I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0321

IDENTIFICATION OF POLYCHORIC CORRELATIONS: A COPULA APPROACH

C. ALMEIDA and M. MOUCHART

http://www.stat.ucl.ac.be

Identification of polychoric correlations: A copula approach*

Carlos Almeida AND Michel Mouchart†

Institut de statistique
Université catholique de Louvain
20, Voie du Roman Pays
1348 Louvain-la-Neuve
Belgium

Abstract

The traditional model underlying the polychoric correlations among ordinal variables is revisited. This model relies on the idea of considering ordinal variables as discretization of corresponding continuous latent variables. The non-identification of the marginal distributions of the latent vector naturally leads to a copula approach; by so-doing, the role of the multivariate normality hypothesis of the latent vector is re-assessed.

Keywords: Ordered Variables, Polychoric Correlations, Identification, Gaussian multidimensional copula.

*Financial support from the contract 'Projet d'Actions de Recherche Concertées' nr 98/03–217 from the Belgian government, and from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged. Without implications, the authors gratefully acknowledge useful comments on earlier versions by Ph. Lambert.

†Corresponding author: mouchart@stat.ucl.ac.be

1 Introduction

Under the usual hypotheses of Normality and Linearity, the LISREL model boils town to a structural specification of the variance matrix of the observed data, but in the case of ordinal variables, the normality is obviously not acceptable and the covariance matrix is not anymore a good summary of data; in particular, it is not a sufficient statistic.

A standard practice consists in using polychoric correlations the same way as Pearson's correlations, see *e.g.* Muthén (1983), Muthén (1984), Jöreskog *et al.* (2002). Polychoric correlations are based on the interpretation of each ordinal variable as a discretization of a corresponding continuous latent variable (hereafter called *ideally measured* variable), assumed to follow a multivariate normal distribution.

So, under normality, the parameters describing the model are the parameters of the multivariate normal distribution (mean and variance) and the thresholds defining the discretization. Some identifiability restrictions are typically imposed fixing the means to zero and the variances to one; the remaining correlation matrix is precisely the matrix of polychoric correlations. Some estimation methods have been developed based on a fit function (Maximum Likelihood or Generalized Linear Square) in order to find consistent estimators of parameters and their asymptotic covariance matrix, see Olsson (1979), Jöreskog (1994), Poon and Lee (1987).

The interpretation (or modelling) of ordinal variables as a discretization of continuous latent variables is clearly invariant under increasing transformations: not only the scale and the origin are arbitrary but also the choice of coordinates. This suggests that the content of the structural equations in a LISREL model should mean that there exists a choice of coordinates that makes the linearity of the structural equation and the exogeneity assumption meaningful.

For testing purposes, the comparison of the polychoric correlations model with the multinomial saturated model has been implemented in statistical software as PRELIS, EQS, etc., and given the fact that the univariate normality of the ideally measured variables is not identified, the question is: What is the actual object of these tests? In order to understand the role of the normality hypothesis, we propose a specification of the model using the copula concept which separates the marginal structure of the latent variables and the dependence one. In the light of this specification the normality hypotheses is re-examined.

In next section, we remind the model leading to the discretization approach for ordinal variables and consider a first step in the identification analysis; we also recast the concept of polychoric correlation from an identi-

fication point of view. Section 3 particularizes the identification analysis to the case of a normal specification, Section 4 suggests an approach based on copula and contains the main results of this paper. A final section gathers some concluding remarks. The analytical subtleties of the paper are relegated to an Appendix that contains the proof of the main theorem.

2 A General Specification

We start with a satured multinomial model on which the discretization hypothesis is introduced in order to take into account the order of the observed variables. A general result on identifiability will be presented.

2.1 Multinomial Model

Let X be a vector of K categorical ordinal variables coded with labels $1, \ldots, r_k$:

$$X = (X_1, \ldots, X_K)' \in \prod_{1 \leq k \leq K} \{1, \ldots, r_k\} \equiv R_X, \quad \text{card } R_X = \prod_{1 \leq k \leq K} r_k,$$

where R_X denotes the range of X. A disjunctive coding may be obtained by defining: $Z_{\underline{j}} = \mathbf{1}_{\{X=\underline{j}\}}$ for each $\underline{j} = (j_1, \ldots, j_K) \in R_X$; more specifically:

$$Z = (Z_{\underline{j}} : \underline{j} \in R_X) \in \left\{ z \in \{0,1\}^{\text{card } R_X} : \sum_{\ell \in R_X} z_\ell = 1 \right\} \equiv R_Z. \tag{1}$$

The *saturated model* is provided by a generalized Bernoulli distribution:

$$Z \sim GBe_{(d+1)}(\theta_{SA}), \tag{2}$$

where

$$\begin{aligned}
&d = \prod_{1 \leq k \leq K} r_k - 1 = \text{card } R_X - 1, \\
&\theta_{\underline{j}} = \mathbb{E}[Z_{\underline{j}}] \in [0,1], \\
&\theta_{SA} = (\theta_{\underline{j}} : \underline{j} \in R_X) \in \Theta_{SA} = \left\{ u \in [0,1]^{d+1} : \sum_{\underline{j} \in R_X} u_{\underline{j}} = 1 \right\},
\end{aligned}$$

where the subscript SA statnds for "saturated". Thus $\Theta_{SA} = \mathcal{S}_d$ ($\equiv$ Simplex of dimension d).

In other words:

$$\mathbb{P}_{\theta_{SA}}(\ Z = z\ ) = \prod_{\underline{j} \in R_X} \theta_{\underline{j}}^{z_{\underline{j}}} \qquad z \in R_Z. \tag{3}$$

Let us now consider an i.i.d. sampling of (2):

$$Z_{(i)} \sim\ ind\ GBe_{(d+1)}(\theta_{SA}) \qquad i = 1, \ldots, n,$$

the sufficient statistic is the sum of the data:

$$Y = \sum_{1 \le i \le n} Z_{(i)} \quad \text{with} \quad Y = (Y_{\underline{j}}; \underline{j} \in R_X) \quad Y_{\underline{j}} \in \{0, \ldots, n\} \quad \sum_{\underline{j} \in R_X} Y_{\underline{j}} = n.$$

Thus, the data Y on K ordinal variables may be viewed as a K-dimensional contingency table distributed as a multinomial distribution:

$$Y \sim MN_{(d+1)}(n, \theta_{SA}).$$

2.2 Discretization

Now we are going to model the ordering property of the categorical variables using a discretization hypothesis.

Each ordered categorical variable is interpreted as a discretization of a corresponding continuous variable; more specifically for each X_k there is posited a continuous latent random variable (*ideally measured variable*) X_k^*, and a vector of thresholds

$$a^{(k)} = (a_j^{(k)} : -\infty = a_0^{(k)} < a_1^{(k)} < \ldots < a_{r_k}^{(k)} = \infty), \tag{4}$$

such that:

$$\forall\ j \in \{1, \ldots, r_k\} \qquad X_k \le j \Leftrightarrow X_k^* \le a_j^{(k)}. \tag{5}$$

For the vector $X^* = (X_1^*, \ldots, X_K^*)$ we introduce the array:

$$\mathcal{A} = \{a^{(k)} : k = 1, \ldots, K\}. \tag{6}$$

For this array we define a vector of thresholds corresponding to each $\underline{j} \in R_X$, as follows:

$$a_{(\underline{j})} = (a_{j_1}^{(1)}, \ldots, a_{j_K}^{(K)}). \tag{7}$$

The *Discretization model on the latent variables* is the statistical model obtained by marginalizing, on the observable variables, the distribution of the latent variables, namely:

$$\forall\ \underline{j} \in R_X : \mathbb{P}_{\theta_{DL}}(\ X \le \underline{j}\ ) = \mathbb{P}\,(\ X^* \le a_{(\underline{j})}\ ), \tag{8}$$

where " $\leq$ " is the coordinate-wise order and the subscript DL stands for "Discretized Latent". The array $\mathcal{A}$ operates a decomposition of $\mathbb{R}^K$ into $\prod_{k=1}^K r_k = d+1$ cubes:

$$c_{\underline{j}} = c_{j_1,\ldots,j_K} = \prod_{k=1}^{K} (a_{j_k-1}^{(k)}, a_{j_k}^{(k)}].$$

Therefore,

$$\mathbb{P}_{\theta_{DL}}(\ X = \underline{j}\ ) = F(c_{\underline{j}}), \tag{9}$$

with θ_{DL} the parameterization given by:

$$\theta_{DL} = (F, \mathcal{A}) \in \Theta_{DL}, \tag{10}$$

where F is the multivariate probability distribution of the ideally measured variables X^*, $\mathcal{A}$ gathers the thresholds as given in (6) and Θ_{DL} is the parameter space for this parametrization.

2.3 A first identification problem

The correspondence between the parametrization of the saturated model and that of (10) is accordingly given by:

$$F(c_{\underline{j}}) = \theta_{\underline{j}} \tag{11}$$

Let us notice a first identification problem in model (8). Let G be the group of continuous strictly increasing functions $g : \mathbb{R} \mapsto \mathbb{R}$, $G_{(K)}$ be the group of coordinate-wise transformations defined as: $\underline{g} = (g_1, \ldots, g_K)$ with $g_j \in G$ and define the corresponding transformation of θ_{DL}:

$$\begin{aligned} \mathcal{A}_{\underline{g}} &= \{g_j(a_j^{(k)}) : j = 1, \ldots r_k - 1,\ k = 1, \ldots, K\}, \\ F_{\underline{g}} &= F \circ \underline{g}^{-1}, \\ \theta_{DL,\underline{g}} &= (F_{\underline{g}}, \mathcal{A}_{\underline{g}}), \end{aligned}$$

then,

$$\mathbb{P}_{\theta_{DL}} = \mathbb{P}_{\theta_{DL,\underline{g}}}. \tag{12}$$

In particular, the marginal distributions $F_{j,\underline{g}} = F_j \circ g_j^{-1}$ are not identified.

As a consequence, a measure of association for data generated by the discretization model (5)-(8) should not depend on the marginal distributions of the latent variables X_j^*. A natural measure of association is the Spearman

coefficient, which is the Pearson correlation among the corresponding latent variables transformed by their own distribution function:

$$\rho_S(X_i, X_j) = \text{Corr } (F_i(X_i^*), F_j(X_j^*)),$$

where Corr (,) stands for the Pearson correlation. Remember that, by the integral transform theorem, $F_j(X_j^*)$ follows a uniform distribution on $[0, 1]$. Furthermore, in case of a finite sample on observable variables U and V, Corr $(F_U(U), F_V(V))$ may be estimated, non parametrically, by plugging the empirical marginal distributions; this estimator is the rank correlation.

In the discretization model (5)-(8), the marginal distribution functions of the latent variables, F_k, are not identified and can therefore not be meaningfully estimated. Thus an operational alternative to the unestimable Corr $(F_i(X_i^*), F_j(X_j^*))$ could be Corr (X_i^*, X_j^*), this is precisely the polychoric correlation. Hence next definition:

Definition 2.1. The *matrix of polychoric correlations* is defined as the $k \times k$ correlation matrix of the continuous real valued random variables $\{X_k^* : k = 1, \ldots, K\}$.

$$\Gamma = (\gamma_{ij}) \qquad \text{where } \gamma_{ij} = \text{Corr } (X_i^*, X_j^*) \tag{13}$$

This concept is clearly not invariant under strictly increasing (non linear) transformations of X_i^*, in spite of the identification problem raised in (12). In next section we consider the identification of Γ under a normality assumption for the joint distribution of latent vector $(X_1^*, \ldots, X_K^*)$. Thereafter we examine the role of the normality assumption from a copula point of view.

3 Identifiability under Normality

The concept of polychoric correlations is most natural under the hypothesis:

$$X^* \sim N(\mu, \Sigma), \tag{14}$$

or, in terms developed in the above section, $F = N(\mu, \Sigma)$. In order to preserve the normality, we consider the set of affine transformations included in $G_{(K)}$; these are the transformations such as:

$$x \in \mathbb{R}^K \mapsto g(x) = Bx + c,$$

with $B = Diag\{b_1, \ldots, b_K\}, b_k > 0$ and $c \in \mathbb{R}^K$. This group is denoted by $G_{(LK)}$ where L refers to the "Linear" feature of these transformations.

According to the above section, the array $\mathcal{A}$ is transformed into $\mathcal{A}_{B,c} = \{b_k a_j^{(k)} + c_k : j = 1, \ldots r_k - 1, \ k = 1, \ldots, K\}$. Then $(N(\mu, \Sigma), \mathcal{A})$ and

$(N(B\mu + c, B'\Sigma B), \mathcal{A}_{B,c})$, are observationally equivalent. This identification problem is solved by fixing $b_k = [Var(X_k^*)]^{-\frac{1}{2}}$ and $c_k = -\mathbb{E}(X_k^*)$ for $k = 1, \ldots, K$. Under $X_k^* \sim N(0,1)$, the parameter in (10) is reduced to:

$$\theta_{NPO} = (\Gamma, \mathcal{A}) \in \Theta_{NPO}, \tag{15}$$

where Γ is a correlation matrix, $\mathcal{A}$ is an array giving the thresholds and NPO stands for "Normal Polychoric". The matrix of polychoric correlations is precisely the matrix Γ. The model (9) now becomes:

$$\mathbb{P}_{\theta_{NPO}}(\ X = \underline{j}\ ) = \Phi_\Gamma(c_{\underline{j}}), \tag{16}$$

where Φ_Γ is the multivariate normal distribution with zero mean, unit variance and Γ correlation matrix.

The dimension of the parameter space in (15) is equal to:

$$\text{Dim } \Theta_{NPO} = \frac{K(K-1)}{2} + \sum_{k=1}^{K}(r_k - 1), \tag{17}$$

(we use the usual notation where Dim $\mathcal{C}$ stands for the dimension of the smallest affine space containing the parameter space indexing $\mathcal{C}$).

Provided that $min\{K, r_1, \ldots, r_K\} \geq 2$, we have that:

$$\text{Dim } \Theta_{NPO} = \frac{K(K-1)}{2} + \sum_{k=1}^{K}(r_k - 1) \leq \prod_{k=1}^{K} r_k - 1 = \text{Dim } \mathcal{S}_d, \tag{18}$$

with the equality if and only if $K = r_1 = r_2 = 2$.

As θ_{SA}, the parameter of the saturated model, is obviously identified and a smooth function of θ_{NPO}, condition (18) says that a necessary condition of identification of θ_{NPO} is always satisfied. A complete characterization of the identifiability of this model is given by the following result:

Theorem 3.1. *Under the normality hypothesis, $\theta_{NPO} = (\Gamma, \mathcal{A})$ in (15) is identified, or equivalently, the mapping $\theta_{NPO} \mapsto \mathbb{P}_{\theta_{NPO}}$, defined in (16), is one-to-one, if the polychoric correlations matrix Γ is not singular.* ■

A proof of this theorem is given in the appendix. We also have the following corollary:

Corollary 3.2. *Under the hypothesis of theorem 3.1, θ_{NPO} is just identified if $K = r_1 = r_2 = 2$; otherwise, the model is overidentified.* ■

This corollary provides possibilities of statistical testing for the restrictions implied, on the saturated model, by the discretization model. These tests have also been considered as tests of normality. However a unique test of the joint normality becomes rapidly unmanageable when K or d is increasing. A simple second best is to test only bivariate normalities; these are the procedures programmed in several packages such as LISREL or EQS; alternative procedures are also available, as for instance in Muthén and Hofacker(1988).

4 Copula Specification

4.1 Copula: a short overview

Once it is recognized that the marginal distributions of the latent variables are not identified in the discretization model, it is natural to decompose the characterization of the multivariate distribution of the latent variables into, on the one hand, the set of its marginal distributions and, on the other hand, a residual aspect that would be variation-free with respect to the first one and would capture the association between the latent variables. This is exactly the idea of a copula, a precise definition and characterization of which are now given.

Definition 4.1. A copula C is the distribution function of a multivariate probability distribution with uniform margins on the unit interval. ■

We need for later use two main results about copulas, the first one is a simple application of the integral transform theorem and the other one is its reciprocal affirmation. Hereafter we use C or F indifferently as probability measure or as distribution function if there is not ambiguity. The proof of these two theorems, along with a detailed study of copulas may be found in Nelsen (1999).

Theorem 4.2. *Let $F_1, \ldots, F_K$ be K univariate distributions and C a K-dimensional copula. Then $F(x_1, \ldots, x_K) = C(F_1(x_1), \ldots, F_K(x_K))$ defines a multivariate distribution F with margins $F_1, \ldots, F_K$.* ■

Theorem 4.3. *(Sklar's theorem) Let F be a K-dimensional distribution with continuous margins $F_1, \ldots, F_K$. Then F has a unique copula representation*

$$F(x_1, \ldots, x_K) = C(F_1(x_1), \ldots, F_K(x_K)).$$

■

The last two theorems give us a bijection between the set of K-dimensional multivariate distributions and the set of K univariate distributions and K-dimensional copulas. Thus, with the copula concept, we decompose the marginal and the joint aspects of a multivariate distribution. A copula of a particular interest is the following:

Definition 4.4. For any correlation matrix Γ the corresponding gaussian copula is defined by:

$$C_\Gamma^G(u_1,\ldots,u_K) = \Phi_\Gamma(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_K)) \quad (u_1,\ldots,u_K) \in [0,1]^K$$

where Φ is the standard normal univariate distribution function. ■

Thus a multivariate normal distribution may be viewed as a gaussian copula along with normally distributed margins.

In the present context an important property of copulas is their invariance to the group of strictly increasing coordinate-wise transformations; more precisely the following result is immediate:

Proposition 4.5. *If F is a multivariate distribution on $\mathbb{R}^K$ and $g \in G_{(K)}$, then F and $F \circ g^{-1}$ have the same copula. More specifically, if*

$$F(x_1,\ldots,x_K) = C(F_1(x_1),\ldots,F_K(x_K)),$$

then for any $g \in G_{(K)}$:

$$F \circ g^{-1}(y_1,\ldots,y_K) = C(F_1 \circ g_1^{-1}(y_1),\ldots,F_K \circ g_K^{-1}(y_K))$$

■

4.2 A copula approach to the discretization model.

Using Sklar's theorem, the discretization model (5)-(8) can be parametrized, instead of (10), as follows:

$$\theta_{DLC} = (\{F_k : k = 1,\ldots,K\}, C, \mathcal{A}), \tag{19}$$

where the subscript DLC stands for "Discrete Latent Copula" and C represents the unique copula (in the form of distribution function) such that:

$$F(x_1,\ldots,x_K) = C(F_1(x_1),\ldots,F_K(x_K)).$$

As the margins of a copula are uniformly distributed on $[0,1]$, the corresponding threshold values are also included in $[0,1]$; the thresholds corresponding to a copula are accordingly denoted as $p_j^{(k)}$ (instead of $a_j^{(k)}$). The model (5)-(8) may now be rewritten as:

$$\forall\ \underline{j} \in R_X \quad X \leq \underline{j} \Leftrightarrow \{X_k^* \leq p_{j_k}^{(k)}, \quad k = 1, \ldots, K\}, \tag{20}$$

and therefore:

$$\mathbb{P}(\ X \leq \underline{j}\ ) = C(p_{j_k}^{(k)} : 1 \leq k \leq K). \tag{21}$$

Thus the non-identifiability of the margins F_k implies that $\theta_{CO} = (C, \mathcal{P})$, where the subscript CO stands for "Copula", is a sufficient parametrization, instead of (19). Therefore (6) may be reparametrized into:

$$\mathcal{P} = \{p_j^{(k)} : j = 1, \ldots, r_k - 1,\ k = 1, \ldots, K\}$$

through:

$$p_j^{(k)} = F_k(a_j^{(k)}) \in [0,1]. \tag{22}$$

Note that $p_j^{(k)}$ may be viewed as the probability that the ordinal variable takes a value equal than or inferior to j, and from (22), is such that $a_j^{(k)}$ corresponds to the $p_j^{(k)}$-quantile of the unidentified marginal distribution F_k.

Furthermore, the threshold parameters $p_j^{(k)}$ are defined independently of the copulas. Therefore $p_j^{(k)}$ may be unbiasedly and consistently estimated by the sample proportions whereas $a_j^{(k)}$ may be consistently, but not unbiasedly (except in very particular cases), estimated but only relatively to an arbitrary specification of F_k.

When the family $\mathcal{C}$ of copulas is finitely parametrized, the necessary condition of identification (18) now becomes:

$$\text{Dim } \mathcal{C} \leq \prod_{k=1}^{K} r_k - 1 - \sum_{k=1}^{K} (r_k - 1),$$

where $\mathcal{C}$ is a subset of K-dimensional copulas' space defining the model.

In the case of a gaussian copula, the model (5)-(8) becomes:

$$\mathbb{P}_{\theta_{NCO}}(\ X \leq \underline{j}\ ) = C_\Gamma^G(p_{j_k}^{(k)} : 1 \leq k \leq K), \tag{23}$$

the parametrization of which becomes $\theta_{NCO} = (\Gamma, \mathcal{P})$, where the subscript NCO stands for "Normal Polychoric". As the parametrization θ_{NCO} is clearly in bijection of θ_{NPO}, the conditions of application of theorem 3.1 remain the same.

5 Conclusions

The usual definition of polychoric correlations, for ordinal variables, is based on a multivariate normal hypothesis. This hypothesis has been decomposed into two variation-free components: a gaussian copula and standard normal margins. On the one hand, the hypothesis on the margins is not identified, therefore not testable, and is justified on two grounds. Firstly to give substance to an association measured by means of Pearson's correlation and secondly by a structural model positing linear models based on the (unidentified) choice of normally distributed coordinates of the latent variable. On the other hand the gaussianity of the copula is an identified and testable hypothesis.

One basic object of identification is to check the empirical meaning of the parameters in a structural model: an unidentified parameter may not be interpreted as the expectation or as the probability limit of some statistic. The copula specification (20)-(21) endows the threshold values $p_j^{(k)}$ with a simple interpretation of the expected value (or probability limit) of a sample proportion. Note however that such an interpretation does not imply that the marginal distributions of the latent variables are uniform on $[0, 1]$: it only refers to the always true (for continuous distribution) and therefore unrestrictive fact that the latent variables, transformed by their own distribution functions, are uniformly distributed on $[0, 1]$.

Furthermore, the threshold values $p_j^{(k)}$ are easily estimated, unbiasedly and consistently, by the sample proportions without requiring arbitrary specifications of the marginal distributions F_k. This is different from the array $\mathcal{A}$ where $a_j^{(k)}$ can be interpreted relatively to an arbitrary specification of F_k only.

As already mentioned, the polychoric correlation is based on a contextual rather than a precise empirical consideration, leaving therefore open the question of how to measure the association among ordinal variables. There is a considerable literature, particularly in social sciences, concerning this topic in general. However, in the framework of the discretization model (5)-(8), it should be clear that any intrinsic measure should be based on the copula of the latent variables, rather than on their (complete) joint distribution. How to estimate an intrinsic measure of association based on the copula of the latent variables is a topic that deserves future work.

Proof of theorem 3.1

Proof. From (15), we need to prove:

$$\mathbb{P}_{\Gamma,\mathcal{A}} = \mathbb{P}_{\tilde{\Gamma},\tilde{\mathcal{A}}} \qquad \Longrightarrow \qquad (\Gamma, \mathcal{A}) = (\tilde{\Gamma}, \tilde{\mathcal{A}}).$$

The proof will be split in two parts:

(i) $\mathbb{P}_{\Gamma,\mathcal{A}} = \mathbb{P}_{\tilde{\Gamma},\tilde{\mathcal{A}}} \qquad \Longrightarrow \qquad \mathcal{A} = \tilde{\mathcal{A}}.$

The assumption $\mathbb{P}_{\Gamma,\mathcal{A}} = \mathbb{P}_{\tilde{\Gamma},\tilde{\mathcal{A}}}$ is by definition equivalent to:

$$\forall\, \underline{j} \in R_X \qquad \mathbb{P}_{\Gamma,\mathcal{A}}(\, X \leq \underline{j}\,) = \mathbb{P}_{\tilde{\Gamma},\tilde{\mathcal{A}}}(\, X \leq \underline{j}\,).$$

Let us consider $a_j^{(k)} \neq \tilde{a}_j^{(k)}$ for some $k = 1, \ldots, K \quad j = 1, \ldots, r_k$.

Let $\underline{j} = (r_1, \ldots, r_{k-1}, j, r_{k+1}, \ldots, r_K)$, and define $\overline{k} = \{1, \ldots, K\} \setminus \{k\}$.

By the injectivity of Φ, the normal standard distribution function, we obtain for any value of Γ and $\tilde{\Gamma}$:

$$\begin{aligned}\mathbb{P}_{\Gamma,\mathcal{A}}(X \leq \underline{j}) &= \mathbb{P}(X_\ell^* < \infty,\ \ell \in \overline{k}, X_k^* \leq a_j^{(k)}) = \Phi(a_j^{(k)}) \\ &\neq \mathbb{P}_{\tilde{\Gamma},\tilde{\mathcal{A}}}(X \leq \underline{j}) = \mathbb{P}(X_\ell^* < \infty,\ \ell \in \overline{k}, X_k^* \leq \tilde{a}_j^{(k)}) = \Phi(\tilde{a}_j^{(k)}).\end{aligned}$$

Therefore, $\mathcal{A}$ is identified.

(ii) $\mathbb{P}_{\Gamma,\mathcal{A}} = \mathbb{P}_{\tilde{\Gamma},\mathcal{A}} \qquad \Longrightarrow \qquad \Gamma = \tilde{\Gamma}.$

As $|\Gamma| \neq 0$, one has $\quad |\gamma_{k_2 k_2}| \neq 1 \quad \forall\, k_1 < k_2$. Let $\gamma_{\ell k} \neq \tilde{\gamma}_{\ell k}$ and consider $\underline{j} = (r_1, , \ldots, r_{\ell-1}, i, r_{\ell+1}, \ldots, \ldots, r_{k-1}, j, r_{k+1}, \ldots, r_K)$; then, we have that:

$$\begin{aligned}\mathbb{P}_{\Gamma,\mathcal{A}}(X \leq \underline{j}) = \mathbb{P}(\ & X_m^* < \infty, m \in \{1, \ldots, K\} \setminus \{\ell, k\}, \\ & X_\ell^* \leq a_i^{(\ell)}, X_k^* \leq a_j^{(k)}) = \Phi_{\Gamma_{\ell k}}(a_i^{(\ell)}, a_j^{(k)}),\end{aligned}$$

where $\Phi_{\Gamma_{\ell k}}$ is the bivariate normal distribution function with mean zero and variance matrix $\Gamma_{\ell k}$, with

$$\Gamma_{\ell k} = \begin{pmatrix} 1 & \gamma_{\ell k} \\ \gamma_{\ell k} & 1 \end{pmatrix}$$

For all $\gamma_{\ell k} \in]-1, 1[$ and $(a, b) \in \mathbb{R}^2$; one has:

$$\frac{\partial}{\partial \gamma_{\ell k}} \Phi_{\Gamma_{\ell k}}(a, b) = \varphi_{\Gamma_{\ell k}}(a, b) > 0$$

where $\varphi_{\Gamma_{\ell,k}}$ is the corresponding bivariate normal density function with the mean and variance matrix as above. (see Johnson and Kotz (1972) or Tallis (1962)). Thus, for all $(a, b) \in \mathbb{R}^2$, the mapping $\gamma_{\ell k} \mapsto \Phi_{\Gamma_{\ell k}}(a, b)$ is a strictly increasing continuous function, so it is injective.

Then by the injectivity of this function:

$$\mathbb{P}_{\Gamma,\mathcal{A}}(X \leq \underline{j}) \neq \mathbb{P}_{\tilde{\Gamma},\mathcal{A}}(X \leq \underline{j})$$

So Γ is identified.

Finally (i) and (ii) prove the theorem. ■

References

Johnson, N. L. and Kotz, S. (1972). *Distributions in statistics: Continuous multivariate distributions* Wiley: New York.

Jöreskog, K. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix, *Psychometrika* **59**, 381–389.

Jöreskog, K., Sörbom, D., du Toit, S. and du Toit, M. (2002). *LISREL8: New Statistical Features.* SSI Scientific International Inc.

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

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

Muthén, B. and Hofacker, Ch. (1988). Testing the assumptions underlying tetrachoric correlations, *Psychometrika*, **83**, 563–578.

Nelsen, Roger B. (1999). *An introduction to copulas. Properties and applications.* Lecture Notes in Statistics (Springer) **139**.

Olsson, Ulf (1979). Maximum likelihood of the polychoric correlation coefficient, *Psychometrika* **44**, 443–460.

Poon, Wai-Yin and Lee, Sik-Yum (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients, *Psychometrika* **52**, 409–430.

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