (pp. 69-104). New Brunswick, NJ: Transaction Publishing.
Hartman, John, & Hsiao, Wey. (1988). Inequality and violence: Issues of theory and measurement in Muller. *American Sociological Review, 53*, 794-800.
Huntington, Samuel. (1991). *The third wave: Democratization in the late twentieth century*. Norman: University of Oklahoma Press.
Jackman, Robert W. (1993). *Power without force: The political capacity of nation-states*. Ann Arbor: University of Michigan Press.
Jaggers, Keith, & Gurr, Ted Robert. (1995). Tracking democracy's third wave with the polity III data. *Journal of Peace Research, 4*, 469-482.
Jöreskog, Karl G., & Sörbom, Dag. (1993). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.
Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review, 98*, 155-163.
Knack, Stephen, & Keefer, Philip. (1997). Does inequality harm growth only in democracies? A replication and extension. *American Journal of Political Science, 41*, 323-332.
Lenski, Gerhard, & Lenski, Jean. (1982). *Human societies*. New York: McGraw-Hill.
Lerner, Daniel. (1958). *The passing of traditional society: Modernizing the Middle East*. Glencoe, IL: Free Press.
Lipset, Seymour Martin. (1994). The social requisites of democracy revisited. *American Sociological Review, 59*, 1-22.
Muller, Georg P. (1988). *Comparative world data*. Baltimore, MD: Johns Hopkins University Press.
Muller, Edward N., & Seligson, Mitchell A. (1994). Civic culture and democracy: The question of causal relationships. *American Political Science Review, 88*, 635-652.
Nagle, John D. (1985). *Introduction to comparative politics: Political systems and performance in three worlds*. Chicago: Nelson-Hall.
New York Times Index. (1980). New York: New York Times.
New York Times Index. (1984). New York: New York Times.
New York Times Index. (1988). New York: New York Times.
Poe, Steven C., & Tate, C. Neal. (1994). Repression of human rights to personal integrity in the 1980's: A global analysis. *American Political Science Review, 88*, 853-872.
Przeworski, Adam, Michael Alvarez, Jose Antonio Cheibub, and Fernando Limongi. (1996). What makes democracies endure? *Journal of Democracy, 7*, 39-55.
Rossi, Peter H., & Anderson, Andy B. (1982). The factorial survey approach: An introduction. In P. H. Rossi & S. L. Nock (Eds.), *Measuring social judgements: The factorial survey approach* (pp. 15-67). Beverly Hills, CA: Sage.

Satorra, Albert. (1990). Robustness issues in structural equation modeling: A review of recent developments. *Quality and Quantity, 24*, 367-386.
Schumpeter, Joseph A. (1950). *Capitalism, socialism, and democracy.* New York: Harper & Row.
Scoble, Harry M., & Wiseberg, Laurie S. (1981). Problems of comparative research on human rights. In Ved P. Nanda, James R. Scarritt, and George W. Shepherd, Jr. (Eds.), *Global human rights: Public policies, comparative measures, and NGO strategies* (pp. 147-171). Boulder, CO: Westview.
Shaver, K. G. (1970). Defensive attribution: Effects of severity and relevance on the responsibility assigned for an accident. *Journal of Personality and Social Psychology, 14*, 101-113.
Singleton, Royce Jr., Straits, Bruce C., Straits, Margaret M., & McAllister, Ronald J. (1988). *Approaches to social research*. Oxford, UK: Oxford University Press.
Small, Melvin, & Singer, J. David. (1982). *Resort to arms: International and civil wars, 1816-1980* (2nd ed.). Beverly Hills, CA: Sage.
Steiger, J. H., & Lind, J. M. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
Sussman, Leonard R. (1982). The continuing struggle for freedom of information. In Raymond D. Gastil (Ed.), *Freedom in the world, 1982* (pp. 101-119). Westport, CT: Greenwood.
Taylor, Charles L., & Jodice, David. (1983). *World handbook of political and social indicators III*. New Haven, CT: Yale University Press.
Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10.
United Nations. 1991. *United Nations Statistical Yearbook*. New York: Author.
Vanderbilt Television News Index and Abstracts. (1980). Nashville, TN: Vanderbilt Television News Archive.
Vanderbilt Television News Index and Abstracts. (1984). Nashville, TN: Vanderbilt Television News Archive.
Vanderbilt Television News Index and Abstracts. (1988). Nashville, TN: Vanderbilt Television News Archive.
Vanhanen, Tatu. (1990). *The process of democratization*. New York: Crane Russak.
The world almanac and encyclopedia. (1990). New York: Press.

Kenneth A. Bollen is the Zachary Smith Professor of Sociology at the University of North Carolina (UNC) at Chapel Hill and is a fellow of the Carolina Population Center at UNC. Liberal democracy remains a focus of this substantive research. He also has several methodological projects in progress on structural equation modeling and longitudinal analysis. He is the author of Structural Equations With Latent Variables *(Wiley) and* Testing Structural Equation Models *(Sage). Recent articles have been published in the* American Sociological Review, Social Forces, Sociological Methods & Research, *and the* Encyclopedia of Biostatistics.

Pamela Paxton is an assistant professor at Ohio State University. Her research interests include political sociology, stratification, and quantitative methods. Some of her recent publications have appeared in the American Sociological Review, American Journal of Sociology, *and* Social Science Research.