Testing Categorized Bivariate Normality With Two-Stage Polychoric Correlation Estimates

Alberto Maydeu-Olivares, Carlos García-Forero, David Gallardo-Pujol, and Jordi Renom

Faculty of Psychology, University of Barcelona, Spain

Abstract. Structural equation modeling (SEM) with ordinal indicators rely on an assumption of categorized normality. This assumption may be tested for pairs of variables using the likelihood ratio G^2 or Pearson's X^2 statistics. For increased computational efficiency, SEM programs usually estimate polychoric correlations in two stages. However, two-stage polychoric estimates are not asymptotically efficient and G^2 and X^2 need not be asymptotically chi-square when the estimator is not efficient. Recently, Maydeu-Olivares and Joe (2005) have introduced a new statistic, M_n, that is asymptotically chi-square even for estimators that are not efficient. We investigate the behavior of G^2, X^2, and M_n when testing underlying bivariate normality with polychoric correlations estimated in two stages.

Keywords: LISREL, MPLUS, categorical data analysis, goodness of fit

Structural equation modeling (SEM) has become a mainstream procedure in the social sciences. In its beginnings, SEM revolved around modeling continuous dependent variables. Later, these models were extended to accommodate ordinal indicators (such as responses to ratings). Thus, for instance, when fitting a factor model to rating data it was no longer necessary to assume that the relationship between the ratings and the latent factors was linear; rather, the ordinal nature of the ratings could be taken into account (Coffman, Maydeu-Olivares, & Arnau, 2008). To do so, it is assumed that each observed ordinal variable is related to a latent response variable via a threshold response process, and that latent response variables are linearly related to the common factors. By assuming that the latent response variables are normally distributed, the model can be very efficiently estimated using a series of stages. First, the correlations between the latent response variables are computed. These are polychoric correlations. In a subsequent stage, the parameters of the model (for instance, factor loadings and interfactor correlations) are estimated from the polychoric correlations using a least squares procedure. Several computer programs for SEM implement multi-stage procedures for estimating models involving ordinal indicators (e.g., Lisrel: Jöreskog & Sörbom, 2001; Mplus: Muthén & Muthén, 2007; Eqs: Bentler, 2006). In their standard output, these programs provide a goodness of fit test of the restrictions imposed by the model on the polychoric correlations. However, these tests are only meaningful if the assumptions involving the use of polychoric correlations hold. In other words, the goodness of fit tests provided by standard software programs are only valid if the observed ordinal data arise by discretizing a multivariate normal distribution (Muthén, 1993).

There is now an extensive literature on the use of sequential procedures to estimate and test the structural restrictions (i.e., the restrictions imposed by the model on the polychoric correlations) in structural equation models when all the indicators are ordinal. Some key references are Bermann (1993), Christoffersson (1975, 1977), Christoffersson and Gunsjö (1983), Gunsjö (1994), Jöreskog (1994), Jöreskog and Moustaki (2001), Küsters (1987), Lee, Poon, and Bentler (1990, 1992, 1995), Maydeu-Olivares (2006), Muthén (1978, 1982, 1984, 1993), Muthén and Christoffersson (1981), Muthén and Satorra (1995), Muthén, du Toit, and Spisic (1997), and Quiroga (1992). However, very little research has focused on how to assess the distributional restrictions (i.e., whether the ordinal data arise by discretizing a multivariate normal distribution). Recently, Maydeu-Olivares (2006) – building on work by Satorra and Bentler (1994) – has proposed a procedure that may be used to assess whether the ordinal data arise by discretizing a multivariate normal distribution. However, this procedure has not been implemented in standard SEM software. Furthermore, if the test proposed by Maydeu-Olivares (2006) rejects the model of underlying normality it is necessary to assess the source of the misfit, that is, to determine the variables for which the assumption of underlying normality is violated. In the case of variables involving more than two response alternatives, underlying normality can be assessed for pairs of variables. When the variables consist only of two response alternatives, the assessment may be performed using the procedure suggested by Muthén and Hofacker (1988) for triplets of variables.

The aim of this study is to identify the best statistic for assessing underlying bivariate normality. The statistic to

DOI: 10.1027/1614-2241.5.4.131

be used depends on how the polychoric correlations are estimated. For increased computational efficiency, in Mplus and Lisrel, polychoric correlations are estimated in two stages using a procedure due to Olsson (1979). In the first stage, the thresholds are estimated for each variable separately by maximum likelihood. In the second stage, each polychoric correlation is estimated separately for each pair of variables also by maximum likelihood given the thresholds estimated in the first stage. This study focuses on how to assess bivariate normality when polychoric correlations are estimated using Olsson's (1979) two-stage procedure, as implemented in Mplus and Lisrel.

Testing Underlying Bivariate Normality With Two-Stage Polychoric Correlation Estimates

Consider two ordinal variables consisting of I and J response alternatives, respectively. The observed data can therefore be gathered in an $I \times J$ contingency table. When estimating a polychoric correlation from this table we assume that the ordinal variables arise from a bivariate standard normal distribution categorized according to $(I - 1)$ and $(J - 1)$ thresholds, respectively. The parameters to be estimated, $\boldsymbol{\kappa}$, consist of the $(I - 1)$ thresholds from one variable, $\boldsymbol{\tau}_1$, and the $(J - 1)$ thresholds from the other variable, $\boldsymbol{\tau}_2$, plus the estimated polychoric correlation, ρ. There are $q = (I - 1) + (J - 1) + 1$ parameters to be estimated in this model.

The two most widely used goodness of fit statistics in categorical data analysis are Pearson's X^2 statistic and the likelihood ratio statistic G^2,

$$X^2 = N \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(p_{ij} - \pi_{ij}(\hat{\boldsymbol{\kappa}})\right)^2}{\pi_{ij}(\hat{\boldsymbol{\kappa}})}, \tag{1}$$

$$G^2 = N \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \ln \frac{p_{ij}}{\pi_{ij}(\hat{\boldsymbol{\kappa}})}, \tag{2}$$

where N denotes sample size, and p_{ij} and π_{ij} denote the bivariate proportions and probabilities. The latter depend on the q-dimensional vector of estimated parameters (thresholds and polychoric correlation). Also, in Equation 2 by convention when $p_{ij} = 0$, $p_{ij} \ln \frac{p_{ij}}{\pi_{ij}(\hat{\boldsymbol{\kappa}})} = 0$. To assess underlying bivariate normality, G^2 is used in Lisrel, whereas Mplus does not currently provide tests (Muthén, 2007).

Olsson (1979) considered one- and two-stage maximum likelihood approaches to estimate the thresholds and polychoric correlation from the observed bivariate contingency table. In the one-stage approach, all parameters are estimated simultaneously. In the two-stage approach, the thresholds are estimated separately from each univariate marginal, then the polychoric correlation is estimated from the bivariate table using the thresholds estimated in the first stage. The one- and two-stage polychoric estimates are most often quite close. Since two-stage polychoric estimates are faster to compute, polychoric correlations are estimated using this two-stage procedure in Lisrel and Mplus. When polychoric correlations are estimated in one stage the estimates are consistent, asymptotically normal and asymptotically efficient (they have minimum variance asymptotically). When they are estimated in two stages, the estimates are still consistent and asymptotically normal (Jöreskog, 1994; Muthén & Satorra, 1995; Olsson, 1979) but they are no longer asymptotically efficient. Thus, the two-stage estimates trade increasing computational efficiency for some (hopefully small) loss of statistical efficiency.

Now, from standard theory (e.g., Agresti, 2002), when the parameters are simultaneously estimated by maximum likelihood – as in Olsson's one-stage procedure – both X^2 and G^2 are asymptotically distributed as a chi-square with $r = IJ - q - 1 = IJ - I - J$ degrees of freedom (*df*). However, when the estimator is not asymptotically efficient, X^2 and G^2 still have the same asymptotic distribution, which is generally a mixture of independent chi-squares with one degree of freedom. This means that when polychoric correlations are estimated in two stages, a chi-square need not be the appropriate reference distribution for these statistics.

Maydeu-Olivares and Joe (2005, 2006) have recently introduced a test statistic, M_n, which is asymptotically chi-square for any consistent and asymptotically normal estimator with *df* equal to the number of probabilities minus the number of estimated parameters minus one. Thus, M_n is in principle more suitable than X^2 and G^2 for assessing underlying bivariate normality when the polychoric correlations have been estimated using Olsson's two-stage estimator. Let $\boldsymbol{\pi}$ be the IJ vector of probabilities, $\mathbf{p}$ be its associated vector of sample proportions, $\mathbf{D}$ be a diagonal matrix with $\boldsymbol{\pi}$ along its diagonal, and $\boldsymbol{\Delta}$ be the $IJ \times q$ matrix of derivatives $\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}'}$. M_n can be written as

$$M_n = N(\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\kappa}}))' \left(\hat{\mathbf{D}}^{-1} - \hat{\mathbf{D}}^{-1} \hat{\boldsymbol{\Delta}} \left(\hat{\boldsymbol{\Delta}}' \hat{\mathbf{D}}^{-1} \hat{\boldsymbol{\Delta}} \right)^{-1} \hat{\boldsymbol{\Delta}}' \hat{\mathbf{D}}^{-1} \right) \times (\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\kappa}})). \tag{3}$$

This statistic is closely related to Pearson's X^2. In matrix form, X^2 can be written as

$$X^2 = N(\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\kappa}}))' \hat{\mathbf{D}}^{-1} (\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\kappa}})). \tag{4}$$

Thus, M_n applies a correction to X^2 to take into account that the estimator need not be statistically efficient. In the special case of maximum likelihood estimation, M_n equals X^2 algebraically (Maydeu-Olivares & Joe, 2005). But for estimators that are not efficient – such as Olsson's two-stage estimator – M_n is stochastically smaller than X^2 and therefore the use of X^2 instead of the correct statistic M_n may yield an impression of poor fit. The difference between M_n and X^2, however, will depend on the efficiency of the estimator. The more efficient the estimator, the closer M_n and X^2 will be. In the next sections we investigate whether M_n actually yields better results than X^2 and G^2 for assessing underlying bivariate normality when polychoric correlations have been estimated in two stages.

A Numerical Example

Agresti (1992) asked 61 respondents to compare the taste of Coke, Classic Coke, and Pepsi using a five-point preference scale in a paired comparison design {Coke vs. Classic Coke, Coke vs. Pepsi, Classic Coke vs. Pepsi}. The categories were {"Strong preference for *i*", "Mild preference for *i*", "Indifference", "Mild preference for *i′*", and "Strong preference for *i′*"}. For each pair of variables, we shall test the assumption that the observed 5 × 5 table arises by categorizing a standard bivariate normal density. That is, we are interested in testing a substantive hypothesis of normally distributed continuous preferences for the soft drinks in the population.

In Table 1 we provide the thresholds and polychoric correlation for each pair of variables estimated in two stages and the asymptotic standard errors of these parameters. The latter were computed as in Jöreskog (1994).

In Table 1 we also provide goodness of fit results for the two-stage estimates using G^2, X^2, and M_n. Inspecting the goodness of fit tests, we first notice that for all three bivariate tables all test statistics suggest that the assumption of categorized bivariate normality is reasonable. We also notice that the estimated G^2 statistics are larger than the M_n and X^2 statistics. This is because we purposely chose a numerical example with a very small sample size to highlight the differences between the statistics. The estimated G^2 statistics are larger because there are some empty cells in the observed bivariate table which are not included in the computation of G^2. On the other hand, all cells are used in the computation of both M_n and X^2.

The most surprising fact in Table 1 is that the values for the asymptotically correct M_n and the asymptotically incorrect X^2 are rather close. This is because the values of M_n and X^2 will be very close if the estimator used is very close to being efficient. With these data, this is precisely the case. The two-stage estimator is so highly efficient that it is irrelevant for practical purposes whether M_n or X^2 is used. To see this, in Table 2 we provide the results obtained when the model is estimated using one-stage maximum likelihood. Note that in this case, the thresholds estimated from different bivariate tables need not be the same across tables. We see in Table 2 that the parameter estimates and their standard errors for these tables are indeed very similar to those obtained using the two-stage approach. Because the one and two-stage estimates shown in Tables 1 and 2 are so close, the G^2 and X^2 values obtained using the one- and two-stage estimates are also very close.

To gain further insight into the efficiency of the two-stage estimator relative to the one-stage estimator, we computed the population asymptotic covariance matrix of the one-stage and two-stage estimators at population values similar to those encountered in the example using the formulae from Jöreskog (1994). The population parameter values used were $\boldsymbol{\tau}_1 = (-1, -0.5, 0.5, 1)'$, $\boldsymbol{\tau}_2 = (-1, -0.5, 0.5, 1)'$, and $\rho = 0.3$. At these values, the determinant of the asymptotic covariance matrix of the two-stage estimates is only 2.5% larger than the determinant of the asymptotic covariance matrix of the one-stage estimates. Also, the population asymptotic variances of the two-stage parameter estimates are less than 1% larger than for the one-stage parameter estimates. Yet, in our implementation, the two-stage estimates are on average 17 times faster to compute than the one-stage estimates. Thus, the two-stage estimator is considerably faster than the one-stage estimator with a very small loss of asymptotic efficiency.

A Small Simulation Study

To further investigate the small sample performance of G^2, X^2, and M_n we performed a simulation study using the above population values: $\boldsymbol{\tau}_1 = (-1, -0.5, 0.5, 1)'$, $\boldsymbol{\tau}_2 = (-1, -0.5, 0.5, 1)'$, and $\rho = 0.3$. We considered six conditions obtained by crossing three levels of sample size (N = 50, 100, and 1,000) and two model conditions. In the first condition, data were generated using a bivariate normal distribution. In the second condition, data were generated using a bivariate t distribution with 5 *df*. The first model condition enables us to investigate the behavior of the statistics when the model is correctly specified. In this condition, empirical rejection rates should be as close as possible to the nominal rates. The second model condition enables us to investigate empirically the power of the statistics when the model is mispecified. In this case, rejection rates should be as large as possible. Notice, however, that it is only meaningful to compare the rejection rates of two

Table 1. Two-stage parameter estimates, estimated standard errors, and goodness of fit tests for Agresti's soft drink data

	1	2	3	4
Thresholds				
Var. 1	−0.796 (0.180)	0.062 (0.161)	0.539 (0.169)	1.510 (0.248)
Var. 2	−0.914 (0.187)	−0.062 (0.161)	0.446 (0.166)	1.202 (0.211)
Var. 3	−1.202 (0.211)	−0.492 (0.168)	0.103 (0.161)	0.853 (0.184)

Vars.	Corr.	G^2	p value	X^2	p value	M_n	p value
Correlations and test statistics							
(2,1)	0.103 (0.140)	21.286	0.128	16.478	0.351	16.478	0.351
(3,1)	−0.347 (0.129)	18.484	0.238	14.358	0.499	14.352	0.499
(3,2)	0.005 (0.141)	18.569	0.234	15.898	0.389	15.898	0.389

Note. N = 61; standard errors in parentheses; 15 *df*.

Table 2. Joint parameter estimates, estimated standard errors, and goodness of fit tests for Agresti's soft drink data

	1	2	3	4
Thresholds				
Var. 2	−0.915 (0.187)	−0.063 (0.161)	0.445 (0.166)	1.202 (0.211)
Var. 1	−0.797 (0.180)	0.061 (0.161)	0.539 (0.161)	1.511 (0.249)
Var. 3	−1.201 (0.211)	−0.484 (0.167)	0.104 (0.160)	0.849 (0.184)
Var. 1	−0.800 (0.180)	0.064 (0.160)	0.538 (0.160)	1.510 (0.250)
Var. 3	−1.202 (0.211)	−0.492 (0.168)	0.103 (0.161)	0.853 (0.184)
Var. 2	−0.914 (0.187)	−0.062 (0.161)	0.446 (0.166)	1.202 (0.211)

Vars.	Corr.	G^2	p value	X^2	p value
Correlations and test statistics					
(2,1)	0.103 (0.144)	21.29	0.128	16.476	0.351
(3,1)	−0.347 (0.121)	18.48	0.238	14.371	0.498
(3,2)	0.005 (0.152)	18.57	0.234	15.915	0.388

Note. $N = 61$; standard errors in parentheses; 15 *df*.

statistics under model misspecification when the empirical rejection rates for correct model misspecification are similar across the two statistics. Also, here we considered a very small model misspecification. The t distribution with 5 *df* used in the model misspecification condition is very similar to the standard normal distribution assumed by the statistics. Like the normal distribution, the $t(5)$ distribution is symmetric, but its kurtosis is larger: 3 for a standard normal distribution and 9 for a $t(5)$ distribution. One thousand replications were used for each of the six conditions. The results of the simulation are shown in Table 3. The mean and variance of each statistic are displayed in this table, along with empirical rejection rates at selected rates: 1%, 5%, 10% and 20%.

We see in this table that when underlying bivariate normality holds, G^2 tends to reject too often the null hypotheses when $N = 50$ and also when $N = 100$. On the other hand, the behavior of X^2 and M_n is acceptable even when $N = 50$ in these 5×5 contingency tables. We also see in this table that the empirical distributions of X^2 and M_n are very similar for all sample sizes, with M_n taking slightly smaller values, as stated by the asymptotic theory. Thus, the small sample behavior of M_n relative to X^2 matches the asymptotic efficiency results for the two-stage estimates at these parameter values.

When the underlying model is incorrect, the statistics have no power to detect the very minor model misspecification used when sample size is 50 and 100. Notice, however, that the statistics are quite powerful when sample size is 1,000; empirical rejection rates are 50% at the nominal rate 5%. We also see in this table that rejection rates for M_n and X^2 are again virtually undistinguishable. It does not matter which of these two statistics is used in terms of power.

Discussion

The purpose of this research was to investigate which statistic should be used to assess categorized bivariate normality when – as is most often the case – polychoric correlations are estimated in two stages. Three statistics were considered, the likelihood ratio test G^2, Pearson's X^2, and a new statistic,

Table 3. Simulation results: Mean, variance of the statistics, and rejection rates when the null model is correctly specified (underlying bivariate normal distribution) and when it is mispecified (underlying bivariate t distribution with 5 *df*)

		Bivariate normal distribution						Bivariate t distribution with 5 *df*					
N	Stat.	Mean	Var.	1%	5%	10%	20%	Mean	Var.	1%	5%	10%	20%
50	G^2	17.67	28.92	1.8	9.3	18.2	35.9	18.37	32.03	3.4	11.3	21.4	39.5
	X^2	15.45	26.86	1.1	4.2	9.4	20.6	15.80	27.71	1.5	5.8	10.3	22.4
	M_n	15.42	26.75	1.1	4.0	9.2	20.5	15.73	27.65	1.5	5.8	10.0	21.8
100	G^2	16.94	32.16	1.9	8.8	16.1	29.2	17.74	34.47	2.5	11.7	19.9	35.2
	X^2	15.13	26.39	0.5	4.0	9.0	18.9	15.98	29.29	1.5	6.2	11.1	25.0
	M_n	15.10	26.29	0.5	3.9	8.7	18.9	15.96	29.19	1.4	6.2	11.0	24.7
1,000	G^2	15.04	30.71	0.8	5.8	10.9	20.8	25.75	70.71	25.8	50.4	63.6	76.3
	X^2	14.92	29.99	0.8	5.7	10.6	19.5	25.94	73.11	25.7	50.7	63.7	76.5
	M_n	14.89	29.89	0.8	5.6	10.5	19.3	25.92	73.00	25.7	50.5	63.7	76.5

Note. 1,000 replications per condition; 15 *df*; $\tau_1 = (-1,-0.5,0.5,1)'$, $\tau_2 = (-1,-0.5,0.5,1)'$, $\rho = 0.3$.

M_n due to Maydeu-Olivares and Joe (2005, 2006). M_n is in principle better suited to the task as G^2 and X^2 are not asymptotically chi-square when polychoric correlations are estimated in two stages because the estimator is not asymptotically efficient, but M_n is asymptotically chi-square even when estimators are not asymptotically efficient. Yet, our – admittedly limited – results suggest that in many instances all three statistics investigated provide good results in this setting. Of the three statistics considered, G^2 yields a poorer performance in small samples, but the behavior of X^2 and M_n is – for practical purposes – indistinguishable. G^2 yields inflated p values when the contingency table is sparse. This result is well known in the literature (e.g., Koehler & Larntz, 1980) and we have shown here that it may also occur when assessing bivariate normality. The behavior of X^2 should not be affected by sparseness as only bivariate tables are considered here, but it may be affected by the lack of statistical efficiency of the estimator. M_n corrects X^2 for the lack of statistical efficiency of the estimator. The lower the efficiency of the estimator, the larger the discrepancy between M_n and X^2. If the estimator is not highly efficient, using X^2 in place of M_n may yield a false impression of poor fit. Previous research (e.g., Jöreskog, 2002; Olsson, 1979) suggests that the two-stage polychoric estimator is highly efficient. Here, using a numerical example and a small simulation we have shown that the lack of statistical efficiency does not in fact affect the behavior of X^2. Indeed, for parameter values where the two-stage estimator is highly efficient (and hence the two-stage estimates are very close to the one-stage estimates) the estimated X^2 and M_n statistics will be very close. In this case, the empirical type I errors of the statistics under the null hypothesis, as well as their power under violations of bivariate normality, will be very close.

To summarize, our results suggest that X^2 (but not G^2) may be safely used to assess bivariate normality when polychoric correlations have been estimated in two stages. Yet, it may be inappropriate to present results from a few points in the parameter space and to attempt draw generalizable conclusions from them. Indeed, there may be some points in the parameter space where the two-stage estimator is less efficient than in the cases considered here. Because the expressions involved in computing M_n are actually a side product of the computations needed to obtain the two-stage estimates and their asymptotic covariance matrix (see Jöreskog, 1994), it may be worthwhile to use the asymptotically correct M_n instead of X^2 to routinely assess bivariate normality when two-stage polychorics are used.

Acknowledgment

This research was supported by grant SEJ2006-08204 from the Spanish Ministry of Education.

References

Agresti, A. (1992). Analysis of ordinal paired comparison data. *Applied Statistics, 41*, 287–297.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York, NY: Wiley.

Bentler, P. M. (2006). *EQS structural equation modeling software*. Encino, CA: Multivariate Software.

Bermann, G. (1993). *Estimation and inference in bivariate and multivariate ordinal probit models*. Uppsala, Sweden: Acta Universitatis Upsaliensis, Studia Statistica Upsaliensia 1.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40*, 5–32.

Christoffersson, A. (1977). Two-step weighted least squares factor analysis of dichotomized variables. *Psychometrika, 42*, 433–438.

Christoffersson, A., & Gunsjö, A. (1983). *Analysis of structures for ordinal data*. Uppsala, Sweden: University of Uppsala. Department of Statistics (Research Report 83-2).

Coffman, D. L., Maydeu-Olivares, A., & Arnau, J. (2008). Asymptotic distribution free interval estimation for an intraclass correlation coefficient with applications to longitudinal data. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 4*, 4–9.

Gunsjö, A. (1994). *Faktoranalys av ordinala variabler*. Stockholm, Sweden: Almqvist & Wiksell, Acta Universitas Upsaliensis. Studia Statistica Upsaliensia 2.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59*, 381–390.

Jöreskog, K. G. (2002). *Structural equation modeling with ordinal variables using LISREL*. Available from www.ssicentral.com/lisrel/techdocs.ordinal.pdf.

Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research, 36*, 347–387.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8*. Chicago, IL: Scientific Software.

Küsters, U. L. (1987). *Hierarchische Mittelwert- und Kovarianzstrukturmodelle mit nichtmetrischen endogenen Variablen*. Heidelberg: Physica-Verlag.

Koehler, K., & Larntz, K. (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials. *Journal of the American Statistical Association, 75*, 336–344.

Lee, S. Y., Poon, W. Y., & Bentler, P. M. (1990). A three-stage estimation procedure for structural equation models with polytomous variables. *Psychometrika, 55*, 45–57.

Lee, S. Y., Poon, W. Y., & Bentler, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika, 57*, 89–105.

Lee, S. Y., Poon, W. Y., & Bentler, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology, 48*, 339–358.

Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika, 71*, 57–77.

Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association, 100*, 1009–1020.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*, 713–732.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551–560.

Muthén, B. (1982). Some categorical response models with continuous latent variables. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation* (Vol. 1, pp. 65–79). Amsterdam: North Holland.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115–132.

Muthén, B. (1993). Goodness of fit with categorical and other non normal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205–234). Newbury Park, CA: Sage.
Muthén, B. (2007). *Testing bivariate normal assumption* Retrieved, April 8, 2007 from Mplus Discussion. http://www.statmodel.com/discussion/messages/9/2674.html1193413810.
Muthén, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46*, 407–419.
Muthén, B., & Hofacker, C. (1988). Testing the assumptions underlying tetrachoric correlations. *Psychometrika, 53*, 563–578.
Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished manuscript.
Muthén, L., & Muthén, B. (2007). *MPLUS (version 5)*. Los Angeles, CA: Muthén & Muthén.
Muthén, B., & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika, 60*, 489–503.
Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*, 443–460.
Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications to developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
Quiroga, A. M. (1992). Studies of the polychoric correlation and other correlation measures for ordinal variables. Universitatis Upsaliensis, Unpublished doctoral dissertation.

Alberto Maydeu-Olivares

Faculty of Psychology
University of Barcelona
P. Valle de Hebrón
171. 08035 Barcelona
Spain
E-mail amaydeu@ub.edu

