TESTING CATEGORIZED BIVARIATE NORMALITY WITH TWO-STAGE POLYCHORIC CORRELATION ESTIMATES

IE Working Paper **MK8-106-I** **08 / 05 / 2003**

Alberto Maydeu Olivares

Instituto de Empresa
Marketing Dept.
C / María de Molina 11-15,
28006 – Madrid
España
Alberto.Maydeu@ie.edu

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 test statistic, M_n, that is asymptotically chi-square in this situation. M_n, may have a wide range of applications beyond the one considered here as it is asymptotically chi-square for a broad class of consistent and asymptotically normal estimators. M_n equals X^2 with an adjustment to take into account that the estimator is not asymptotically efficient. Also, $M_n \leq X^2$ where $M_n = X^2$ in the case of the one-stage maximum likelihood estimator.

Keywords
LISREL, MPLUS, Categorical data analysis, pseudo-maximum likelihood estimation, multinomial models

* This research was supported by grant BSO2000-0661 of the Spanish Ministry of Science and Technology, by a Distinguished Young Investigator Award of the Catalan Government, and by a grant from the Fundación Instituto de Empresa. I am indebted to Karl Jöreskog and Harry Joe for their comments to a previous draft.

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 the likelihood ratio statistic G^2 or Pearson's X^2 test statistic. 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, the distribution of G^2 and X^2 when the two-stage approach is employed remains to be investigated. Yet, 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 PRELIS/LISREL (Jöreskog & Sörbom, 1993) and MPLUS (Muthén & Muthén, 1998). To assess the goodness of fit of the moel, G^2 is used in PRELIS/LISREL (Jöreskog, 2001, July 26, personal communication). No goodness of fit test is currently implemented in MPLUS. The purpose of this paper is to investigate the asymptotic distribution of G^2 and X^2 when the two-stage estimator is employed.

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,

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

for certain (implicitly defined) matrices $\mathbf{T}_1$ and $\mathbf{T}_2$. For instance, for $I = 2$ and $J = 3$,

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

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,

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

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

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

Then, under typical regularity conditions, it follows that

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

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

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

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

2.1 One-stage estimation

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

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

where $\boldsymbol{\Delta}=\dfrac{\partial\boldsymbol{\pi}_{12}}{\partial\boldsymbol{\kappa}'}=\begin{pmatrix}\dfrac{\partial\boldsymbol{\pi}_{12}}{\partial\boldsymbol{\tau}'} & \dfrac{\partial\boldsymbol{\pi}_{12}}{\partial\rho}\end{pmatrix}$ 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

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

Second stage: Estimate the polychoric correlation by maximizing

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

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

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

where $\mathbf{B}_{11} = \left(\boldsymbol{\Delta}_{11}'\mathbf{D}_1^{-1}\boldsymbol{\Delta}_{11}\right)^{-1}\boldsymbol{\Delta}_{11}'\mathbf{D}_1^{-1}$, $\mathbf{D}_1 = \mathrm{Diag}\left(\boldsymbol{\pi}_1\right)$, $\boldsymbol{\Delta}_{11} = \dfrac{\partial \boldsymbol{\pi}_1}{\partial \boldsymbol{\tau}'}$, 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

$$\sqrt{N}\left(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}\right) \stackrel{a}{=} \mathbf{B}_1\sqrt{N}\left(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}\right). \quad (12)$$

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

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

where $\mathbf{B}_{22} = \left(\boldsymbol{\Delta}_{22}'\mathbf{D}_{12}^{-1}\boldsymbol{\Delta}_{22}\right)^{-1}\boldsymbol{\Delta}_{22}'\mathbf{D}_{12}^{-1}$, 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 $\sqrt{N}\left(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}\left(\rho, \hat{\boldsymbol{\tau}}\right)\right)$ to proceed. In Appendix 1, we show that

$$\sqrt{N}\left(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}\left(\rho, \hat{\boldsymbol{\tau}}\right)\right) \stackrel{a}{=} \left(\mathbf{I} - \boldsymbol{\Delta}_{21}\mathbf{B}_1\right)\sqrt{N}\left(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}\right) \quad (14)$$

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

$$\sqrt{N}\left(\hat{\rho} - \rho\right) \stackrel{a}{=} \mathbf{B}_{22}\left(\mathbf{I} - \boldsymbol{\Delta}_{21}\mathbf{B}_1\right)\sqrt{N}\left(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}\right). \quad (15)$$

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

$$\sqrt{N}\left(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}\right) \stackrel{a}{=} \mathbf{G}\sqrt{N}\left(\mathbf{p}_{12} - \boldsymbol{\pi}_{12}\right) \qquad \mathbf{G} = \begin{pmatrix}\mathbf{B}_1 \\ \mathbf{B}_{22}\left(\mathbf{I} - \boldsymbol{\Delta}_{21}\mathbf{B}_1\right)\end{pmatrix}. \quad (16)$$

and since as shown in the Appendix,

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

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

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

3. Goodness of fit testing

We shall first obtain 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

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

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

$$\sqrt{N}\hat{\mathbf{e}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}) \qquad \boldsymbol{\Omega} = (\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})\boldsymbol{\Gamma}(\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})'. \tag{20}$$

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

$$X^2 = N\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(p_{ij} - \pi_{ij}(\hat{\boldsymbol{\kappa}})\right)^2}{\pi_{ij}(\hat{\boldsymbol{\kappa}})} = N\hat{\mathbf{e}}'\hat{\mathbf{D}}_{12}^{-1}\hat{\mathbf{e}} \tag{21}$$

$$G^2 = N\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij}\ln\frac{p_{ij}}{\pi_{ij}(\hat{\boldsymbol{\kappa}})}, \tag{22}$$

where $\hat{\mathbf{D}}_{12} = \mathrm{Diag}\left(\boldsymbol{\pi}_{12}(\hat{\boldsymbol{\kappa}})\right)$, and by convention when $p_{ij} = 0$, $p_{ij}\ln\dfrac{p_{ij}}{\pi_{ij}(\hat{\boldsymbol{\kappa}})} = 0$.

From standard theory (e.g., Agresti, 1990), when the model parameters are estimated simultaneously, $G^2 \stackrel{a}{=} X^2 \xrightarrow{d} \chi^2_{IJ-I-J}$. When they are estimated in two stages, it also holds that $G^2 \stackrel{a}{=} X^2$ (see Agresti, 1990: p. 434). Therefore, we only consider here the asymptotic distribution of X^2. Now, again using standard results (Agresti, 1990: p. 432), $X^2 = N\hat{\mathbf{e}}'\hat{\mathbf{D}}_{12}^{-1}\hat{\mathbf{e}} \stackrel{a}{=} N\hat{\mathbf{e}}'\mathbf{D}_{12}^{-1}\hat{\mathbf{e}}$. A necessary and sufficient condition for X^2 to be asymptotically chi-square distributed is (e.g., Schott, 1997: Theorem 9.10)

$$\boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\boldsymbol{\Omega} = \boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\boldsymbol{\Omega} \tag{23}$$

In the Appendix we show that when the two-stage estimator is employed

$$\left(\boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\right)^2 = (\mathbf{I} - \mathbf{K})(\mathbf{I} - \mathbf{J}\mathbf{K})(\mathbf{I} - \mathbf{J}) - \mathbf{C} \neq \boldsymbol{\Omega}\mathbf{D}_{12}^{-1} = (\mathbf{I} - \mathbf{K})(\mathbf{I} - \mathbf{J}) - \mathbf{C} \tag{24}$$

where $\mathbf{K} = \boldsymbol{\Delta}\mathbf{G}$, $\mathbf{J} = \mathbf{D}_{12}\mathbf{K}'\mathbf{D}_{12}^{-1}$ and $\mathbf{C} = \boldsymbol{\pi}_{12}\boldsymbol{\pi}_{12}{}'\mathbf{D}_{12}^{-1}$. Thus, (23) 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). Thus, an alternative test statistic must be sought that it is asymptotically chi-square under more general conditions than X^2 and G^2 to test the goodness of fit of the model when the two-stage estimator is employed.

Let $\tilde{\boldsymbol{\vartheta}}$ be a consistent estimator satisfying

$$\sqrt{N}\left(\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\right) \overset{a}{=} \mathbf{G}\sqrt{N}\left(\mathbf{p} - \boldsymbol{\pi}\right) \tag{25}$$

for some $q \times C$ matrix $\mathbf{G}$ satisfying

$$\mathbf{G}\boldsymbol{\Delta} = \mathbf{I}. \tag{26}$$

Now, let $\tilde{\mathbf{e}} = \left(\mathbf{p} - \boldsymbol{\pi}\left(\tilde{\boldsymbol{\vartheta}}\right)\right)$ and consider the test statistic

$$M_n = N\tilde{\mathbf{e}}'\tilde{\mathbf{U}}\tilde{\mathbf{e}} \qquad \mathbf{U} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\boldsymbol{\Delta}\left(\boldsymbol{\Delta}'\mathbf{D}^{-1}\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}'\mathbf{D}^{-1} \tag{27}$$

where $\tilde{\mathbf{U}}$ denotes $\mathbf{U}$ evaluated at $\tilde{\boldsymbol{\vartheta}}$. We show in the Appendix that under these conditions

$$M_n \xrightarrow{d} \chi^2_{C-q-1}. \tag{28}$$

We note that M_n can be written as

$$M_n = X^2 - N\tilde{\mathbf{e}}'\tilde{\mathbf{D}}^{-1}\tilde{\boldsymbol{\Delta}}\left(\tilde{\boldsymbol{\Delta}}'\tilde{\mathbf{D}}^{-1}\hat{\boldsymbol{\Delta}}\right)^{-1}\hat{\boldsymbol{\Delta}}'\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{e}}. \tag{29}$$

Thus, $M_n \leq X^2$. M_n equals X^2 with an adjustment to take into account that $\tilde{\boldsymbol{\vartheta}}$ is not asymptotically efficient. Note that the second term in (29) becomes zero when $\hat{\mathbf{e}}'\hat{\mathbf{D}}_{12}^{-1}\hat{\boldsymbol{\Delta}} = \mathbf{0}'$. Since this is the gradient vector that maximum likelihood estimates satisfy, in the case of one-stage maximum likelihood estimation $M_n = X^2$.

The two-stage estimator under consideration is consistent and with $\mathbf{G}$ given by (16) it satisfies (26) (see Appendix). Thus, with two-stage parameter estimates $M_n \xrightarrow{d} \chi^2_{IJ-I-J}$.

4. Numerical results

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 and the asymptotic standard errors of these parameters. The standard errors 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

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 and these 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 as reflected in (29), the values of M_n and X^2 will be very close if the estimator used is highly efficient, yet not fully efficient. With these data, 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. As a result, in this example the G^2 and X^2 values obtained using the one and two-stage estimates are very close.

To further illustrate the high asymptotic relative efficiency of the two stage estimator we computed the population asymptotic covariance matrix of the one-stage and two-stage estimators using (8) and (18) at population values similar to those encountered in the example: $\boldsymbol{\tau}_1 = (-1,-0.5,0.5,1)'$, $\boldsymbol{\tau}_2 = (-1,-0.5,0.5,1)'$ and $\rho = 0.3$. At these values, the determinant of the asymptotic covariance matrix of the two-stage estimates (18) 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.

To investigate the small sample performance of G^2, X^2 and M_n we performed a simulation study using the above population values. 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, in the critical region {1% to 10%} G^2 tends to reject too often the null hypotheses when $N = 50$ and $N = 100$. The behavior of X^2 and M_n is acceptable even when $N = 50$ in these 5×5 contingency tables. This is remarkable. Also, we 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, in accordance to (29). 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.

5. Discussion

The purpose of this research was to investigate whether it was theoretically justified the present use of G^2 to test categorized bivariate normality when the model parameters are estimated in two stages. We have shown that with two-stage parameter estimates G^2 is not asymptotically chi-square distributed, and neither is X^2. With two-stage parameter estimates, G^2 and X^2 are asymptotically equivalent, and they are distributed as a mixture of one-degree of freedom chi-squares.

We have proposed a new test, M_n, that is asymptotically distributed as a chi-square with two-stage parameter estimates. 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). Our numerical results suggest that G^2 yields on average smaller p-values than M_n, particularly in small samples. We have also shown that M_n reduces algebraically to X^2 in the case of one-stage parameter estimates. The more efficient the two stage estimates, the closer M_n will be to X^2. At the population values used in our simulation study, the two-stage estimates are so efficient that the empirical distributions of M_n and X^2 are very similar. However, at parameter values where the two-stage estimates are not so efficient, that is, with large polychoric correlations, M_n is not so close to X^2. Given the straightforward computation of M_n, we recommend that this asymptotically correct statistic be used to assess the goodness of fit of the model when two-stage parameter estimation is employed.

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.

Khatri, (1966). A note on a Manova model applied to growth curves. Annals of the Institute of Mathematical Statistics, 18, 75-86.

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.	M_n	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%*
	M_n	15.42	26.75	1.1	4.0	9.2	20.5	32.1	43.2	54.4	64.1	74.7	84.6	93.8
50	G^2	17.67	28.92	1.8	9.3	18.2	35.9	49.2	60.8	71.5	80.3	87.2	93.8	97.8
	X^2	15.45	26.86	1.1	4.2	9.4	20.6	32.7	43.5	54.6	64.5	74.8	84.7	93.8
	M_n	15.10	26.29	0.5	3.9	8.7	18.9	29.8	41.5	53.4	63.4	73.4	82.2	91.6
100	G^2	16.94	32.16	1.9	8.8	16.1	29.2	45.5	56.5	66.4	73.3	81.8	88.3	94.4
	X^2	15.13	26.39	0.5	4.0	9.0	18.9	29.9	41.8	54.0	63.7	73.6	82.2	91.8
	M_n	14.89	29.89	0.8	5.6	10.5	19.3	29.2	39.2	49.1	58.5	68.5	80.0	89.5
1000	G^2	15.04	30.71	0.8	5.8	10.9	20.8	29.9	40.2	50.5	59.0	69.4	80.2	89.9
	X^2	14.92	29.99	0.8	5.7	10.6	19.5	29.3	39.3	49.4	58.6	68.8	80.1	89.7

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 $\boldsymbol{\Delta}_{21}=\frac{\partial \boldsymbol{\pi}_{12}}{\partial \boldsymbol{\tau}'}$. 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,

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

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 $\boldsymbol{\Delta}=\frac{\partial \boldsymbol{\pi}_{12}}{\partial \boldsymbol{\kappa}'}$. 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,

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

Proof of Equation (24):

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,

$$\mathbf{D}_{12}^{-1}\boldsymbol{\pi}_{12}=\mathbf{1} \qquad \boldsymbol{\Delta}'\mathbf{1}=\mathbf{0} \qquad \boldsymbol{\pi}_{12}'\mathbf{D}_{12}^{-1}=\mathbf{1}' \qquad \mathbf{1}'\boldsymbol{\Delta}=\mathbf{0}' \qquad (30)$$

and (17) we first notice that

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

Thus, $\mathbf{AC} = \mathbf{C}$ and $\mathbf{CA} = \mathbf{C}$. Furthermore, using (30) 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

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

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, $\left(\boldsymbol{\Omega}\mathbf{D}_{12}^{-1}\right)^2 \neq \boldsymbol{\Omega}\mathbf{D}_{12}^{-1}$. □

<u>Proof of Equation (28)</u>:

First, using a Taylor expansion we find analogously to the proof of (19) that

$$\sqrt{N}\tilde{\mathbf{e}} \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Omega}) \qquad \boldsymbol{\Omega} = (\mathbf{I}-\boldsymbol{\Delta}\mathbf{G})\boldsymbol{\Gamma}(\mathbf{I}-\boldsymbol{\Delta}\mathbf{G})'. \tag{31}$$

Then, by Lemma 1 in Khatri (1966), $\mathbf{U}$ in (27) can be written as

$$\mathbf{U} = \boldsymbol{\Delta}_c\left(\boldsymbol{\Delta}_c'\mathbf{D}_{12}\boldsymbol{\Delta}_c\right)^{-1}\boldsymbol{\Delta}_c' \tag{32}$$

where $\boldsymbol{\Delta}_c'$ is $(C - q) \times C$ matrix satisfying $\boldsymbol{\Delta}_c'\boldsymbol{\Delta} = \mathbf{0}$. Now, we write $\boldsymbol{\Omega}$ in (31) as $\boldsymbol{\Omega} = \mathbf{Y}\boldsymbol{\Gamma}\mathbf{Y}'$, where $\mathbf{Y} = \mathbf{I} - \boldsymbol{\Delta}\mathbf{G}$. Using (32), $\mathbf{Y}'\mathbf{U} = \mathbf{U}$ and $\mathbf{UY} = \mathbf{Y}$. Thus,

$$\boldsymbol{\Omega}\mathbf{U}\boldsymbol{\Omega} = \mathbf{Y}\boldsymbol{\Gamma}\mathbf{U}\boldsymbol{\Gamma}\mathbf{Y}' = \mathbf{Y}\boldsymbol{\Gamma}\mathbf{Y}' - \mathbf{Y}\boldsymbol{\Delta}\left(\boldsymbol{\Delta}'\mathbf{D}^{-1}\boldsymbol{\Delta}\right)\boldsymbol{\Delta}'\mathbf{Y}', \tag{33}$$

where in (33) we have used (27). When (26) holds, the second term is zero and therefore $\boldsymbol{\Omega}\mathbf{U}\boldsymbol{\Omega} = \boldsymbol{\Omega}$.

The degrees of freedom available for testing are given by $\operatorname{rank}(\boldsymbol{\Omega}\mathbf{U})$. Using the expression of $\mathbf{U}$ in (27) and (26), $\boldsymbol{\Omega}\mathbf{U} = \mathbf{I} - \boldsymbol{\Delta}\mathbf{G} - \boldsymbol{\pi}\mathbf{1}'$. This matrix is idempotent and therefore its rank equals its trace. The number of degrees of freedom available for testing is using (26) $\operatorname{tr}(\boldsymbol{\Omega}\mathbf{U}) = \operatorname{tr}(\mathbf{I}) - \operatorname{tr}(\boldsymbol{\Delta}\mathbf{G}) - \operatorname{tr}\left(\boldsymbol{\pi}\mathbf{1}'\right) = C - q - 1$.

Equation (26) can be verified for the two-stage estimator using $\mathbf{T}_1\boldsymbol{\Delta}_{21} = \boldsymbol{\Delta}_{11}$, $\mathbf{T}_2\boldsymbol{\Delta}_{21} = \boldsymbol{\Delta}_{12}$, so that $\mathbf{B}_1\boldsymbol{\Delta}_{21} = \mathbf{I}$. Also, $\mathbf{B}_{22}\boldsymbol{\Delta}_{22} = \mathbf{I}$. Finally, $\mathbf{T}_1\boldsymbol{\Delta}_{22} = \mathbf{0}$, $\mathbf{T}_2\boldsymbol{\Delta}_{22} = \mathbf{0}$, so that $\mathbf{B}_1\boldsymbol{\Delta}_{22} = \mathbf{0}$. □

NOTAS

NOTAS

NOTAS

Depósito Legal: M-20073   I.S.S.N.: 1579-4873