0701B 19-11-1986

A MULTIVARIATE ESTIMATION MODEL FOR NORMAL CORRELATION IN DISCRETE ORDERED VARIABLES **)

Jaap van der Pol & Jan de Leeuw *)

*) Adresses:
J. van der Pol,
Postbank N.V.,
P.O. Box 21009,
1000 EX AMSTERDAM,
The Netherlands.

Prof. Dr. J. de Leeuw,
Dept. of Data Theory,
Leiden University,
Hooigracht 15,
2312 KM LEIDEN,
The Netherlands.

**) This research was supported in part by the Netherlands Organisation for the Advancement of Pure Research, grant nr. 56-150.

Abstract

A multivariate estimation technique for normal correlation in discrete variables is introduced. It is shown to give efficient estimates under fairly general conditions. The performance of the technique is discussed, and illustrated with a bootstrap simulation study on a real-life dataset.

Keywords: tetrachoric correlation, polychoric correlation, minimum chi square estimation, joint-bivariate estimation.

1. INTRODUCTION

With the recent developments in multivariate moment structure modelling of discrete variables (e.g. Muthén & Christoffersson, 1981, Muthén, 1978, 1982, 1984), there is a renewed interest in the problem of efficient estimation of normal correlation in discrete, ordered variables. In the LISREL-program for instance (Jöreskog & Sörbom, 1981), an option for the calculation of polychoric correlations is incorporated and in Muthén (1984) the LISREL-methodology for structural equation modelling is extended to mixed sets of continuous and discrete variables, giving polychoric correlations for each pair of variables, GLS-estimates of the covariances of the polychorics and a chi-square diagnostic check on the multinormal model. Another important contribution is Poon & Lee (1987) who present a ML-estimation technique for the correlation matrix of a mixed set of multivariate normally distributed continuously and discretely observed random variables. All methods and models mentioned here give a consistent estimate of the polychoric correlation matrix, but only the Poon & Lee (1987) estimate is a full information ML-estimate, and therefore efficient if the condition of multivariate normality is met. So a statistically efficient solution to the multivariate polychoric correlation problem exists, though at the price of a laborious computing process because of the complicated likelihood function requiring the repeated evaluation of multiple integrals.

Using the likelihood principle, however, is not the only possibility to obtain efficient estimates of the polychoric correlation matrix. In this paper it will be shown that a multivariate extension of the Ritchie Scott (1918) polychoric correlation estimation model (i.e. the minimum variance weighted average of tetrachoric correlations) gives (almost) equally efficient polychorics and in a computationally easier way. A chi-squared test statistic on the model residuals provides a diagnostic check on the appropriateness of the model assumption of multivariate normality.

The discussion in this paper is structured as follows. In section 2 a short review of the Ritchie-Scott (1918) methodology for solving the normal correlation problem in discrete, ordered random variables is presented. In Section 3 the multivariate estimation model is defined and its statistical features are discussed. Section 4 gives some computational considerations with respect to the estimation technique. In Section 5, finally, the technique is illustrated in the spirit of K. Pearson, presenting results of a bootstrap simulation study on a recently analysed and published set of father and son IQ-data.

2. ESTIMATION OF NORMAL CORRELATION IN BIVARIATE TABLES

Let u and v be two random variables having a standard bivariate normal distribution with correlation parameter ρ. Suppose further that u and v are observed as discrete variables x and y with n and m categories, respectively. Given thresholds $-\infty<\alpha_1<\alpha_2<..<\alpha_{n-1}<\infty$ for u, x is defined by

$$\begin{aligned} x &= 1 \text{ if } & u &< \alpha_1, \\ x &= 2 \text{ if } & \alpha_1 \leq u &< \alpha_2, \\ &\;\vdots & &\;\vdots \\ x &= n \text{ if } & \alpha_{n-1} &\leq u, \end{aligned}$$

and y analogously by v and thresholds $-\infty<\beta_1<\beta_2<..<\beta_{m-1}<\infty$. Now if π_{ij} denotes the joint probability of $u<\alpha_i$ and $v<\beta_j$ then the model states that

$$\pi_{ij} = \text{prob}(x \leq i \;\&\; y \leq j) = \text{prob}(u < \alpha_i \;\&\; v < \beta_j) = \Phi_\rho(\alpha_i,\beta_j) \qquad i = 1,2,..,n, \quad j = 1,2,..,m, \tag{1}$$

with $\alpha_n = \beta_m = \infty$,

$$\Phi_\rho(\alpha_i,\beta_j) = \int_{-\infty}^{\alpha_i} \int_{-\infty}^{\beta_j} \phi_\rho(u,v)\, dv du, \tag{2}$$

and $\phi_\rho(u,v)$ the standard bivariate normal density with correlation ρ. If $\pi_{i+} = \text{prob}(u<\alpha_i)$ and $\pi_{+j} = \text{prob}(v<\beta_j)$ then (1) implies that $\pi_{i+} = \Phi(\alpha_i)$ and $\pi_{+j} = \Phi(\beta_j)$, with $\Phi(.)$ the standard normal distribution. The parameters of the model are $\rho,\alpha_1,..,\alpha_{n-1},\beta_1,..,\beta_{m-1}$.

A simple count in (1) shows nm - 1 independent probabilities which are to be explained by (n - 1) + (m - 1) + 1 parameters. Thus unless we have a fourfold table of observed frequencies (n = m = 2), the model is restrictive, and the number of restrictions it imposes is (n - 1)(m - 1) - 1. In general we cannot solve for $\rho, \alpha_1, .., \alpha_{n-1}, \beta_1, .., \beta_{m-1}$, given the π_{ij}, π_{i+}, and π_{+j}. Therefore, actual methods for computing estimates of these parameters concentrate on finding approximate solutions to the equations $p_{ij} = \Phi_\rho(\alpha_i, \beta_j)$, $p_{i+} = \Phi(\alpha_i)$, and $p_{+j} = \Phi(\beta_j)$ (Pearson & Pearson, 1922, Tallis, 1962, Lancaster & Hamdan, 1964, Martinson & Hamdan, 1971, Olsson, 1979). Here p_{ij}, p_{i+}, and p_{+j} are, of course, the sample analogues of the cumulative probabilities defined above.

The oldest, and for our purposes most interesting method was proposed by Ritchie-Scott (1918). This is the method that is generalized to the multivariate case below. Ritchie-Scott took as his point of departure the computation of the (n - 1)(m - 1) tetrachoric coefficients corresponding with the different points of dichotomy in the table. Each of these tetrachorics is a consistent, but not an efficient estimate of ρ in the n x m-table. The next step is to compute the linear combination of the tetrachorics (with weights adding up to one) which has the smallest possible variance. For this we obviously need the dispersion matrix of the (n - 1)(m - 1) tetrachorics, and as a consequence the method of Ritchie-Scott could be applied, until recently, only to very small tables.

Here a somewhat modernized and streamlined account is given of Ritchie-Scott's (1918) method. For this purpose (1) is 'reduced' to the tetrachoric model (n = m = 2) for a typical point of dichotomy (α_i, β_j) by collapsing all categories up to and including i, and likewise j, and similarly the categories above i, and j. In this tetrachoric model let $\alpha_i = \Phi^{-1}(\pi_{i+})$ and $\beta_j = \Phi^{-1}(\pi_{+j})$. Then, by the fact that π_{ij} is strictly increasing in ρ for fixed α_i and β_j, and by employing the implicit function theorem, it follows that an unique solution ρ_{ij} of $\pi_{ij} = \Phi_\rho(\alpha_i, \beta_j)$ exists. Since this can be done for each point of dichotomy (α_i, β_j) i = 1,.., n - 1, j = 1,..,m - 1, a one-to-one transformation of the π_{ij}, π_{i+}, π_{+j} to ρ_{ij}, α_i, β_j exists. In these new coordinates the polychoric model for the table is identical to the model that the (n - 1)(m - 1) values of ρ_{ij} are all equal. The transformation has consequently produced a linear model in the new coordinates.

Now apply the same transformations to the observed proportions, which gives estimates $\hat{a}_i$, $\hat{\beta}_j$ and $\hat{\rho}_{ij}$. These estimates are all consistent, but there are $(n-1)(m-1)$ estimates of the single parameter ρ, which is undesirable. Collect the $(n-1)+(m-1)$ estimates $\hat{\alpha}_i$ and $\hat{\beta}_j$ in the vector $\hat{\Theta}_1$, and the $(n-1)(m-1)$ estimates $\hat{\rho}_{ij}$ in the vector $\hat{\Theta}_2$. The joint dispersion of the asymptotic distribution of $(\Theta'_1, \Theta'_2)'$ is V, with submatrices V_{11}, V_{12}, V_{22}. Note that V is positive definite, unless $|\rho| = 1$, in which case the model degenerates. A consistent estimate $\hat{V}$ of V is asymptotically positive definite with probability one. Its inverse, $\hat{V}^{-1}$ has submatrices written as $\hat{V}^{11}$, $\hat{V}^{12}$, and $\hat{V}^{22}$. By standard statistical large sample theory (Ferguson, 1958, Wijsman, 1959, 1960) efficient estimates of the parameters are obtained from a sample of N observations by minimizing the chi square function

$$g(\Theta_1,\Theta_2) = N\{(\hat{\Theta}_1 - \Theta_1)'\hat{V}^{11}(\hat{\Theta}_1 - \Theta_1) + 2(\hat{\Theta}_1 - \Theta_1)'\hat{V}^{12}(\hat{\Theta}_2 - \Theta_2) + (\hat{\Theta}_2 - \Theta_2)'\hat{V}^{22}(\hat{\Theta}_2 - \Theta_2)\}, \tag{3}$$

with Θ_1 the vector of threshold parameters α_i and β_j, and with Θ_2 the vector of tetrachorics ρ_{ij}. There are no restrictions on Θ_1, while all elements of Θ_2 must be equal. Given Θ_2 (3) is minimized over Θ_1 by setting

$$\tilde{\Theta}_1 = \hat{\Theta}_1 + (\hat{V}^{11})^{-1}\hat{V}^{12}(\hat{\Theta}_2 - \Theta_2). \tag{4}$$

The value of the minimum over Θ_1, given Θ_2, is

$$g(*,\Theta_2) = N(\hat{\Theta}_2 - \Theta_2)'\{\hat{V}^{22} - \hat{V}^{21}(\hat{V}^{11})^{-1}\hat{V}^{12}\}(\hat{\Theta}_2 - \Theta_2) = $$

$$= N(\hat{\Theta}_2 - \Theta_2)'\hat{V}_{22}^{-1}(\hat{\Theta}_2 - \Theta_2). \tag{5}$$

This result is of importance, because it shows that for efficient estimation of the polychoric correlation coefficient only V_{22}, a consistant estimate of the the dispersion matrix of the tetrachorics, is needed. Covariances between tetrachorics and boundary points are not needed.

By writing Θ_2 in the form $\Theta_2 = \rho u$, where u is a vector with all elements equal to +1, and by differentiating (5) with respect to ρ, it follows that the estimate of the polychoric correlation is

$$\hat{\rho} = (u'\hat{V}_{22}^{-1}\hat{\Theta}_2)/(u'\hat{V}_{22}^{-1}u). \tag{6}$$

The minimum value of (5) is

$$g(*,*) = N\{\hat{\Theta}_2'\hat{V}_{22}^{-1}\hat{\Theta}_2 - (u'\hat{V}_{22}^{-1}\hat{\Theta}_2)^2/(u'\hat{V}_{22}^{-1}u)\}, \tag{7}$$

which is asymptotically a chi square with (n - 1)(m - 1) - 1 degrees of freedom. This statistic can be used to test (discreticized) bivariate normality in a general n x m table. It remains to be shown that this is actually the same procedure as the one discussed by Ritchie-Scott (1918). A linear combination $h'\Theta_2$ of the tetrachorics has dispersion $h'V_{22}h$. This must be minimized under the condition that $h'u = 1$. The solution is $\hat{h} = \hat{V}_{22}^{-1}u/(u'\hat{V}_{22}^{-1}u)$, and from (6) it follows that $\hat{\rho}$ is an estimate of $h'\Theta_2$.

The theory in this section is quite satisfactory and useful, but in order to apply it an estimate of V_{22}, the dispersion of the asymptotic distribution of the (n - 1)(m - 1) tetrachorics is needed. This matrix is obtained using the results derived in the Appendix. It is perhaps important to emphasize that if also efficient estimates of the α_i and β_j are wanted, then the matrices V_{11} and V_{12} must be estimated as well. These formulas are not given here explicitly, but the general approach of the Appendix makes it relatively simple to derive them.

3. JOINT BIVARIATE ESTIMATION IN THE BLOCK MULTINORMAL MODEL

An obvious next step in the development of the Pearson-approach to multivariate analysis with discrete variables is an attempt to generalize the models and techniques from the previous sections to the multivariate case. The model is easy to generalize, and this is done first.

If there are k observed variables, then their probability distribution can be displayed as a k-dimensional contingency table. Each cell in the table has probability equal to the integral of the multivariate normal density $\phi_{\Gamma}(x)$ over the parallelopiped in R^k defined by the cell boundaries (some of which may be infinite). Here Γ is the correlation matrix of the k variables, which without loss of generality may be assumed to have zero mean and unit variance. De Leeuw (1983) calls the resulting model the block-multinormal model.

The next step in the previous section was to find a one-to-one transformation of the probabilities in the table which simplified the model. In the multivariate case this is much more difficult for various reasons. The most important one is that the parameters of the multivariate normal distribution are - in standardized form - the $\frac{1}{2}k\,(k-1)$ correlations of all different sets of two variables. So the parameter structure is essentially bivariate. As a consequence the bivariate normal model can be fitted exactly to a 2 x 2 table, but the multivariate normal model cannot be fitted exactly to a 2^k table with $k > 2$. Moreover, an estimation method based on the multivariate tetrachoric series (Kendall, 1941) would be of limited value since the series diverges for many correlation structures (Harris and Soms, 1980). It appears that the transformation approach of the previous section does not work in the block-multinormal model, at least not adequately.

On the other hand it is clear that, in principle at least, the maximum likelihood method (or a minimum chi-square method) can be applied to the block-multinormal model directly, i.e. without transformations. We have expressions for the cell probabilities, which is all we need (see Poon & Lee, 1987). But with such a technique the complications simply occur at a different level. In order to evaluate a cell probability we have to compute a multivariate integral, and this is in most cases difficult and laborious. And if k is at all large, we deal unavoidably with a large number of empty cells in the observed table, which makes it difficult to rely on the asymptotic results of maximum likelihood estimation. Even if we choose to do so, we may at least expect the sampling errors of the resulting estimates to be extremely high.

The above considerations suggest that it is unwise, at least for the moment, to apply the block-multinormal model in situations where we deal with, say, five or more variables. It seems that the general model is beyond our present capacities, both computationally and statistically. Consequently in this paper a 'derivative' of the general model is considered in which the unobserved continuous variables $u_1,\ldots,u_k$ are not assumed to be jointly multivariate normally distributed, but merely that all $\binom{k}{2}$ bivariate marginals are normally distributed. Of course, multivariate normality implies joint bivariate normality, but multivariate normality is much stronger because it also implies relations between higher order probabilities. It is also clear that the empty cell problem will be less serious for the joint bivariate block-normal model, because the bivariate marginal tables are much better filled than the multivariate table. Here follows the analysis for the joint bivariate model.

In this case the appropriate transformation is easy to find. All bivariate tables are simply transformed to boundary points and tetrachoric correlations, and all higher order frequencies are not transformed, but left as they are. This defines a one-one transformation again, and the joint bivariate block-normal model is equivalent to the linear model which asserts that all tetrachorics corresponding to the same bivariate marginal are equal. If variable j has n_j categories, then we have $(n_j - 1)(n_\ell - 1)$ tetrachorics for the pair (j,ℓ).Collect all $\sum_{j<\ell}(n_j-1)(n_\ell-1)$ tetrachorics in the vector Θ. Then the model says that $\Theta = S\gamma$, with γ the vector of $\binom{k}{2}$ correlation coefficients and S a binary selection matrix which connects the separate tetrachorics with their respective bivariate tables. S is chosen such that each column of S corresponds with one table, say of pair (j,ℓ). It has $(n_j - 1)(n_\ell - 1)$ elements equal to +1 and all other elements in the column are zero. Using exactly the same reasoning as in the previous section shows that efficient estimates of the vector γ can be found by minimizing the chi square function

$$g(\gamma) = N(\hat{\Theta} - S\gamma)' \hat{V}^{-1} (\hat{\Theta} - S\gamma), \qquad (8)$$

where $\hat{V}$ estimates the dispersion matrix of all tetrachorics jointly (compare the Appendix). Observe that the estimates are fully efficient in the joint bivariate block-normal model, and the minimum value of the chi square statistic gives a test for the appropriateness of this model. If interpreted in terms of the multivariate block-normal model the estimates are limited information estimates, which are consistent but not fully efficient. The chi square statistic then merely tests a necessary condition for block multivariate normality.

The minimum of (8) can be found once the dispersion matrix V is computed. The estimate of γ, a vector of $\binom{k}{2}$ elements, is simply

$$\hat{\gamma} = (S'\hat{V}^{-1}S)^{-1}S'\hat{V}^{-1}\hat{\theta}, \tag{9}$$

and the minimum value of the loss function is

$$g(*) = N\hat{\Theta}'(\hat{V}^{-1} - \hat{V}^{-1}S(S'\hat{V}^{-1}S)^{-1}S'\hat{V}^{-1})\hat{\Theta} \tag{10}$$

which is the chi square for goodness-of-fit of the joint bivariate block-normal model. It follows directly from (9) that the dispersion of the asymptotic distribution of $\hat{\gamma}$ is given by

$$E((\hat{\gamma} - \gamma)(\hat{\gamma} - \gamma)') = (S'\hat{V}^{-1}S)^{-1}. \tag{11}$$

Because the procedure presented here requires more computation that the bivariate procedure of Olsson (1979), used in LISREL, it should at least be indicated what its theoretical advantages are. In the first place, and most importantly, there is information about the covariance of the asymptotic distribution of the polychoric correlations. It follows directly from (8) that $N^{1/2}(\hat{\gamma} - \gamma)$ has asymptotic dispersion $(S'V^{-1}S)^{-1}$. Of course the covariance of the correlations is needed if parameters of a structural model must be fitted efficiently to the polychoric correlations at a later stage in our investigations (see, for example, De Leeuw, 1983, Muthén, 1984).

A second advantage of our method, already indicated above, is that the procedure is fully efficient in the joint bivariate model. Since the method of Olsson does not take covariation between different bivariate distributions into account, it is a limited information method for the joint bivariate model, and consequently not efficient.

In the third place it follows from the usual asymptotic statistical theory that our method is efficient in the block multinormal model, provided we suitably restrict the class of competing estimates. More precisely, if we define the class C of estimates which are differentiable functions of the bivariate marginals, and which are Fischer-consistent for the polychoric correlations, then our method has the same asymptotic dispersion as the best method in this class (see De Leeuw, 1983). Olsson's method is in C, and so our method generally will be better. Observe, however, that our method itself is not in C, because it uses fourth order frequencies in the estimation of V (see the Appendix).

How much better our method will be, is relatively easy to estimate. Olsson's estimate of γ, asymptotically equivalent to γ, is the generalized least squares estimate of (8) with V replaced by V_B, the block-diagonal matrix in which the diagonal blocks of V are copied. Thus

$$\tilde{\gamma} = (S'\hat{V}_B^{-1}S)^{-1}S'\hat{V}_B^{-1}\hat{\theta}. \tag{12}$$

Its asymptotic dispersion, $(S'V_B^{-1}S)^{-1}S'V_B^{-1}VV_B^{-1}S(S'V_B^{-1}S)^{-1}$, is larger than $(S'V^{-1}S)^{-1}$, the dispersion of γ. How much larger depends on the particular application.

4. COMPUTATIONAL CONSIDERATIONS

Formulas (9), (10), and (11) give the desired answer to the statistical (first order) problems associated with the joint bivariate model. For practical purposes however, these formula's are not well-suited. Suppose we analyse 10 variables with, say, 5 categories each, then each bivariate table gives 16 tetrachorics, and the total number of tetrachorics is 45 x 16 = 720. Consequently about 260,000 different covariances between tetrachorics have to be computed, and the resulting matrix of order 720 must be inverted. This last operation may cause trouble as to computing time, storage, and numerical instability. Therefore an alternative formulation of the problem is considered, which seems more friendly in these respects. It is also seen directly to be the joint bivariate generalization of the method of Ritchie-Scott (1918).

Consider the minimization of $\text{tr}\, A'\hat{V}A$ of all matrices A that satisfy $A'S = I$. A must have the same dimensions as S, i.e. the number of tetrachorics times the number of polychorics. The solution for A is

$$\hat{A} = \hat{V}^{-1}S(S'\hat{V}^{-1}S)^{-1}. \tag{13}$$

Thus, from (9), $\hat{\gamma} = \hat{A}'\hat{\Theta}$. Because $A'S = I$ we see that the polychorics are a linear combination of tetrachorics, with coefficients adding up to one. $\hat{A}$ is computed iteratively, by using a conjugate gradient method with linear restrictions (Pshenichny and Danilin, 1978, chapter 3). The method is self-correcting, it can be truncated if there is no subsequent improvement of the objective function $\text{tr}\, \hat{A}'\hat{V}\hat{A}$, and it does not require explicit inversion of $\hat{V}$. A convenient and relatively cheap initial estimate of A is found if we restrict A to be of the form $a_1 \dot{+} \dots \dot{+} a_K$ with $\dot{+}$ the direct sum and with $K = \binom{k}{2}$. Thus the initial A has non-zero elements in a row only for the tetrachorics corresponding to that row. In other words, we initially compute polychorics for each table separately, which has the additional advantage that we also have a partitioning of the overall chi square into components, one component for each bivariate distribution. It will also be clear from a comparison of initial and final estimates how much is gained by going from a separate to a joint bivariate estimation method.

5. A BOOTSTRAP SIMULATION STUDY ON REAL LIFE DATA

It will be clear from the definition of model (8) that the accuracy of the polychoric estimates largely depends on the (un)biasedness and robustness of the tetrachoric estimator. Indications for the existence and magnitude of either effect in the latter can be found in contributions of Pearson & Heron (1913), Castellan (1966), Froemel (1971), Kirk (1973), and Brown & Benedetti (1977). Generally speaking all estimators prove to be more or less susceptible to bias in case of tables with small cell-frequencies (less than five), whether or not in combination with a "high" correlation.

As to robustness against violations of normality, less is known. Again, in general, one may say that the use of extreme points of dichotomization should be avoided (if necessary, this can be done by regrouping the data). Nevertheless, the tetrachoric estimator may be unbiased and robust whereas the polychoric may not be. In order to examine the possibility of such effects a Monte-Carlo study was performed on both the tetrachoric and polychoric estimator in a small example. The tetrachoric correlations were computed with Digvi's routine (Digvi, 1979).

To save space only the main results are mentioned here. A more detailed description is presented in Van der Pol & De Leeuw (1986). Analysing 400 samples of respectively 200, 400 and 1600 observations of computer generated normally distributed data on three variables with three categories, uniform marginal observed distributions, and correlations .8, .6, and .5, it showed that on average the tetrachoric estimates are slightly negatively biased. The polychorics have a similar tendency in the positive direction. Both effects, however, are not marked enough to be disturbing. So with regard to the polychorics the simulation results are in close agreement with those of Olsson (1979).

In a second example the effect of regrouping data on the polychoric correlation estimator has been examined. Again only main results are mentioned. Recoding an 18 x 18 table of heights of fathers and daughters (Pearson & Lee, 1903) according to the types presented in Lancaster & Hamdan (1964), and comparing several correlation estimators it showed that the joint-bivariate correlation coefficients and the Olsson (1979) ML-estimates identical are within two decimal places for various types of collapsing adjacent rows and adjacent columns. Collapsing opposite rows and/or opposite columns, however, gives highly divergent and inconsistent results for all types of estimators. Again, details are in Van der Pol & De Leeuw (1986).

A third example more appropriate for the joint-bivariate model presented here is a bootstrap simulation study on a set of (non-verbal) intelligence (IQ), and social economic status variables (SES) analysed in Vroon et al. (1986). The data set used contains 2776 IQ-scores of fathers and sons measured at the age of eighteen years by means of army intelligence tests for Dutch conscripts. A third variable added is a combined measure for the social economic status of the household both belong to. The SES-variable is a combination of educational level of the father, and of the mother, and occupational level of the father. To keep the presentation short, the three variables were recoded to three categories by collapsing adjacent categories such that approximately a uniform observed marginal distribution was obtained for each variable. From this dataset 100 bootstrap samples were generated and analysed. The averaged results are given in Table 1. It shows tetrachoric correlations for the three bivariate tables, their computed and estimated standard errors, the bivariate and the joint-bivariate polychoric correlations and their computed and estimated standard errors, three teststatistics for the bivariate, and one for the joint-bivariate estimation model. Note that in this particular simulation study on the average only a small improvement for the joint-bivariate polychorics over the strictly bivariate polychorics has been established. Note also the close agreement between the estimated and computed standard deviations of the coefficients.

In Table 2 and Table 3 estimates of the polychorics' large sample covariances are shown. The estimates in Table 2 are the consistent large sample covariance estimates as computed from the bivariate polychorics and the tetrachorics' covariance matrix, i.e.

$$C = (S_o' V^{-1} S_o)^{-1}, \tag{14}$$

with S_o the bivariate solution of (13) and initial matrix for the joint-bivariate estimation algorithm. These estimates are asymptotically identical to Muthén's (1984) covariance estimates of the polychorics (see also Browne, 1984). In Table 3 the joint-bivariate efficient estimates are shown. Comparing the various results provides an insight into the improvement achieved when passing from the strictly bivariate to the joint-bivariate estimation model. As to that aspect of this particular simulation the overall conclusion is obvious: only a very small improvement can be noticed in the correlation, and standard errors are on average slightly smaller than the strictly bivariate estimates. Moreover, the test statistics indicate that on average both the strictly bivariate and the joint-bivariate estimation model has to be rejected (in both cases 56 out of 100 times). Clearly this result is partly due to the choice of data and the processing of the IQ-, and SES-variables. But the fact that in Poon & Lee (1987), using artificially generated data, somewhat similar findings are shown, might indicate that the effect is possibly relatively common. Further exploration on this topic is needed.

Table 1
Bootstrap simulation study of tetrachoric and polychoric correlations.
Chi squared test statistics of the bivariate and joint-bivariate estimation model

		Mean	S.e.			Mean	S.e.
<u>IQ-father x IQ-son</u>							
Tetrachoric:	(1,1)	0.3330	0.0328	S.e. tetrachoric:	(1,1)	0.0294	0.0006
	(1,2)	0.3350	0.0281		(1,2)	0.0293	0.0004
	(2,1)	0.3000	0.0309		(2,1)	0.0311	0.0005
	(2,2)	0.2912	0.0281		(2,2)	0.0296	0.0004
z-value	(1)	- 0.4164	0.0265	z-value	(1)	- 0.5052	0.0278
IQ-father	(2)	0.4809	0.0221	IQ-son	(2)	0.3486	0.0228
Polychoric (j.biv.)		0.3162	0.0213	Polychoric (biv.)		0.3167	0.0219
S.e.		0.0219	0.0002	X^2 (DF = 3)		5.4503	3.5730
<u>IQ-father x SES</u>							
Tetrachoric:	(1,1)	- 0.0051	0.0336	S.e. tetrachoric:	(1,1)	0.0355	0.0004
	(1,2)	0.0679	0.0334		(1,2)	0.0359	0.0004
	(2,1)	0.0434	0.0325		(2,1)	0.0308	0.0001
	(2,2)	0.0879	0.0303		(2,2)	0.0310	0.0002
z-value	(1)	- 0.4164	0.0265	z-value	(1)	- 0.8739	0.0276
IQ-father	(2)	0.4809	0.0221	SES	(2)	0.1388	0.0221
Polychoric (j.biv.)		0.0521	0.0234	Polychoric (biv.)		0.0517	0.0234
S.e.		0.0239	0.0001	X^2 (DF = 3)		8.2639	5.0844
<u>IQ-son x SES</u>							
Tetrachoric:	(1,1)	0.0509	0.0360	S.e. tetrachoric:	(1,1)	0.0358	0.0005
	(1,2)	0.0837	0.0333		(1,2)	0.0351	0.0003
	(2,1)	0.1107	0.0317		(2,1)	0.0311	0.0002
	(2,2)	0.0869	0.0277		(2,2)	0.0304	0.0001
z-value	(1)	- 0.5052	0.0278	z-value	(1)	- 0.8739	0.0276
IQ-son	(2)	0.3486	0.0228	SES	(2)	0.1388	0.0222
Polychoric (j.biv.)		0.0951	0.0221	Polychoric (biv.)		0.0869	0.0222
S.e.		0.0238	0.0001	X^2 (DF = 3)		6.3962	4.1163
X^2 (j.biv.)(DF = 9)		19.7956	7.6304				

Table 2
Covariance matrix polychorics (+ standard errors)
Bivariate model

	IQ-father		IQ-son		SES	
IQ-father	1.5802	(0.0175)				
IQ-son	0.0261	(0.0331)	1.3400	(0.0340)		
SES	0.3809	(0.0361)	0.1319	(0.0301)	1.5848	(0.0161)

Table 3
Covariance matrix polychorics (+ standard errors)
Joint-bivariate model

	IQ-father		IQ-son		SES	
IQ-father	1.5661	(0.0190)				
IQ-son	0.0282	(0.0330)	1.3359	(0.0337)		
SES	0.3815	(0.0368)	0.1317	(0.0299)	1.5801	(0.0161)

REFERENCES

Brown, M.B. & Benedetti, J.K. (1977). On the mean and the variance of the tetrachoric correlation coefficient. Psychometrika, 42, 347-355.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62-83.

Castellan, N.J. Jr. (1966). On the estimation of the tetrachoric correlation coefficient. Psychometrika, 31, 67-73.

De Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. Annals of Applied Econometrics, 22, 113-138.

Digvi, D.R. (1979). Calculation of the tetrachoric correlation coefficient. Psychometrika, 44, 169-172.

Ferguson, T.S. (1958). A method for generating best asymptotically normal estimates with application to the estimation of bacterial desities. Annals of Methematical Statistics, 29, 1046-1062.

Froemel, E.C. (1971). A comparison of computer routines for the calculation of the tetrachoric correlation coefficient. Psychometrika, 36, 167-174.

Harris, B. & Soms, A.P. (1980). The use of the Tetrachoric Series for Evaluating Multivariate Normal Probabilities. Journal of Multivariate Analysis, 10, 252-267.

Jöreskog, K.G. & Sörbom, D. (1981). LISREL V user's guide. Chicago: International Educational Resources.

Kendall, M.G. (1941). Proof of Relations connected with the Tetrachoric Series and its Generalization. Biometrika, 32, 196-198.

Kirk, D.B. (1973). On the numerical approximation of the bivatiate normal (tetrachoric) correlation coefficient. Psychometrika, 38, 259-268.

Lancaster, H.O. & Hamdan, M.A. (1964). Estimation of the correlation coefficient in contingency tables with possibly nonmetrical characteristics. Psychometrika, 29, 383-391.

Martinson, E.O. & Hamdan, M.A. (1971). Maximum Likelihood and some other Asymptotically Efficient Estimators of Correlation in two way Contingengy Tables. Journal of Statistical Computation and Simulation, 1, 45-54.

Martinson, E.O. & Hamdan, M.A. (1975). Calculation of the Polychoric Estimate of Correlation in Contingency tables. Applied Statistics, 24, 272-278.

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

Muthén, B. (1979). A structural probit model with latent variables. Journal of the American Statistical Association, 74, 807-811.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. Journal of Econometrics, 22, 43-65.

Muthén, B. & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. Psychometrika, 46, 407-419.

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

Pearson, K. & Heron, D. (1913). On theories of association. Biometrika, 9, 159-315.

Pearson, K. & Lee, A. (1903). On the laws of inheritance in man. Biometrika, 14, 127-156.

Poon, W.Y. & Lee, S.Y. (1987). Maximum Likelihood Estimation of Multivariate Polyserial and Polychoric Correlation Coefficients. Psychometrika, 52, forthcoming.

Pshenichny, B.N. & Danilin, Y.M. (1978). Numerical methods in extremal problems. Moscow: Mir Publishers.

Ritchie-Scott, A. (1918). The correlation coefficient of a polychoric table. Biometrika, 12, 93-133.

Tallis, G.M. (1962). The maximum likelihood estimation of correlation from contingency tables. Biometrics, 18, 342-353.

Van der Pol, J. & De Leeuw, J. (1986). Aspects of the estimation of the polychoric correlation coefficient. Unpublished paper. Amsterdam: Postbank N.V.

Vroon, P.A., De Leeuw, J. & Meester, A.C. (1986). Distributions of intelligence and educational level in fathers and sons. British Journal of Psychology, 77, 137-142.

Wijsman, R.A. (1959). On the theory of BAN estimates. Annals of Mathematical Statistics, 30, 185-191.

Wijsman, R.A. (1960). Correction to "On the theory of BAN estimates" Annals of Mathematical Statistics, 31, 1268-1270.

APPENDIX

Asymptotic distribution of the tetrachoric correlation coefficient.

Suppose we have a fourfold table with ~~observed~~ (non-vanishing) cell proportions $p_{00}, p_{01}, p_{10}, p_{11}$.
Then the tetrachoric correlation coefficient can be computed by solving the system of equations

$$\begin{aligned} \Phi(\alpha) - p_{00} - p_{01} &= 0 \\ \Phi(\beta) - p_{00} - p_{10} &= 0 \\ \Phi_\rho(\alpha,\beta) - p_{00} &= 0 \end{aligned} \qquad \text{(A1)}$$

with Φ, Φ_ρ, α, and β as defined in section 2.
If we rewrite system (A1) as

$$f\,(\Theta,p) = 0 \qquad \text{(A2)}$$

where $f = (f_1, f_2, f_3)'$, $\Theta = (\alpha, \beta, \rho)'$ and $p = (p_{00}, p_{01}, p_{10})'$ then - by the implicit function theorem -

$$\frac{\partial \Theta}{\partial p} = -\left(\frac{\partial f}{\partial \Theta}\right)^{-1} \cdot \frac{\partial f}{\partial p} \qquad \text{(A3)}$$

The partial derivatives can be calculated in a straightforward fashion with the following well-known identities

$$\Phi_\rho(\alpha,\beta) = \int_\alpha^\infty \phi\,(x) \int_w^\infty \phi(w)\, dwdx, \quad w = (\beta - \rho x)/(1-\rho^2)^{1/2}, \qquad \text{(A4)}$$

$$\frac{\partial \Phi_\rho(\alpha,\beta)}{\partial \rho} = \phi_\rho(\alpha,\beta)\,, \qquad \text{(A5)}$$

where ϕ is the standard normal density and ϕ_ρ the bivariate standard normal density with correlation ρ. The derivatives are

$$\frac{\partial\alpha}{\partial p_{00}} = \frac{\partial\alpha}{\partial p_{01}} = \phi(\alpha)^{-1}$$

$$\frac{\partial\beta}{\partial p_{00}} = \frac{\partial\beta}{\partial p_{10}} = \phi(\beta)^{-1}$$

$$\frac{\partial\rho}{\partial p_{00}} = [1 - K_\rho(\alpha,\beta) - K_\rho(\beta,\alpha)]/\phi_\rho(\alpha,\beta) \qquad \text{(A6)}$$

$$\frac{\partial\rho}{\partial p_{01}} = -K_\rho(\alpha,\beta)/\phi_\rho(\alpha,\beta)$$

$$\frac{\partial\rho}{\partial p_{10}} = -K_\rho(\beta,\alpha)/\phi_\rho(\alpha,\beta)$$

and $K_\rho(x,y) = \Phi((y - \rho x)/(1 - \rho^2)^{1/2})$. By substituting the derivatives from (A6) we find the following expansion for the observed tetrachoric correlation $\hat{r}$

$$N^{1/2}(\hat{r} - \rho) = \{[1-K_\rho(\alpha,\beta) - K_\rho(\beta,\alpha)]z_{00} - K_\rho(\alpha,\beta)z_{01}$$

$$- K_\rho(\beta,\alpha)z_{10}\}/\phi_\rho(\alpha,\beta) + o_p(1) \qquad \text{(A7)}$$

with $z_{00} = N^{1/2}(\hat{p}_{00} - \pi_{00})$, $z_{01} = N^{1/2}(\hat{p}_{01} - \pi_{01})$, and $z_{10} = N^{1/2}(\hat{p}_{10} - \pi_{10})$ where π_{00}, π_{01}, π_{10} stand for true cell proportions and N the number of observations.

Now suppose $\hat{r}_1$ and $\hat{r}_2$ are two tetrachoric correlation coefficients computed from the same multivariate table. It does not matter if they are computed from the same bivariate marginal or not, i.e. it does not matter if they estimate the same polychoric correlation coefficient. Using the notation of (A7) we can write

$$N^{1/2}(\hat{r}_1 - \rho_1) = g_1'\hat{z}_1 + o_p(1), \qquad \text{(A8)}$$

$$N^{1/2}(\hat{r}_2 - \rho_2) = g_2'\hat{z}_2 + o_p(1). \qquad \text{(A9)}$$

Here g_1 and g_2 are vectors with three elements containing the three derivatives from (A7), and $\hat{z}_1$ and $\hat{z}_2$ contain normalized proportions from the fourfold tables corresponding with the tetrachorics. In terms of the vector π and $\hat{p}$, which contain all the expected and observed cell probabilities from the complete multivariate table, respectively, we can write

$$\hat{z}_1 = N^{1/2} A_1 (\hat{p} - \pi), \qquad \text{(A10)}$$

$$\hat{z}_2 = N^{1/2} A_2 (\hat{p} - \pi), \qquad \text{(A11)}$$

with A_1 and A_2 appropriate summation matrices.

If the multivariate table has L cells, then A_1 and A_2 are 3 x L matrices with elements equal to zero or one. The effect of applying A_1 or A_2 to the multivariate probabilities, or proportions, is to add the relevant cell entries in such a way that the bivariate marginals for the fourfold tables are formed. It follows from (A8) - (A11) that the asymptotic covariance of $N^{1/2}(\hat{r}_1 - \rho_1)$ and $N^{1/2}(\hat{r}_2 - \rho_2)$ is equal to

$$g'_1 A_1 (\Pi - \pi\pi') A'_2 g_2, \qquad \text{(A12)}$$

with Π = diag (π). By substituting $r_1 = r_2$ and $\rho_1 = \rho_2$, formula (A12) can be used to estimate the variances of the tetrachorics.

Of course this development generalizes directly to any vector with tetrachoric correlations, selected from the same, or different bivariate tables. All we have to know to compute the elements of the asymptotic dispersion matrix $\hat{V}$ of the tetrachorics, are the vector of derivatives g, which has just three elements per tetrachoric, and the corresponding matrix A. For each tetrachoric there is a consistent estimate $\hat{g}$ of g, obtained by inserting the sample quantities in the expression for g, and a (unique) A. Computationally, as in other sparse matrix expressions, we do not think of A as a matrix but as a recipe to select indices. Then formula (A12) reduces considerably, and only frequencies up to the fourth order are required when computing the dispersion of the tetrachorics.