Quality & Quantity 21: 349–360 (1987)

Corrections for coarsely categorized measures: LISREL's polyserial and polychoric correlations

ROBERT M. O'BRIEN & PAMELA HOMER
Department of Sociology, College of Arts and Sciences, University of Oregon, Eugene, OR 97403, USA

Department of Advertising, College of Communications, The University of Texas at Austin, TX 78712, USA

Abstract. Versions 5 and 6 of LISREL (Joreskog and Sorbom, 1981) contain procedures that estimate the underlying correlation between continuous variables on the basis of crude rank category measures. The procedures assume that the distribution of the measured variables would have been bivariate normal if they had not been categorized. Using survey data and simulations, the accuracy of these polyserial/polychoric (P/P) based estimates of the underlying correlations are compared with those based on simple equal distance scoring of the categories. The results indicate that under some conditions, e.g., nearly normally distributed variables and moderate to high correlations, the polyserial/polychoric based estimates are better. Under other conditions, e.g., a moderate to high degree of skew and kurtosis and low correlations, the equal distance score based estimates are better. Under all conditions, the amount of error decreases fairly rapidly as the number of categories is increased from two to five.

Introduction

A number of recent studies in the social science literature have discussed problems involving the use of rank category measures of underlying continuous variables (e.g., Bollen and Barb, 1981; Gephart, 1983; Henry, 1982; Johnson and Creech, 1983; Martin, 1975, 1978; O'Brien 1981a, 1981b, 1983).[1] These authors note that many commonly used variables may be conceived of as having a continuous interval level underlying distributions, while being measured by a crude ranking of categories, e.g., a Likert-type item measuring job satisfaction or the rating of a topic as salient or nonsalient.

These studies have confirmed the common sense expectation that the cruder the categorizations involved (e.g., two categories rather than seven), the greater the discrepancy between observed correlations and underlying correlations. Further, they provide an indication of the potential amount of error involved and suggest that one way to ameliorate the distorting effects

[1] I have not cited related studies that deal with rank-order measures of underlying variables (see, O'Brien 1983).

of employing rank category measures of underlying continuous variables is to construct measures in a way that increases the number of rank order categories into which a respondents' answers can be placed. For example, constructing measures with seven response categories rather than two, (e.g., Lehman and Hulbert, 1972; Neumann, 1983) forming a summative scale of items (e.g., Cox, 1970; Givon and Shapira, 1984; Ramsay, 1973), or the use of a multiple indicator model in LISREL (Johnson and Creech, 1983). These solutions suggest the use of more refined measures through the inclusion of more items and/or response categories.

There are, however, proposed solutions to the problem of using crude rank category measures (of underlying continuous variables) that do not require multiple measures or a finer grade of measurement. These include the standard treatment of dichotomous dependent variables using probit or logit analysis (Hanusek and Jackson, 1977).[2] These models may be extended to the case of polythomous variables (e.g., Aitchison and Silvey, 1957; Amemiya, 1975; McCullagh, 1980; McKelvey and Zavoina, 1975). Much of this literature has ben cogently summarized for sociologists in two articles by Winship and Mare (1983, 1984). They demonstrate how these techniques can be used to estimate relationships involving polytomous ordinal variables in structural equation and path models. Each of these solutions to the problem of using crude rank category variables involves an assumption about the distribution of the unobserved (uncategorized) variables. For example, probit models assume an underlying normal distribution, logit models assume an underlying logistic distribution, and tetrachoric correlations an underlying bivariate normal distribution.

Version 5 of LISREL (Joreskog and Sorbom, 1981) provides an even more comprehensive set of procedures than those suggested by Winship and Mare (1983, 1984). LISREL's polyserial/polychoric (P/P) procedures allow for the factor analysis of dichotomous or polythomous items, multivariate probit regression, the combining of structural and measurement models that use rank order measures, and other sophisticated analyses.

The LISREL procedure accomplishes this by estimating the bivariate correlations between continuous variables based on their polytomous measures using maximum likelihood estimating procedures developed by Olsson and associates (Olsson, 1979; Olsson, Drasgow and Dorans, 1982). LISREL provides an estimate of the correlation between a continuous and a polytomous measured variable, which is labeled a polyserial correlation (this

[2] The logit analysis assumes an underlying logistic distribution while the probit analysis assumes an underlying normal distribution. Hanusek and Jackson (1977: 204) note that the maximum likelihood logistic estimator is very similar to the probit estimator and that the choice between the two is largely one of convenience and program availability.

includes the biserial correlation as a special case, e.g., when the polytomous variable is a dichotomy). This is an estimate of the correlation that exists between the underlying random variables Y and Z, where Y has been measured on a continuum and Z has been measured as a polytomy. The procedure is based on the assumption that Y and Z have a bivariate normal distribution. LISREL also provides estimates of the underlying correlation when both variables are measured as polytomies. These estimates are labeled polychoric correlations (they include the tetrachoric correlation as a special case, e.g., when both variables are measured as dichtomomies). Again it is assumed that the two underlying variables share a bivariate normal distribution. The LISREL procedure calculates polyserial and polychoric correlations for rank category variables having from two to nine categories. these estimated bivariate correlations can then be used as "improved" inputs into any of the large variety of models that can be estimated by LISREL.

The *potential* importance of these LISREL procedures is difficult to overemphasize, since they make available to social scientists a general solution to the problem of using crude rank category measures. By assuming a bivariate normal underlying distribution, crude rank category measures can be used to estimate the relationship among underlying continuous variables. Enthusiasm for this method must be tempered, however, by asking: "What happens if the underlying distribution between the two latent variables is not bivariate normal?" Certainly, in this situation, the P/P estimates of the underlying correlation will not be correct: see Bollen and Barb (1981), Henry (1982) and O'Brien (1981a, 1981b, 1985) for the effects of categorization on nonnormally distributed underlying variables. One might suppose, however, that the P/P estimates would be better than uncorrected correlations based on the common approach of assigning integer values to each rank order category, i.e., equal distance scores (EDS). A second question is thus generated: "How great is the improvement that results from using P/P estimates rather than EDS? In this paper, these two questions are examined by comparing the accuracy of estimates of underlying correlations between two continuous underlying variables that are based on the EDS with those estimates produced by LISREL's P/P procedures.

Methods

The obvious strategy to use in this research is a simulation of some sort, but a major decision involves choosing parameters for the simulation. We could proceed by taking data from an underlying bivariate normal distribution with a known correlation and then categorize these data and input the resulting data into the LISREL procedure for polyserial and polychoric correlations; the program would then produce highly accurate estimates of the underlying correlations (Olsson, 1979; Olsson, et al., 1982). But what if the underlying bivariate distribution is not normal? And, if it is not a bivariate normal distribution, how are we to fairly represent the typical type(s) of underlying distribution(s)? The problem is, of course, ultimately not solvable, because we almost never know the unobserved underlying distributions of variables.[3] However, if the simulations do not represent the sort of situations typical of the real-world, the findings will be of little use to researchers.

The strategy used involved the selection of a number of variables from the 1980 American National Election Studies[4] that were measured somewhat continuously, e.g., respondents' NORC occupational prestige and their attitudes as measured by thermometer scales (thermometer scales allow responses from zero to one hundred). These variables were used to represent the distribution of underlying variables. We do not contend that respondents' scores on these variables are the same as their true underlying scores, but that the distribution of such variables may not be unreasonable to use as representative of those for underlying variables. At a minimum, we have not tried to select extreme distributions. Later in this paper more extreme distributions are examined.

Specifically, nine variables were selected from the American National Election Studies conducted in 1980. Each of the variables selected might be assumed to provide some degree of interval level measurement. Seven thermometer scales were used that asked respondents to record their feelings toward poor people, people on welfare, big business, businessmen and businesswomen, the federal government, federal government workers and Black militants. The respondents' NORC coded occupational prestige and their personal income were also included.[5] Each of these nine variables was

[3] In some cases we have an idea of the underlying continuous distribution, e.g., income is positively skewed while test scores on an easy test are negatively skewed.

[4] Neither the Inter-University Consortium for Political and Social Research (which generously made the data available) nor the original collectors of the data are responsible for the analyses and conclusions in this paper.

[5] These particular variables were somewhat arbitrarily selected from the American National Election Studies data. We selected occupational prestige because it is the most finely graded of all the measures, income because of its known nonnormal distribution and several pairs of thermometer scale variables, e.g., those for poor people, the federal government and business. Income was coded as a rank category variable with 22 categories. These were recoded using the midpoints of each category and the mean income for the open-ended category was estimated using a procedure outlined in Shryock and Siegel (1971: 365–366). There were 882 cases for which there was information on all nine variables and these cases served as our data base.

then categorized in various ways. Five dichotomous versions were created, for example, a version with approximately 10% the cases in the low scoring category and 90% in the high scoring category (a 10–90 version), as well as 30–70, 50–50, 70–30 and a 90–10 version. Two trichotomous versions, a 33–33–33 version and a 25–25–50 version were constructed.[6] Thus, there were seven versions of each variable in addition to its continuous version. Each of these eight versions of the seven thermometer scales was correlated with each of the eight versions of the income and NORC variables, and each version of income and NORC was correlated with the other. This resulted in fifteen eight by eight matrices of unique correlations. These correlations were of two types: (1) Pearson product moment correlations using EDS for each category, and (2) LISREL P/P correlations.

Each eight by eight matrix of correlations contains one "regular" correlation (based on uncategorized data), 14 polyserial correlations (10 based on dichotomies, and 4 based on trichotomies), 25 tetrachoric correlations (based on the correlations between two dichotomous variables), 20 polychoric correlations based on a dichotomy and a trichotomy and four polychoric correlations based on two trichotomies. Multiplying each of these by 15 gives the number correlations of each type upon which the results are based. In all there were 945 correlations estimated by each method.[7] Five criteria were used to compare the EDS and P/P based estimates of the "true" underlying correlation (i.e., the correlation between the uncategorized variables): (1) the average signed bias, (2) the mean squared error (efficiency), (3) the average absolute deviation, (4) the percentage of times the EDS correlations were closer to the actual value than the P/P estimates and (5) the percentage of times the EDS estimates and the P/P estimates were "too high." In the next section the results of these anslyses are presented.

Results

Table 1 presents some results of these comparisons. In terms of bias, the pattern is fairly consistent. Not unexpectedly, the correlation based on EDS underestimate the underlying correlations on the average (the average signed bias is -0.042), while the P/P estimates tend to overestime the underlying correlations (average signed bias is 0.015). This pattern holds for the

[6] Selected comparisons (described later) were made with a five category version (20–20–20–20–20).

[7] We should note that the LISREL calculations of polyserial and polychloric correlations were not inexpensive in terms of either money spent or computer time, and that the maximum likelihood estimates failed to converge on a solution when categorized versions were correlated with other versions of the same variable (i.e., in the case of an underlying correlation of 1.00).

Table 1. Comparison of errors using equal distance and polyserial/polychoric (P/P) estimates

	Bias		Mean squared error		Absolute error	
	Equal distance	P/P	Equal distance	P/P	Equal distance	P/P
Dichotomy–Dichotomy	−0.058	0.014	0.007	0.008	0.067	0.071
Dichotomy–Trichotomy	−0.038	0.016	0.004	0.005	0.052	0.058
Trichotomy–Trichotomy	−0.017	0.019	0.002	0.003	0.036	0.044
Dichotomy–Continuous	−0.033	0.137	0.004	0.005	0.044	0.056
Trichotomy–Continuous	−0.055	0.171	0.001	0.002	0.022	0.032
Total	−0.042	0.015	0.005	0.006	0.054	0.060

estimates based on a polytomous and continuous variable, two dichotomies, and a dichotomy and trichotomy. Only in the case of two trichotomies (where there are only 60 nonindependent cases) is the bias fairly even (albeit in opposite directions). There appears to be a tendency for the P/P estimates to be overestimates and there is a clear tendency for EDS estimates to be underestimates. If the degree of bias were the sole criterion, the P/P estimates might be declared a success, since their bias is almost two–thirds less than that of estimates based on EDS. There are other criteria, however, and by these criteria the P/P procedure does not provide better estimates.

The mean squared error (the mean of the squared deviations of the estimated correlations from the underlying correlations) is a measure of efficiency and, as Blalock states (1979: 207), when discussing bias and efficiency, "Of the two criteria discussed efficiency is the more important. . . . Knowledge that estimates will average out to the correct figure in the long run is little consolation if, for any given sample, the estimate is quite likely to be far from the parameter." When the mean squared errors in Table 1 are compared, the correlations based on EDS perform slightly, though probably not significantly, better than the estimates generated by the P/P procedure; i.e., the average squared distance of these estimates from the underlying correlations is as small as those provided by the P/P procedure. The interpretation of this result is perhaps clearer if we examine the results based on the average of the absolute values (not squared values) of these errors. For example, for 945 estimates the EDS correlations deviate (in terms of absolute value) from the "true value" on the average by 0.054 while P/P based estimates are off by an average of 0.060. Again the results are consistent across all conditions, i.e., dichotomy–dichotomy, dichotomy–trichotomy, etc.

The EDS correlations show more bias because they are almost always too low (797 out of 945 times) while the P/P estimates, though typically too high (547 out of 945 times), exhibit a better balance of high and low values. At the same time, the EDS estimates are closer to the actual value in almost one-half of the cases (469 out of 945 cases). Thus, for these data, there is little or no advantage to using the P/P based estimates, rather than the EDS based estimates.

Some important (though not surprising) additional results can be gleaned from Table 1. For example, for both the EDS and P/P estimates involving two polytomies the mean squared error (and the average absolute error) is greatest in the case of two dichotomies, next highest for the dichotomy–trichotomy and least for the case of two trichotomies. In the situation involving a polytomy and a continuous variable, dichotomies result in greater error than trichotomies. For two polytomies the EDS based estimates are less biased as the number of categories increases while the number of categories does not effect the bias of the P/P estimates in the same way. If anything, there is a tendency toward greater bias as the number of categories increases, but this trend is probably not significant.

Additional results, which do not appear in Table 1, indicate that both P/P and EDS based correlations do poorly when the category marginals are extreme. For example, when one variable has a 10–90 split and the other has a 90–10 split, the absolute error using P/P based estimated averaged 0.073 and using EDS based estimates averaged 0.107, when the splits were 90–10 and 90–10 (or 10–90 and 10–90) the average absolute error for P/P correlations was 0.1032 and for the EDS based correlations was 0.082. These marginals result in the most dramatic increases in error, while the 50–50 marginals on both variables showed the least error (P/P 0.055 and the EDS based correlations 0.045).

Our final analyses of the survey data involved examining a 20–20–20–20–20 split of each of the variables. In this case the estimates showed great improvement on all criteria for both the EDS and P/P estimates. Bias was −0.006 for the EDS based estimates and 0.003 for the P/P based estimates; the mean squared error was 0.001 for both of them and the average absolute error was 0.022 for the EDS estimates and 0.021 for the P/P estimates. This result strongly supports the argument for increasing the number of response categories in measurement.

Simulated data

One problem with making generalizations from the survey data used in this study is that both their skewness and kurtosis were not very extreme. The skewness of all variables (except respondent's income) ranged from 0.061 to 0.321 in absolute value and averaged 0.230 (in absolute value), kurtosis (except for respondent's income) ranged from −0.409 to 0.274.[8] It is crucial to examine what happens with greater degrees of skewness and kurtosis. Second, the correlations among most of the survey variables were low: the average correlation was 0.124. It is essential to examine the performance of the P/P and EDS estimates across a larger range of correlations.

The first step in our simulations involved generating variables having a multivariate nonnormal distribution with specified values of skew, kurtosis and correlation. This was accomplished by using a procedure outlined by Vale and Maurelli (1983). For one set of simulations, a trivariate distribution with a skew of 0.25 and kurtosis of 0.50 for each variable and intercorrelations among them of 0.12, 0.38 and 0.80 was generated.[9] For a second set of simulations, the skewness was set at 1.50, the kurtosis at 3.00 and the correlations were again set at 0.12, 0.38 and 0.80. Vale and Maurelli's (1983) procedure was used to randomly generate 1,000 cases for each set of data. Since these values are generated with a random component, the exact values of skewness, kurtosis and correlation vary somewhat from those desired.[10] The scores were then dichotomized in five different ways (90–10, 70–30, 50–50, 30–70 and 10–90), trichotomized in one way (33–33–33) and split into five equal categories (20–20–20–20–20).

Summary results from our analyses appear in Table 2. When the underlying correlation is 0.12 and there is a low degree of skewness and kurtosis, the findings parallel those for data from the survey, e.g., more bias for the EDS based estimates (−0.0383) than for the P/P based estimates (0.0067), but fairly similar mean squared and absolute errors. Again the bias for P/P based correlations is positive and for the EDS is negative. As the correlation increases (for these nearly normally distributed scores), the EDS based correlations perform worse in comparison with the P/P based correlations. When the continuous correlation is 0.38 the absolute error for the EDS based correlations is 0.1349 and for the P/P based correlations is only 0.0308; when the underlying correlation is 0.80, this difference is even greater: 0.2935 versus 0.0368. Further, the bias of the P/P estimate is not dramatically affected as the underlying correlation increases.

[8] Respondent's income exhibited a high degree of both skewness (3.166) and kurtosis (13.377). All measures of skewness and kurtosis were based on version 9 of SPSS, which uses Fisher's statistics for skewness and kurtosis (Bliss, 1967).

[9] The correlation of 0.12 is close to the average correlation found in our survey data, that of 0.38 is close to the highest correlation we found (that between respondent's income and occupational prestige) and that of 0.80 was chosen arbitrarily to represent a very high correlation.

[10] Over the three levels of correlations, the average skewness and kurtosis for the low skewness and kurtosis simulations was 0.212 and 0.520 respectively. The correlations were 0.12, 0.38 and 0.80. For the high skewness and kurtosis simulations, the average skewness was 1.49 and the average kurtosis was 3.15. The correlations were 0.12, 0.38 and 0.81.

moderate ($r = 0.38$) correlations with a degree of skew and kurtosis similar to those found in our survey data, the P/P procedure provided better estimates than the EDS. With a higher degree of skew and kurtosis the P/P based correlations were poorer than the EDS based correlations for low correlations ($r = 0.12$), somewhat better for moderate correlations ($r = 0.38$), and substantially better for a high underlying correlation ($r = 0.80$). Thus, we tentatively conclude that with highly correlated variables the P/P procedure is useful (at least for the levels of skewness and kurtosis that we investigated) and that it may also be for moderately correlated variables. The situation with low values of correlations is more ambiguous, since our results show that the P/P based estimates may actually be poorer than the EDS based estimates in this situation. This last set of conclusions was suggested as a possibility by Kim, Nie and Verba (1977) for the case of dichotomously measured variables for use in factor analysis, although they did simulations only with normally distributed variables. They stated (1977: 57): "However, one situation does exist under which ϕ [phi: which is the same as the EDS based correlation for a dichotomy] may be preferred to tetrachoric: when the underlying factor is assumed continuous but not normal, *and* the underlying correlations among the continuous (hypothetical) variables are assumed low."

An important, and we believe preferable, solution to the problem of crude measurement should not be overlooked. That is the use of better measurement to begin with (see, Duncan, 1975: 161 and Blalock, 1974: 424 for similar comments), rather than the use of ad hoc assumptions about the forms of the unknown underlying distributions. In our simulations, with only five categories for each variable (with a 20–20–20–20–20 set of marginals), the EDS based correlations were never off by more than 0.041 (absolute error) for any level of underlying correlation or skewness and kurtosis and the error in the P/P based estimates generally were quite low.

References

J. Aitchison and S.D. Silvey, The generalization of probit analysis to the case of multiple responses. *Biometrika* 57 (1957) 253–262.

T. Amemiya, Qualitiative response models. *Annals of Economic and Social Measurement* 4 (1975) 363–372.

C.I. Bliss, *Statistics in Biology* (Volume 1), New York: McGraw-Hill (1967).

H.M. Blalock, Jr., *Social Statistics* (revised second edition). New York: McGraw-Hill (1979).

—— Beyond ordinal measurement: weak tests of stronger theories. pp. 424–454 In: H.M. Blalock, Jr. (ed.), *Measurement in the Social Sciences: Theories and Strategies*, pp. 424–454. Chicago: Aldine (1974).

K. Bollen and K. Barb, Pearson's R and coarsely categorized measures. *American Sociological Review* 46 (1981) 232–239.

D.R. Cox, *The Analysis of Binary Data*. London: Methuen (1970).

O.D. Duncan, *Introduction to Structural Equation Models*. New York: Academic Press (1975).

R.P. Gephart, Multiple R, the "parametric strategy," and measurement imprecision. *Sociological Perspectives* 26 (1983) 473–500.

M.M. Givon, and Z. Shapira, Response rating scales: a theoretical model and its application to the number of categories problems. *Journal of Marketing Research* 21 (1984) 410–419.

E.A. Hanusek and J.E. Jackson, *Statistical Methods for Social Scientists*. New York: Academic Press (1977).

F. Henry, Multivariate analysis of ordinal data. *American Sociological Review* 47 (1982) 299–304.

P. Homer and R.M. O'Brien, Using LISREL models with crude rank category measures. *Quality and Quantity*, (forthcoming).

D.R. Johnson and J.C. Creech, Ordinal measures in multiple indicator models: a simulation study of categorization error. *American Sociological Review* 48 (1983) 398–407.

K.G. Joreskog and D. Sorbom, *LISREL V: Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Methods*. Chicago: International Educational Services (1981).

Jae-On Kim, Norman Nie and Sidney Verba, A note on factor analyzing dichotomous variables: the case of political participation. Political Methodology 4 (1977) 39–62.

D.R. Lehmann and J. Hulbert, Are three-point scales always good enough? *Journal of Marketing Research* 9 (1972) 444–445.

W.S. Martin, The effects of scaling on the correlation coefficient: a test of validity. *Journal of Marketing Research* 10 (1975) 316–318.

—— Effects of scaling on the correlation coefficient: additional considerations. *Journal of Marketing Research* 15 (1978) 304–308.

P. McCullagh, Regression models for ordinal data. *Journal of the Royal Statistical Society*, Series B 42 (1980) 109–142.

R.D. McKelvey and W. Zavoina, A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 4 (1975) 103–120.

L. Neumann, The effects of scale length on means and correlation coefficients. *Quality and Quantity* 17 (1983) 405–408.

R.M. O'Brien, Reducing grouping distortion in rank category variables. *American Journal of Political Science* 25 (1981) 605–616.

—— Using rank category variables to represent continuous variables: defects of common practice. *Social Forces* 61 (1981) 144–155.

—— Rank order versus rank category measures of continuous variables (comment on Bollen and Barb, *ASR*, April 1981, and Henry, *ASR*, April 1982). *American Sociological Review* 48 (1983) 284–286.

—— The relationship between ordinal measures and their underlying values: why all the disagreement? *Quantity and Quality* 19 (1985) 265–277.

U. Olsson, Maximum likelihood estimation of the polychoric correlation coefficient, *Psychometrika* 44 (1979) 443–460.

U. Olsson, F. Drasgow and N.J. Dorans, The polyserial correlation coefficient. *Psychometrika* 47 (1982) 337–347.

J.O. Ramsay, The effect of number of categories in rating scales on precision of estimation of scale values. *Psychometrika* 38 (1973) 513–532.

H.S. Shryock, J.S. Siegel and Associates, *The Methods and Materials of Demography* (3rd edn.). Washington, D.C.: U.S. Government Printing Office (1971).

C.D. Vale and V.A. Maurelli, Simulating multivariate nonnormal distributions. *Psychometrika* 48 (1983) 465–474.

C. Winship and R.D. Mare, Structural equation and path analysis for discrete data. *American Journal of Sociology* 89 (1983) 54–110.

—— Regression models with ordinal variables. *American Sociological Review* (1984).

Table 2. Comparison of errors using equal distance and polyserial/polychoric (P/P) estimates in distributions with varying skew, kurtosis and correlation

		Low skew and kurtosis		High skew and kurtosis	
Underlying correlation		Equal distance	P/P	Equal distance	P/P
0.12	Bias	−0.0383	0.0067	−0.0162	0.0486
	MSE[a]	0.0028	0.0025	0.0015	0.0049
	ABE[b]	0.0412	0.0386	0.0307	0.0540
0.38	Bias	−0.1349	0.0154	−0.0995	0.0831
	MSE	0.0231	0.0015	0.0155	0.0118
	ABE	0.1349	0.0308	0.1030	0.0870
0.80	Bias	−0.2935	0.0053	−0.3007	0.0175
	MSE	0.1138	0.0025	0.1195	0.0037
	ABE	0.2935	0.0368	0.3007	0.0477

[a] Mean squared error.
[b] Average absolute error.

The results are quite different when the less normally distributed data are examined, i.e., the distribution with higher degrees of skew and kurtosis. When the underlying correlation is 0.12, the EDS based correlations are more accurate than the P/P based correlations in terms of bias (−0.0162 versus 0.0486) and mean squared and absolute error. When the underlying correlation is 0.38 the P/P based estimates are slightly better in terms of bias, mean squared and absolute error, and when the underlying correlation is 0.80 they perform much better in terms of bias, mean squared and absolute error.

This same trend is apparent in the number of times the EDS based correlations are closer to the underlying correlations under the two levels of skewness and kurtosis. When the underlying correlation is 0.12, the EDS based correlations are closer 26 out of 63 times for the underlying distribution with low skew and kurtosis and 44 out of 63 times for the one with high skew and kurtosis. When the underlying correlation is 0.38 the comparable figures are 3 of 63 and 32 of 63 times, and when the underlying correlation is 0.80, the figures are 0 of 63 and 1 of 63 times. Clearly, the degree of skewness and kurtosis, as well as the underlying correlation, affect the relative accuracy of EDS and P/P based estimates of correlations.

A further finding, though not surprising, is encouraging. When as few as five categories were employed to measure each variable, the error using EDS scoring was never greater than 0.041 at any level of underlying correlation for either high or low degrees of skewness and kurtosis. The error for the P/P estimates were also low, except in the high skew and kurtosis condition when the correlations were low and moderate. Here, the absolute errors were 0.0871 and 0.0620 respectively.

Further simulations using both positive and negative values of kurtosis paired with different values of skewness could be carried out and their effects on the EDS and P/P estimates mapped. Such an exercise, however, is less vital when one considers that investigators seldom know the precise distribution of the underlying variables that are crudely measured. When, however, strong theory predicts a particular form of underlying distribution, its parameters can be simulated and the effects of crude rank category measurement assessed using methods similar to those used in our simulations.

Conclusions

The results in this paper address one of two issues concerning the use of LISREL models with crude rank order data, i.e., the use of LISREL's P/P procedures to correct the correlations between variables that will then be used as input to estimate the parameters of LISREL models. A second issue is the use of *multiple measures* to correct for the effects of crude rank order measures on the coefficients of relationships between latent variables. There is evidence that LISREL's performance is impressive in this situation whether EDS (Johnson and Creech, 1983; Homer and O'Brien, forthcoming) or P/P are used (Homer and O'Brien, forthcoming). Thus, the results reported here are most relevant when variables are measured with a *single measure*, as has been typical in discussions of the effects of crude rank category measures (Bollen and Barb, 1981; Henry, 1982; O'Brien, 1981a, 1981b; Winship and Mare, 1983, 1984).

The results based solely on survey data from the American National Elections Studies have fairly straightforward implications, which might have been summarized as follow: To the extent that the bivariate distributions and correlations of the selected variables are representative of those of underlying variables typically examined, the LISREL procedure of calculating P/P correlations does not appear to significantly improve estimates of underlying correlations over estimates based on simple EDS.[11] When the study was extended to include a wider range of correlations between underlying variables, however, we found that for the strong (r = 0.80) and

[11] This is not to say that the use of multiple crude rank category measures of underlying variables in P/P do not lessen the distortion created by the use of only a single such measure. The indication is that the use of such a multiple measure strategy is helpful (Johnson and Creech, 1983). They note, however, (p. 398) that, "Caution is warranted in the use of two-, three- or four-category ordinal indicators, particularly when the sample size is small, as the estimates tend to be biased and inefficient."