Testing categorized bivariate normality with two-stage polychoric correlation estimates

Albert Maydeu-Olivares

Dept. of Psychology

University of Barcelona

Address correspondence to: Albert Maydeu-Olivares. Faculty of Psychology. University of Barcelona. P. Valle de Hebrón, 171. 08035 Barcelona (Spain). E-mail: amaydeu@psi.ub.es.
This research was supported by grant BSO2000-0661 of the Spanish Ministry of Science and Technology. I'm indebted to Karl Jöreskog for his encouragement to write this manuscript and for his comments to a previous draft.

Abstract

We show that when the thresholds and the polychoric correlation are estimated in two stages, neither Pearson's X^2 nor the likelihood ratio G^2 goodness of fit test statistics are asymptotically chi-square. We propose a new statistic that is asymptotically chi-square in this situation.

However, our experience -illustrated with a numerical example- and a small simulation study using a 5×5 table suggest that in finite samples our new test is not able to outperform X^2. Both match rather well a reference chi-square distribution even in small samples. G^2, on the other hand, was found to be conservative in small samples. Thus, in practice it is justified to use X^2 to test bivariate normality when the parameters are estimated in two stages, even in small samples. Larger samples may be needed to use G^2.

We attribute our empirical results to the very small numerical differences found when X^2 and G^2 are computed using two or one-stage estimates, where in the latter case it is theoretically justified to use these statistics.

Keywords: LISREL, LISCOMP, MPLUS, categorical data analysis, pseudo-maximum likelihood estimation, multinomial models, sparse tables.

1. Introduction

Consider a bivariate standard normal density categorized according to (I - 1) and (J - 1) thresholds, respectively. Within a maximum likelihood framework, Olsson (1979) considered one and two-stage approaches to estimate the $q = (I - 1) + (J - 1) + 1$ parameters of this model from the observed $I \times J$ 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.

Of course, after estimating the parameters one must test the model (Muthén, 1993). To this end, one may employ Pearson's X^2 test statistic or the likelihood ratio statistic G^2. From standard theory (e.g, Agresti, 1990), when the one-stage approach is employed both statistics are asymptotically distributed as a chi-square with $r = IJ - q - 1 = IJ - I - J$ degrees of freedom. However, ever since Olsson (1979) concluded that very similar results are obtained with the computationally simpler two-stage approach, this approach has become the standard procedure for estimating this model. As such, it is the procedure implemented in computer programs such as LISCOMP (Muthén, 1987), PRELIS/LISREL (Jöreskog & Sörbom, 1993) and MPLUS (Muthén & Muthén, 1998). Yet, the distribution of X^2 and G^2 is presently unknown when the two-stage approach is employed. This paper fills in this gap using large sample theory.

2. Asymptotic distribution of parameter estimates

Consider a $I \times J$ contingency table. Let $\boldsymbol{\pi}_{12} = \left(\pi_{11}, \cdots, \pi_{1J}, \cdots, \pi_{i1}, \cdots, \pi_{iJ}, \cdots, \pi_{I1}, \cdots, \pi_{IJ}\right)'$ denote its IJ vector of probabilities and $\mathbf{p}_{12}$ its associated vector of sample proportions.

Furthermore, let $\boldsymbol{\pi}_1 = \left(\pi_1, \cdots, \pi_i, \cdots, \pi_I\right)'$ and $\boldsymbol{\pi}_2 = \left(\pi_1, \cdots, \pi_j, \cdots, \pi_J\right)'$ denote the vectors of univariate marginal probabilities, and $\mathbf{p}_1$ and $\mathbf{p}_2$ the vectors of its associated sample proportions.

We note that,

$$\boldsymbol{\pi}_1 = \mathbf{T}_1 \boldsymbol{\pi}_{12} \qquad\qquad \boldsymbol{\pi}_2 = \mathbf{T}_2 \boldsymbol{\pi}_{12}, \tag{1}$$

illustrated here for $I = 2$ and $J = 3$

$$\mathbf{T}_1 = \begin{pmatrix} \mathbf{1}_3' & \mathbf{0}_3' \\ \mathbf{0}_3' & \mathbf{1}_3' \end{pmatrix} \qquad\qquad \mathbf{T}_2 = \begin{pmatrix} \mathbf{I}_3 & \mathbf{I}_3 \end{pmatrix}$$

where $\mathbf{1}_3$ and $\mathbf{0}_3$ denote three-dimensional column vectors of 1's and 0's respectively.

Now, assume the following model for π_{ij},

$$\pi_{ij} = \int_{\tau_{1_{i-1}}}^{\tau_{2_i}} \int_{\tau_{2_{j-1}}}^{\tau_{2_j}} \phi_2\left(\mathbf{z}_{12}^*\right) d\mathbf{z}_{12}^* \tag{2}$$

where $\tau_{1_0} = -\infty, \tau_{2_0} = -\infty, \tau_{1_I} = \infty, \tau_{2_J} = \infty$ and $\phi_n(\bullet)$ denotes a n-variate standard normal density function. Thus, $\mathbf{z}_{12}^*$ has mean zero and correlation matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. In particular,

$$\pi_i = \int_{\tau_{1_{i-1}}}^{\tau_{2_i}} \phi_1\left(z_1^*\right) dz_1^* \qquad\qquad \pi_j = \int_{\tau_{2_{j-1}}}^{\tau_{2_j}} \phi_1\left(z_2^*\right) dz_2^* \tag{3}$$

In the sequel, let $\boldsymbol{\tau} = \left(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2\right)'$, where $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ denote the $(I - 1)$ and $(J - 1)$ dimensional vectors of thresholds implied by the model. Finally, let $\boldsymbol{\kappa} = \left(\boldsymbol{\tau}, \rho\right)'$. We now provide the asymptotic distribution of the one and two-stage parameter estimates using standard results for maximum likelihood estimators for multinomial models. Agresti (1990) is a good source for the relevant theory.

Let $\boldsymbol{\pi}$ and $\mathbf{p}$ be C-dimensional vectors of multinomial probabilities and sample proportions, respectively, and let N denote sample size. Consider a parametric structure for $\boldsymbol{\pi}$, $\boldsymbol{\pi}(\boldsymbol{\vartheta})$, with Jacobian matrix $\boldsymbol{\Delta} = \dfrac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\vartheta}'}$, and suppose we estimate $\boldsymbol{\vartheta}$ by maximizing

$$L(\boldsymbol{\vartheta}) = N \sum_{c=1}^{C} p_c \ln \pi_c(\boldsymbol{\vartheta}). \tag{4}$$

Then, under typical regularity conditions, it follows that

$$\sqrt{N}(\mathbf{p} - \boldsymbol{\pi}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}) \qquad \boldsymbol{\Gamma} = \mathbf{D} - \boldsymbol{\pi}\boldsymbol{\pi}' \tag{5}$$

$$\sqrt{N}\left(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\right) \overset{a}{=} \mathbf{B}\sqrt{N}(\mathbf{p} - \boldsymbol{\pi}) \tag{6}$$

$$\sqrt{N}\left(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\right) \xrightarrow{d} N\left(\mathbf{0}, \left(\boldsymbol{\Delta}'\mathbf{D}^{-1}\boldsymbol{\Delta}\right)^{-1}\right) \tag{7}$$

where $\mathbf{D} = Diag(\boldsymbol{\pi})$, $\mathbf{B} = \left(\boldsymbol{\Delta}'\mathbf{D}^{-1}\boldsymbol{\Delta}\right)^{-1} \boldsymbol{\Delta}'\mathbf{D}^{-1}$, $\xrightarrow{d}$ denotes convergence in distribution, and $\overset{a}{=}$ denotes asymptotic equality.

2.1 One-stage estimation

Akin to (5) we write, $\sqrt{N}\left(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}\right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma})$, where $\boldsymbol{\Gamma} = \mathbf{D}_{12} - \boldsymbol{\pi}_{12}\boldsymbol{\pi}_{12}'$ and $\mathbf{D}_{12} = Diag\left(\boldsymbol{\pi}_{12}\right)$. Then, when all the parameters are estimated simultaneously by maximizing $L_{12}\left(\boldsymbol{\tau}_i, \boldsymbol{\tau}_j, \rho\right) = N \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \ln \pi_{ij}\left(\boldsymbol{\tau}_i, \boldsymbol{\tau}_j, \rho\right)$, it follows from (7)

$$\sqrt{N}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}) \xrightarrow{d} N\left(\mathbf{0}, \left(\boldsymbol{\Delta}'\mathbf{D}_{12}^{-1}\boldsymbol{\Delta}\right)^{-1}\right) \tag{8}$$

where $\boldsymbol{\Delta} = \dfrac{\partial \boldsymbol{\pi}_{12}}{\partial \boldsymbol{\kappa}'} = \left(\dfrac{\partial \boldsymbol{\pi}_{12}}{\partial \boldsymbol{\tau}'} \quad \dfrac{\partial \boldsymbol{\pi}_{12}}{\partial \rho}\right)$ and all necessary derivatives can be found in Olsson (1979).

2.2 Two-stage estimation

Consider now the following sequential estimator for $\boldsymbol{\kappa}$ (Olsson, 1979):

First stage: Estimate the thresholds for each variable separately by maximizing

$$L_1(\boldsymbol{\tau}_1) = N\sum_{i=1}^{I} p_i \ln \pi_i(\boldsymbol{\tau}_i) \qquad\qquad L_2(\boldsymbol{\tau}_2) = N\sum_{j=1}^{J} p_j \ln \pi_j(\boldsymbol{\tau}_j) \tag{9}$$

Second stage: Estimate the polychoric correlation by maximizing

$$L_{12}(\rho, \hat{\boldsymbol{\tau}}) = N\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} \ln \pi_{ij}(\rho, \hat{\boldsymbol{\tau}}) \tag{10}$$

We shall now provide an alternative derivation of Olsson's results for this estimator closely following Jöreskog's (1994). We first notice that $\hat{\boldsymbol{\tau}}_i$ and $\hat{\boldsymbol{\tau}}_j$ are maximum likelihood estimates, as (9) is the kernel of the log-likelihood function for estimating the thresholds from the univariate marginals of the contingency table. Similarly, (10) is the kernel of the log-likelihood function for estimating the polychoric correlation from the bivariate contingency table given the estimated thresholds. That is, $\hat{\rho}$ is a pseudo-maximum likelihood estimate in the terminology of Gong and Samaniego (1981).

To obtain the asymptotic distribution of the two-stage estimates we first apply (6) to the first stage estimates to obtain

$$\sqrt{N}(\hat{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_1) \overset{a}{=} \mathbf{B}_{11}\sqrt{N}(\mathbf{p}_1 - \boldsymbol{\pi}_1) \qquad\qquad \sqrt{N}(\hat{\boldsymbol{\tau}}_2 - \boldsymbol{\tau}_2) \overset{a}{=} \mathbf{B}_{12}\sqrt{N}(\mathbf{p}_2 - \boldsymbol{\pi}_2) \tag{11}$$

where $\mathbf{B}_{11} = (\boldsymbol{\Delta}_{11}'\mathbf{D}_1^{-1}\boldsymbol{\Delta}_{11})^{-1}\boldsymbol{\Delta}_{11}'\mathbf{D}_1^{-1}$, $\mathbf{D}_1 = Diag(\boldsymbol{\pi}_1)$, $\boldsymbol{\Delta}_{11} = \dfrac{\partial\boldsymbol{\pi}_1}{\partial\boldsymbol{\tau}'}$, and so on. These derivatives can also be found in Olsson (1979). Now, letting $\mathbf{B}_1 = \begin{pmatrix}\mathbf{B}_{11}\mathbf{T}_1\\ \mathbf{B}_{12}\mathbf{T}_2\end{pmatrix}$, by (11) and (1) we get

$$\sqrt{N}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \overset{a}{=} \mathbf{B}_1\sqrt{N}(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}). \tag{12}$$

Similarly, a direct application of (6) to the second stage estimates yields

$$\sqrt{N}(\hat{\rho} - \rho) \overset{a}{=} \boldsymbol{B}_{22}\sqrt{N}(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}(\rho, \hat{\boldsymbol{\tau}})) \tag{13}$$

where $\mathbf{B}_{22} = (\boldsymbol{\Delta}_{22}'\mathbf{D}_{12}^{-1}\boldsymbol{\Delta}_{22})^{-1}\boldsymbol{\Delta}_{22}'\mathbf{D}_{12}^{-1}$, and $\boldsymbol{\Delta}_{22} = \dfrac{\partial\boldsymbol{\pi}_{12}}{\partial\rho}$. Note that $\mathbf{B}_{22}$ and $\boldsymbol{\Delta}_{22}$ are a row and a column vector, respectively, despite the notation. Now, we need the asymptotic distribution of $\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}(\rho,\hat{\boldsymbol{\tau}})\right)$ to proceed. In Appendix 1, we show that

$$\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}(\rho,\hat{\boldsymbol{\tau}})\right)\stackrel{a}{=}\left(\mathbf{I}-\boldsymbol{\Delta}_{21}\mathbf{B}_{1}\right)\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}\right) \qquad (14)$$

where $\boldsymbol{\Delta}_{21}=\frac{\partial\boldsymbol{\pi}_{12}}{\partial\boldsymbol{\tau}'}$. Putting together (13) and (14) we obtain

$$\sqrt{N}\left(\hat{\rho}-\rho\right)\stackrel{a}{=}\mathbf{B}_{22}\left(\mathbf{I}-\boldsymbol{\Delta}_{21}\mathbf{B}_{1}\right)\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}\right). \qquad (15)$$

Finally, putting together (12) and (15) we obtain

$$\sqrt{N}\left(\hat{\boldsymbol{\kappa}}-\boldsymbol{\kappa}\right)\stackrel{a}{=}\mathbf{G}\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}\right) \qquad \mathbf{G}=\begin{pmatrix}\mathbf{B}_{1}\\ \mathbf{B}_{22}\left(\mathbf{I}-\boldsymbol{\Delta}_{21}\mathbf{B}_{1}\right)\end{pmatrix}. \qquad (16)$$

and since as shown in the Appendix,

$$\mathbf{G}\boldsymbol{\pi}_{12}=\mathbf{0}, \qquad (17)$$

$$\sqrt{N}\left(\hat{\boldsymbol{\kappa}}-\boldsymbol{\kappa}\right)\stackrel{d}{\to}N\left(\mathbf{0},\mathbf{G}\mathbf{D}_{12}\mathbf{G}'\right) \qquad (18)$$

where $\mathbf{G}$ and $\mathbf{D}_{12}$ are to be evaluated at the true population values.

3. Goodness of fit testing

We wish to investigate the asymptotic distribution of Pearson's X^2 statistic,

$X^{2}=N\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(p_{ij}-\pi_{ij}(\hat{\boldsymbol{\kappa}})\right)^{2}}{\pi_{ij}(\hat{\boldsymbol{\kappa}})}$ and the likelihood ratio statistic

$G^{2}=N\sum_{i=1}^{I}\sum_{j=1}^{J}p_{ij}\ln\frac{p_{ij}}{\pi_{ij}(\hat{\boldsymbol{\kappa}})}=2\hat{L}_{12}$, where by convention when $p_{ij}=0$, $p_{ij}\ln\frac{p_{ij}}{\pi_{ij}(\hat{\boldsymbol{\kappa}})}=0$.

From standard theory (e.g., Agresti, 1990), when the model parameters are estimated simultaneously, $G^{2}\stackrel{a}{=}X^{2}\stackrel{d}{\to}\chi^{2}_{IJ-I-J}$. When they are estimated in two stages, it also holds that $G^{2}\stackrel{a}{=}X^{2}$ (see Agresti, 1990: p. 434). Thus, we only consider the asymptotic distribution of

X^2. To do so, we first consider the asymptotic distribution of the unstandardized residuals $\sqrt{N}\hat{\mathbf{e}} := \sqrt{N}\left(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}(\hat{\boldsymbol{\kappa}})\right)$ when $\hat{\boldsymbol{\kappa}}$ are two- stage parameter estimates. In the Appendix it is shown that

$$\sqrt{N}\hat{\mathbf{e}} \overset{a}{=} (\mathbf{I} - \mathbf{K})\sqrt{N}\left(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}\right) \tag{19}$$

where $\boldsymbol{\Delta} = \dfrac{\partial \boldsymbol{\pi}_{12}}{\partial \boldsymbol{\kappa}'} = \begin{pmatrix}\boldsymbol{\Delta}_{21} & \boldsymbol{\Delta}_{22}\end{pmatrix}$ and $\mathbf{K} = \boldsymbol{\Delta}\mathbf{G}$. Thus, by (19) and (17),

$$\sqrt{N}\hat{\mathbf{e}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}) \qquad \boldsymbol{\Omega} = (\mathbf{I} - \mathbf{K})\mathbf{D}_{12}(\mathbf{I} - \mathbf{K})' - \boldsymbol{\pi}_{12}\boldsymbol{\pi}_{12}'. \tag{20}$$

Now, we note that X^2 can be written as $X^2 = N\hat{\mathbf{e}}'\hat{\mathbf{D}}_{12}^{-1}\hat{\mathbf{e}}$. Using standard results (Agresti, 1990: p. 432),

$$X^2 = N\hat{\mathbf{e}}'\hat{\mathbf{D}}_{12}^{-1}\hat{\mathbf{e}} \overset{a}{=} N\hat{\mathbf{e}}'\mathbf{D}_{12}^{-1}\hat{\mathbf{e}}. \tag{21}$$

A necessary and sufficient condition for (21) to be asymptotically chi-square distribution is (e.g., Schott, 1997: Theorem 9.10)

$$\boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\boldsymbol{\Omega} = \boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\boldsymbol{\Omega} \tag{22}$$

In the Appendix we show that

$$\left(\boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\right)^2 = (\mathbf{I} - \mathbf{K})(\mathbf{I} - \mathbf{JK})(\mathbf{I} - \mathbf{J}) - \mathbf{C} \neq \boldsymbol{\Omega}\mathbf{D}_{12}^{-1} = (\mathbf{I} - \mathbf{K})(\mathbf{I} - \mathbf{J}) - \mathbf{C} \tag{23}$$

where $\mathbf{J} = \mathbf{D}_{12}\mathbf{K}'\mathbf{D}_{12}^{-1}$ and $\mathbf{C} = \boldsymbol{\pi}_{12}\boldsymbol{\pi}_{12}'\mathbf{D}_{12}^{-1}$. Thus, (22) is not satisfied. Neither X^2 nor G^2 are asymptotically chi-squared. Rather, these statistics converge in distribution to a mixture of $r = IJ - I - J$ independent chi-square variables with one degree of freedom (Box, 1954: Theorem 2.1).

Nevertheless, following Moore (1977) it is easy to construct an alternative quadratic form in the residuals (20) that is asymptotically chi-square. Consider $\hat{\boldsymbol{\Omega}}^-$, a matrix that converges in probability to $\boldsymbol{\Omega}^-$, a generalized inverse of $\boldsymbol{\Omega}$ satisfying $\boldsymbol{\Omega}\boldsymbol{\Omega}^-\boldsymbol{\Omega} = \boldsymbol{\Omega}$. Then, it readily follows that

$$W = N\mathbf{e}'\hat{\boldsymbol{\Omega}}^{-}\mathbf{e} \xrightarrow{d} \chi^2_r \qquad (24)$$

where W is a Wald statistic. To consistently estimate $\boldsymbol{\Omega}$ we evaluate all derivative matrices in $\boldsymbol{\Delta}$ and $\mathbf{G}$ at the two-stage estimates and evaluate all probabilities at the estimated values. Then, $\hat{\boldsymbol{\Omega}}^{-}$ can be readily obtained from $\hat{\boldsymbol{\Omega}}$ by numerical methods.

In closing this section, we note that with one-stage parameter estimates $X^2 = W$ as $\mathbf{D}_{12}^{-1}$ is a generalized inverse of the asymptotic covariance matrix of the unstandardized residuals $\sqrt{N}\hat{\mathbf{e}} := \sqrt{N}\left(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}(\hat{\boldsymbol{\kappa}})\right)$.

Next, we present a numerical example and a very small simulation study to illustrate the small sample performance of X^2, G^2, and W.

4. Numerical Results

4.1 Soft drink data

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*", "Indiference", "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, the asymptotic standard errors of these parameters, and the goodness of fit results. The standard errors for the parameter estimates were obtained as the square root of the diagonal of (18) which was consistently estimated by evaluating all derivative matrices and probabilities at the estimated parameter values.

Insert Tables 1 and 2 about here

As can be seen in this table for all three bivariate tables the test statistics suggest that the assumption of categorized bivariate normality is reasonable. We also observe that the W and X^2 values obtained are extremely close. It is our experience that this is always the case. Thus, although X^2 is not asymptotically chi-squared when the model parameters are estimated in two stages, its finite sample distribution is so close to that of the asymptotically correct W statistic that it may be used in its place. We believe this is because the expected probabilities obtained using the two-stage estimates are very close to those obtained when all model parameters are jointly estimated.

To illustrate this point, in Table 2 we provide the one-stage estimates for these three bivariate tables, along with goodness of fit statistics. We note that when estimating the parameters jointly, the thresholds estimated from different bivariate tables need not be the same across tables. We see in this table that the parameter estimates and their standard errors are very similar to those obtained using the two stage approach. Furthermore, the X^2 values are also essentially identical to those obtained using the two stage approach.

We also note in Tables 1 and 2 that the estimated G^2 statistics are larger than the W and X^2 statistics. This is because we purposely chose a numerical example with a very small sample size relative to the size of the table 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 and these are not included in the computation of G^2. To investigate more systematically this effect we performed a simulation study using $\boldsymbol{\tau}_1 = (-1, -0.5, 0.5, 1)'$, $\boldsymbol{\tau}_2 = (-1, -0.5, 0.5, 1)'$ and $\rho = 0.3$. The results for $N = 50$, $N = 100$, and $N = 1000$ across 1000 replications are presented in Table 3.

Insert Table 3 about here

As can be seen in this table, G^2 yields conservative results when $N = 50$ and $N = 100$, although its behavior in the critical region {1% to 5%} is acceptable when $N = 100$. We also see that the differences between the W and X^2 statistics are negligible, so that the computationally more complex W statistic is not really needed. The X^2 statistic has an acceptable behavior even when $N = 50$ in 5×5 contingency tables. This is remarkable.

5. Discussion

The purpose of this research was to investigate whether it was theoretically justified to use X^2 and G^2 to test categorized bivariate normality when the model parameters are estimated in two stages. We have found that it is not theoretically justified. Consequently, we have proposed an alternative test, W, that is asymptotically chi-squared when two-stage estimation is employed.

However, our experience -illustrated with our small numerical example- and our very small simulation study suggest that there are very small numerical differences when X^2 and G^2 are computed using one and two-stage estimates. In the former case it is theoretically justified their use. Thus, in practice it seems justified to use X^2 and G^2 with two stage estimates. It is not the purpose of the present research to investigate the relative performance of X^2 and G^2 across a variety of sample sizes and true parameter values. However, our limited simulation results suggest that when there are empty cells G^2 is too conservative and that X^2 is to be preferred to G^2. We have also found that W does not outperform X^2 in practice, where the latter is to be preferred on computational grounds.

References

Agresti, A. (1990). Categorical data analysis. New York: Wiley.

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

Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. Annals of Mathematical Statistics, 16, 769-771.

Gong, G., & Samaniego, F.J. (1981). Pseudo maximum likelihood estimation: Theory and applications. Annals of Statistics, 9, 861-869.

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

Jöreskog, K.G. & Sörbom, D. (1993). LISREL 8. User's reference guide. Chicago, IL: Scientific Software.

Moore, D.S. (1977). Generalized inverses, Wald's method, and the construction of chi-squared tests of fit. Journal of the American Statistical Association, 72, 131-137.

Muthén, B. (1987). LISCOMP: Analysis of linear structural equations using a comprehensive measurement model. Mooresville, IN: Scientific Software.

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, L. & Muthén, B. (1998). Mplus. Los Angeles, CA: Muthén & Muthén.

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

Schott, J.R. (1997). Matrix analysis for statistics. New York: Wiley.

TABLE 1

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

Thresholds

	1	2	3	4
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)

Correlations and Test Statistics

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

Notes: $N = 61$; standard errors in parentheses; 15 d.f.

TABLE 2

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

Thresholds

	1	2	3	4
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)

Correlations and Test Statistics

Vars.	Corr.	G^2	p-value	X^2	p-value
(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

Notes: $N = 61$; standard errors in parentheses; 15 d.f.

TABLE 3

Simulation results

				Nominal rates										
N	*Stat.*	*Mean*	*Var.*	*1%*	*5%*	*10%*	*20%*	*30%*	*40%*	*50%*	*60%*	*70%*	*80%*	*90%*
50	W	15.33	27.37	1.2	4.6	9.0	18.8	31.53	42.5	52.3	62.4	73.1	84.9	94.6
	G^2	17.55	29.27	1.8	9.4	17.6	34.5	47.8	59.3	68.9	78.6	87.6	94.1	97.8
	X^2	15.37	27.64	1.4	4.6	9.2	19.0	32.0	43.0	52.6	62.8	73.3	84.8	94.6
100	W	14.90	24.59	0.2	3.1	8.6	18.4	28.6	38.8	51.5	63.2	71.7	81.6	91.9
	G^2	16.75	30.21	1.2	7.8	14.9	28.9	43.3	56.0	65.6	72.2	80.8	89.1	94.5
	X^2	14.93	24.69	0.2	3.2	8.7	18.5	28.8	38.9	52.0	63.4	71.9	81.7	91.9
1000	W	14.65	28.32	0.8	4.3	9.1	18.0	27.4	37.5	49.2	56.9	67.2	79.4	88.6
	G^2	14.81	29.4	1.0	4.8	9.7	18.6	28.7	38.9	49.6	57.3	68.2	79.8	89.2
	X^2	14.67	28.42	0.8	4.5	9.2	18.0	27.6	37.7	48.3	56.9	67.3	79.5	88.8

Notes: 1000 replications; 15 d.f.; $\boldsymbol{\tau}_1 = (-1, -0.5, 0.5, 1)'$, $\boldsymbol{\tau}_2 = (-1, -0.5, 0.5, 1)'$, $\rho = 0.3$.

Appendix: Proofs of key results

Proof of Equation (14):

A first order Taylor expansion of $\boldsymbol{\pi}_{12}(\rho,\hat{\boldsymbol{\tau}})$ around $\boldsymbol{\tau}=\boldsymbol{\tau}_0$ yields

$$\boldsymbol{\pi}_{12}(\rho,\hat{\boldsymbol{\tau}}) \overset{a}{=} \boldsymbol{\pi}_{12}(\rho,\boldsymbol{\tau}) + \boldsymbol{\Delta}_{21}(\hat{\boldsymbol{\tau}}-\boldsymbol{\tau}),$$

where $\boldsymbol{\Delta}_{21}=\dfrac{\partial\boldsymbol{\pi}_{12}}{\partial\boldsymbol{\tau}'}$. Thus, $\sqrt{N}\left(\boldsymbol{\pi}_{12}(\rho,\hat{\boldsymbol{\tau}})-\boldsymbol{\pi}_{12}\right)\overset{a}{=}\boldsymbol{\Delta}_{21}(\hat{\boldsymbol{\tau}}-\boldsymbol{\tau})\overset{a}{=}\boldsymbol{\Delta}_{21}\mathbf{B}_1\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}\right)$, where the last asymptotic equality follows from (12). Now,

$$\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}(\rho,\hat{\boldsymbol{\tau}})\right)=\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}\right)-\sqrt{N}\left(\boldsymbol{\pi}_{12}(\rho,\hat{\boldsymbol{\tau}})-\boldsymbol{\pi}_{12}\right)\overset{a}{=}\left(\mathbf{I}-\boldsymbol{\Delta}_{21}\mathbf{B}_1\right)\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}\right)$$

Proof of Equation (17):

$\mathbf{B}_{11}\mathbf{T}_1\boldsymbol{\pi}_{12}=\mathbf{0}$ because $\boldsymbol{\Delta}_{11}'\mathbf{D}_1^{-1}\mathbf{T}_1\boldsymbol{\pi}_{12}=\boldsymbol{\Delta}_{11}'\mathbf{D}_1^{-1}\boldsymbol{\pi}_1=\mathbf{0}$. $\mathbf{B}_{12}\mathbf{T}_2\boldsymbol{\pi}_{12}=\mathbf{0}$ because $\boldsymbol{\Delta}_{12}'\mathbf{D}_2^{-1}\mathbf{T}_2\boldsymbol{\pi}_{12}=\boldsymbol{\Delta}_{12}'\mathbf{D}_2^{-1}\boldsymbol{\pi}_2=\mathbf{0}$. Thus, $\mathbf{B}_1\boldsymbol{\pi}_{12}=\mathbf{0}$.

Also, $\mathbf{B}_{22}\boldsymbol{\pi}_{12}=\mathbf{0}$ because $\boldsymbol{\Delta}_{22}'\mathbf{D}_{12}^{-1}\boldsymbol{\pi}_{12}=\mathbf{0}$, so the proof is complete

Proof of Equation (19):

A first order Taylor expansion of $\boldsymbol{\pi}_{12}(\hat{\boldsymbol{\kappa}})$ around $\boldsymbol{\kappa}=\boldsymbol{\kappa}_0$ yields

$$\boldsymbol{\pi}_{12}(\hat{\boldsymbol{\kappa}})\overset{a}{=}\boldsymbol{\pi}_{12}(\boldsymbol{\kappa})+\boldsymbol{\Delta}(\hat{\boldsymbol{\kappa}}-\boldsymbol{\kappa}),$$

where $\boldsymbol{\Delta}=\dfrac{\partial\boldsymbol{\pi}_{12}}{\partial\boldsymbol{\kappa}'}$. Thus, $\sqrt{N}\left(\hat{\boldsymbol{\pi}}_{12}-\boldsymbol{\pi}_{12}\right)\overset{a}{=}\boldsymbol{\Delta}(\hat{\boldsymbol{\kappa}}-\boldsymbol{\kappa})\overset{a}{=}\boldsymbol{\Delta}\mathbf{G}\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}\right)$, where the last asymptotic equality follows from (16). Now,

$$\sqrt{N}\left(\mathbf{p}_{12}-\hat{\boldsymbol{\pi}}_{12}\right)=\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}\right)-\sqrt{N}\left(\hat{\boldsymbol{\pi}}_{12}-\boldsymbol{\pi}_{12}\right)\overset{a}{=}(\mathbf{I}-\boldsymbol{\Delta}\mathbf{G})\sqrt{N}\left(\mathbf{p}_{12}-\boldsymbol{\pi}_{12}\right)$$

Proof of Equation (23):

We write $\boldsymbol{\Omega}\mathbf{D}_{12}^{-1} = (\mathbf{I}-\mathbf{K})(\mathbf{I}-\mathbf{J}) - \mathbf{C} = \mathbf{A}-\mathbf{C}$ where $\mathbf{K} = \boldsymbol{\Delta}\mathbf{G}$, $\mathbf{J} = \mathbf{D}_{12}\mathbf{G}'\boldsymbol{\Delta}'\mathbf{D}_{12}^{-1}$ and $\mathbf{C} = \boldsymbol{\pi}_{12}\boldsymbol{\pi}_{12}'\mathbf{D}_{12}^{-1}$. Then, $\left(\boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\right)^2 = \mathbf{A}^2 - \mathbf{AC} - \mathbf{CA} + \mathbf{C}^2$. Using the standard equalities,

$$\mathbf{D}_{12}^{-1}\boldsymbol{\pi}_{12} = \mathbf{1} \qquad \boldsymbol{\Delta}'\mathbf{1} = \mathbf{0} \qquad \boldsymbol{\pi}_{12}'\mathbf{D}_{12}^{-1} = \mathbf{1}' \qquad \mathbf{1}'\boldsymbol{\Delta} = \mathbf{0}' \tag{25}$$

and (17) we first notice that

$$\mathbf{JC} = \mathbf{CJ} = \mathbf{0} \qquad \mathbf{KC} = \mathbf{CK} = \mathbf{0} \tag{26}$$

Thus, $\mathbf{AC} = \mathbf{C}$ and $\mathbf{CA} = \mathbf{C}$. Furthermore, using (25) and $\mathbf{1}'\boldsymbol{\pi}_{12} = 1$ (a scalar), $\mathbf{C}^2 = \mathbf{C}$.

Thus, $\left(\boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\right)^2 = \mathbf{A}^2 - \mathbf{C}$, but noting that

$$\mathbf{J}^2 = \mathbf{J} \qquad \mathbf{K}^2 = \mathbf{K} \tag{27}$$

we find that $\mathbf{A}^2 = \left((\mathbf{I}-\mathbf{K})(\mathbf{I}-\mathbf{J})\right)^2 = (\mathbf{I}-\mathbf{K})(\mathbf{I}+\mathbf{JK})(\mathbf{I}-\mathbf{J})$. Therefore, $\left(\boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\right)^2 \neq \boldsymbol{\Omega}\mathbf{D}_{12}^{-1}$.