Quality & Quantity 22: 191–201 (1988)

Using LISREL models with crude rank category measures

PAMELA HOMER[1] & ROBERT M. O'BRIEN[2]

[1]*Department of Advertising, College of Communication, University of Texas at Austin, Austin, TX 78712, USA*; [2]*Department of Sociology, College of Arts and Sciences, University of Oregon, Eugene, OR 97403, USA*

Abstract. Recent versions of LISREL (Joreskög and Sorbom, 1983) contain procedures for estimating polyserial and polychoric correlations from crude rank category measures. In this paper, the accuracy of these procedures for estimating the relationship between constructs in both a single and a multiple indicator model is compared to that of using Pearsonian correlations based on equal distance scoring of the rank categories. In these comparisons, a variety of multivariate *nonnormal* distributions were simulated and the average bias and average absolute error of the estimated relationships between underlying constructs were calculated. These estimates were affected by the number of rank categories for each measure, the correlations among measures, their skewness and kurtosis, and the correlation between underlying constructs. The most important finding is that although the polychoric procedure can be helpful in estimating the correlation between unobserved variables in single indicator models, it does not improve estimates based on Pearsonian correlations in the multiple indicator model.

Introduction

In recent years the use of LISREL (Joreskög and Sorbom, 1983) has proliferated. LISREL produces estimates of the causal relationships among several underlying constructs measured by multiple indicators. This gives it obvious appeal in both theory building and hypothesis testing.

Versions V and VI of LISREL contain a feature that allows researchers to estimate the correlation between theoretically continuous variables using categorical measurements (e.g., simple yes–no items or five category Likert items). The assumption made is that the underlying distribution is bivariate normal. When these correlations are based on two such rank category measures, the estimated underlying correlation is labeled a polychoric correlation; when they involve a rank category measure and a continuously measured variable, they are labeled polyserial correlations. These labels parallel those for the more familiar tetrachoric and biserial correlations, but polychoric and polyserial correlations can be used with variables having more than two categories.

These procedures should be useful in most of the social sciences, since the level of measurement is often quite crude in these fields even when underlying variables may be conceived of as continuous, e.g., the extensive use of Likert-type items and yes/no answer format to measure attitudes. Since the polychoric procedure assumes an underlying bivariate normal distribution, and many constructs and indicators are not normally distributed, several questions may be raised regarding its applicability.[1]

Our interest focuses on the advantages of using LISREL's polychoric estimates of the relationships between *indicators* for the estimation of the relationship between *underlying constructs*. These estimates are compared to those based on Pearson product moment correlations that use simple equal distance scoring, (i.e., assigning consecutive integer values to the ranked categories). These comparisons involve both multiple and single indicator models, and use categorized data that do not have an underlying bivariate normal distribution.

Previous studies

A number of studies have examined problems involved in the use of rank category measures of underlying continuous variables (e.g., Bollen and Barb, 1981; Gephart, 1983; Lehmann and Hulbert, 1972; Martin, 1975, 1978; O'Brien, 1981a, 1981b, 1983). As expected, these researchers find that the fewer categories used, the greater the discrepancy between observed and underlying correlations, and when there are few categories these discrepancies are often substantial.

One suggested solution to this problem is to use multiple indicator models in LISREL (Johnson and Creech, 1983). However, recent versions of LISREL include a procedure that does not require more than a single indicator per construct in order to correct for categorization error. These estimated bivariate correlations can then be used as direct inputs into a variety of structural equation models with the expectation that the resulting estimates will be more accurate.

In an earlier study, the authors (O'Brien and Homer, 1987) compared the performance of these polyserial and polychoric estimates to uncorrected Pearson product moment correlations; both survey and simulated data with varying levels of correlation, skewness, and kurtosis were used. In the case of survey data from the 1980 American National Elections Studies, the LISREL procedure of calculating polyserial and polychoric correlations did not significantly improve estimates of underlying correlations over estimates based on Pearson product moment correlations using simple equal distance scoring of the rank categories. To capture a broader range of underlying correlations, skewness, and kurtosis, an extensive series of simulations was performed. For strong ($r = 0.80$) and moderate ($r = 0.38$) correlations, with skew and kurtosis similar to that of the survey data, the polyserial and polychoric procedures were useful. With higher levels of skewness and kurtosis these correlations were inferior to the equal distance based Pearson correlations for low underlying correlations ($r = 0.12$), somewhat superior for moderate underlying correlations ($r = 0.38$), and substantially better for high underlying correlations ($r = 0.80$). Under all conditions, the amount of error decreased fairly rapidly as the number of categories increased from two to five.

These findings relate only to the accuracy of Pearson-based and polychoric/polyserial-based estimates of the bivariate correlations between indicators. The present study extends these findings by comparing the accuracy of estimates of the correlation between two *unobserved constructs* produced by a simple multiple indicator model that result when LISREL's polychoric-based correlations are used for input rather than Pearson correlations based on equal distance scoring. These polychoric and equal distance based estimates are compared for both single and multiple indicator models and across a variety of conditions of skew and kurtosis in the variables.

We suspected that only in the case of single indicator models would the polychoric-based correlations among indicators significantly improve estimates of the relationship between underlying constructs. Our reasoning was that multiple indicator models correct for the random error, including that due to categorization. Thus, if the correlations among categorized variables based on equals distance scoring are lower than those using the corrected polychoric estimates, as they must be (O'Brien and Homer, 1987), then these lower correlations among the indicators will be taken into account in the LISREL multiple indicator model, and the relationships between the unobserved constructs will be adjusted for this error. This does not occur in single indicator models, so that correlations based on equal distance scoring will not be corrected for categorization errors, while those based on the polychoric correction will be corrected in the process of obtaining the polychoric correlations.

Method

LISREL's polychoric and polyserial procedures produce unbiased estimates of underlying correlations between variables when data conform to a bivariate normal distribution that has been categorized (Olsson, 1979; Olsson et al., 1982). The accuracy of these estimates, when the underlying bivariate distribution is nonnormal, has not been investigated. To examine the effects

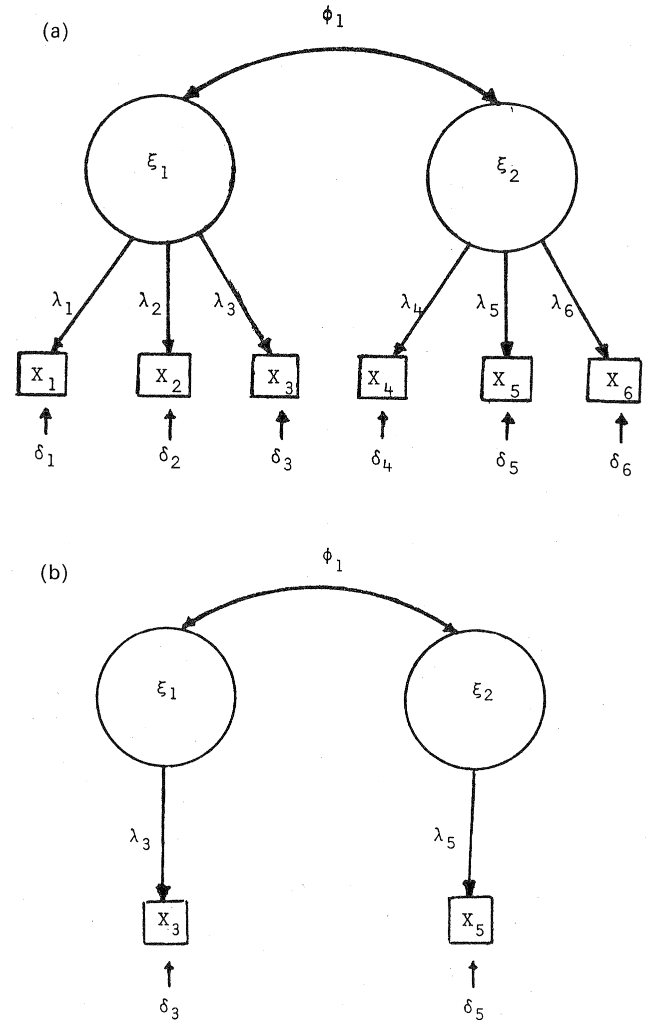

Fig. 1. (a) Multiple indicator model; (b) single indicator model.

of violating the normality assumption, a simple two construct, six indicator model was used (see Fig. 1). Each underlying construct (ksi) is measured by three indicators (X's) with independent error terms (delta's).

The first step involved the generation of sets of variables having multivariate nonnormal distributions with specified values of skewness, kurtosis, and correlations (Vale and Maurelli, 1983). Twenty-seven different distributions were generated. There were three different levels of the unobserved correlation (phi) between the two constructs (ksi's): high (0.80), medium (0.60), and low (0.30). The correlations between the three indicators of each construct and the construct (lambda's) were of three types: all high (0.9, 0.8, 0.9 for the first construct and 0.9, 0.8, 0.9 for the second), all low (0.4, 0.3, 0.4 for the first and 0.4, 0.3, 0.4 for the second construct), or one set of high and one set of low correlations (0.9, 0.8, 0.9 for the first and 0.4, 0.3, 0.4 for the second construct). Finally, there were three different combinations of skew and kurtosis. One-third of the simulations generated six variables (X's) of low positive skew (0.25) and low positive kurtosis (0.50); one third produced six variables of high positive skew (1.50) and high positive kurtosis (3.00), the remaining third simulated distributions containing three variables with high positive skew and high positive kurtosis, and three variables with high negative skew and high positive kurtosis.[2] This last set of simulations was conducted in an effort to determine potential effects of skewness in opposite directions on the estimation of the underlying relationship between constructs. These three levels of phi, three combinations of sets of values for the lambdas, and three combinations of skewness and kurtosis, when completely crossed, yielded 27 different distributions.

The number of cases randomly generated for each of the 27 simulations was 2000. Since these cases were generated with a random component, the exact values of skewness, kurtosis, and correlation vary somewhat from those specified. Comparison of the simulated and targeted correlations showed that the average amount of bias for the 27 simulations was $+0.0019$. In terms of skewness, the simulated variables tended to be slightly more highly skewed than desired with an average bias of 0.0026; the kurtosis of the simulated viarables had an average bias of -0.0125. These comparisons indicate the success of the simulations in generating nonnormal multivariate distributions with expected values of skewness, kurtosis, and correlations.

The scores from each of these 27 distributions were then categorized in two ways: dichotomized at the median (50–50) or split into five equal categories (20–20–20–20–20). Five different correlations derived from each of the 27 distributions were used as input for LISREL, employing the model specified in Fig. 1. Two runs used LISREL's polychoric procedure: one for dichotomized data and one for five category data. Two runs used Pearson product moment correlations based on equal distance scoring: one for dichotomized data and one for five category data. The fifth type of run consisted of using Pearson correlations for the continuous data (raw data from the simulations). With the three levels of skew and kurtosis, three levels of correlation between the indicator variables and constructs (unobserved variables), three levels of correlations between the constructs, and five types of input for each of these conditions, 135 separate LISREL runs were necessary. Two criteria were used to evaluate the accuracy of the polychoric-based and Pearson-based estimates of the simulated value of phi (the

Table 1. Average absolute error in estimating phi using Pearson product moment and polychoric correlations in single and multiple indicator models.

Phi	Single indicator models			Multiple indicator models		
	Pearson	Polychoric	Difference	Pearson	Polychoric	Difference
			Continuous[a]			
0.3	0.162	0.162	0.000	0.050	0.050	0.000
0.6	0.362	0.362	0.000	0.040	0.040	0.000
0.8	0.495	0.495	0.000	0.028	0.028	0.000
Total	0.340	0.340	0.000	0.040	0.040	0.000
			5 Categories			
0.3	0.166	0.153	0.013	0.065	0.067	−0.002
0.6	0.369	0.343	0.026	0.060	0.061	−0.001
0.8	0.488	0.457	0.031	0.053	0.052	0.001
Total	0.341	0.318	0.023	0.059	0.060	−0.001
			Dichotomous			
0.3	0.199	0.156	0.043	0.069	0.080	−0.011
0.6	0.428	0.341	0.087	0.083	0.084	−0.001
0.8	0.572	0.463	0.109	0.088	0.077	0.011
Total	0.400	0.320	0.080	0.080	0.080	0.000
			Skew groups			
Low	0.376	0.340	0.036	0.068	0.059	0.009
High	0.370	0.341	0.029	0.043	0.044	−0.001
Pos/Neg	0.335	0.303	0.032	0.068	0.079	−0.011

[a] These values for the Pearson and polychoric correlations are the same in the continuous case.

correlations between the two ksi's) in the model in Fig. 1: the average signed bias and the average absolute deviation.

Results

Table 1 provides a summary of the average absolute error of the pearson-based and polychoric-based estimates for both the single indicator and multiple indicator models.[3] These results are reported separately for different numbers of categories (i.e., continuous, five categories, and dichotomous) and different correlations between underlying constructs (i.e., 0.3, 0.6, 0.8). The effects of different skewness and kurtosis combinations (i.e., low, high, and high positive-high negative) are summarized in the bottom panel of the table. Differences between the average absolute errors for the Pearson-based and polychoric-based estimates for both single and multiple indicator models also are indicated. These represent the absolute amount of improvement, in terms of absolute error, obtained by using polychoric rather than Pearson correlations. The results obtained using either the absolute error or average bias were very similar, since the estimates of the underlying correletions among the constructs (phi's) were almost always too low. Thus, we report only the average absolute error.

Single indicator models
As expected the polychoric-based estimates provide a substantial improvement in the estimation of phi (over the Pearson-based estimates) in the single indicator models. This improvement is most marked for the dichotomized data where it averages 0.080; the improvement is less in the five category situation where it averages 0.023. The comparisons for the continuous case are not very meaningful, since in the continuous case polychoric and Pearson-based estimates should converge.[4] In spite of the improvement gained from polychoric estimates for the single indicator models, LISREL still grossly underestimates phi in all conditions, whether or not polychoric correlations are used.

Multiple indicator models
These models effectively reduce the error in estimating phi, whether polychoric or Pearson-based correlations are used. For example, in the dichotomous situation the average error is reduced from 0.400 to 0.080 for the Pearson-based estimates and from 0.320 to 0.080 for the polychoric-based estimates. The impressive performance of the multiple indicator model in dealing with crude rank category measures relative to the single indicator model is unambiguous; average absolute error is greatly reduced when a multiple indicator model is substituted for a single indicator model.

Comparisons of the relative accuracy of the polychoric and Pearson-based estimates is quite informative. In the multiple indicator models, there is virtually no difference in their accuracy in estimating the correlation (phi) between the two unobserved variables.

Continuous variables
When the correlations among the observed variables are based on continuous variables, the absolute errors at different levels of correlation between the unobserved constructs reveals some interesting patterns. For single indicator models, the average absolute error of the estimate increases substantially as phi increases; however, this increase is not great in percentage terms. That is, the absolute error when phi is 0.3 is 0.162 or 54% of the value of phi,

when phi is 0.6 the estimate is off on an average of 60%, and when it is 0.80 it is off by an average of 62%. The average absolute error for the multiple indicator model is relatively small at all levels of phi, and in absolute and percentage terms is smaller the greater the underlying value of phi.

Five categories

For five category rank measures, the polychoric procedure performed only somewhat better than the Pearson estimates for the single indicator model (e.g., on the average it decreased the error in estimating phi by 0.023). Again more error was made in an absolute sense the greater the underlying value of phi, but in percentage terms this difference is not too great.

The polychoric and Pearson estimates performed very similarly for the multiple indicator models. Again, in percentage terms, the greater the value of phi the less the error. The better performance of the polychoric-based estimates in the single indicator models occurs because there is no correction for crude rank category measures in this situation for the Pearson-based estimates while there is for the polychoric-based estimates. In the multiple indicator model, however, the Pearson correlations *are* corrected, albeit indirectly, for the measurement error resulting from crude categorization; LISREL multiple indicator models automatically correct for measurement error, which is estimated using the correlations among indicators. Since these correlations are lower for the Pearson-based than polychoric-based estimates, the absolute value of phi is increased accordingly.

Dichotomies

When Pearson-based correlations are used, the average absolute error is greater for dichotomies than continuous or five category measures. Once again, the average amount of error is less for the single indicator models that use polychoric-based correlations rather than Pearsonian correlations at all levels phi, and the same pattern of greater absolute error as phi increases is evident.

The absolute error for the multiple indicator models is quite similar whether Pearson-based or polychoric-based correlations are used. Again, there is a very substantial increase in accuracy when a multiple indicator rather than single indicator model is used.

In all cases, whether Pearson-based or polychoric-based correlations are used, the multiple indicator models, which are clearly the models of choice, work better as we move from dichotomies to five category polytomies to continuous measurement. The increase in precision, however, is not as great as one might expect, since much of the error of estimation is removed in moving from the single indicator to a multiple indicator model.

Skewness levels

Table 1 also presents comparisons of Pearson-based and polychoric-based estimates in single and multiple indicator models with three types of skewness. Across all the conditions of skewness and kurtosis simulated, polychoric-based correlations improve upon Pearson-based estimates for the single indicator models. Multiple indicator models substantially reduce the average absolute error, whether polychoric or Pearson-based estimates are used.

Conclusions

It has been suggested that one method to reduce distortion caused by the use of rank category measures of underlying continuous variables is with a LISREL multiple indicator model (Johnson and Creech, 1983). A series of multivariate nonnormal distributions were generated with a range of skewness and kurtosis, correlations among indicators, and correlations between underlying constructs. In our simulations, multiple indicator models that used crude rank category indicators resulted in significant reductions in the average absolute error of the estimates of underlying correlations between constructs when compared to single indicator models.

The accuracy of the estimates of the correlation between constructs provided by LISREL were compared for both equal distance-based correlations and polychoric correlations in multiple and single indicator models. The results of these comparisons may be summarized as follows:

1. Under all levels of skewness, correlations between underlying constructs, and number of categories of measurement, the multiple indicator model greatly reduced the average absolute error involved in estimating the underlying correlation in comparison to a single indicator model: independent of the method of calculating correlations (i.e., Pearson or polychoric).
2. Polychoric correlations are superior to Pearson correlations in single indicator models, but they still result in large errors. The use of polychoric correlations, rather than Pearson (equal distance based) correlations, does not improve estimates (of phi) in the multiple indicator model. It should be noted that O'Brien and Homer (1987) found that when the correlations between variables were very low (e.g., phi = 0.12), the polychoric estimates were inferior to the Pearsonian estimates.
3. For the single indicator model, the average absolute error of the estimate increases substantially as the underlying "true" correlation increases, but in percentage terms this relationship is much smaller.

4. For the multiple indicator model, the average absolute error of the estimate decreases both absolutely and relatively as the underlying "true" correlation increases for continuous and 5 category rank measures. In the dichotomous case, the average absolute error decreases relatively, i.e., as the value of phi increases, the percentage by which the estimate is in error decreases [percent error = ((phi estimate − phi)/phi) × 100].

This research involved only a sample of the possible combinations of skewness and kurtosis, variable correlations, and underlying correlations between constructs for a single multiple indicator model. The results, however, are impressive evidence of the power of multiple indicator models to reduce measurement error due to the use of crude rank category measures, whether or not LISREL's polychoric procedure is used.

Notes

1. For example, income, a common demographic indicator in many models, tends to follow a Pareto distribution when measured continuously. Other variables with potentially non-normal distributions that are usually measured categorically include: level of education, physical demands of a job, and many attitudinal items.
2. All measures of skewness and kurtosis are based on Fisher's statistics (Bliss, 1967).
3. The multiple indicator and single indicator models used are depicted in Figure 1. Note that for the single indicator model variables X_3 and X_5 were used from the original set of six variables. That is, for all sets of X_1, X_2, and X_3 generated, the highest correlation in that set was used; and for all sets of X_4, X_5, and X_6 the lowest correlation in that set was used. Similar, though slightly less dramatic, results were obtained when variables with the highest correlations in each set were used.
4. We used the Pearson-based estimates for the polychoric ones, since LISREL will estimate polychoric based estimates for variables that have a minimum of two and a maximum of nine categories.

References

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

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

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

D.R. Johnson & J.C. Creech. (1983). "Ordinal measures in multiple indicator models: a simulation study of categorization error", *American Sociological Review* 47: 398–407.

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

D.R. Lehmann & J. Hulbert. (1982). "Are three-point scales always good enough?" *Journal of Marketing Research* 19: 444–445.

W.S. Martin. (1973). "The effects of scaling on the correlation coefficient: a test of validity", *Journal of Marketing Research* 10: 316–318.

W.S. Martin. (1978). "Effects of scaling on the correlation coefficient: additional considerations", *Journal of Marketing Research* 15: 304–308.

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

R.M. O'Brien. (1981b). "Using rank category variables to represent continuous variables: defects of common practice", *Social Forces* 61: 144–155.

R.M. O'Brien. (1983). "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: 284–286.

R.M. O'Brien & P.H. Homer. (1987). "Corrections for coarsely categorized measures: LISREL's polyserial and polychoric correlations", *Quality and Quantity* 21: 349–360.

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

U. Olsson, F. Drasgow & N.J. Dorars. (1982). "The polyserial correlation coefficient", *Psychometrika* 47: 337–347.

C. Vale, C. David & V.A. Maurelli. (1983). "Simulating multivariate nonnormal distributions", *Psychometrika* 48: 465–474.