International Conference

The many dimensions of poverty

Brasilia, Brazil – 29-31 August 2005
Carlton Hotel

Methods of Factor Analysis for Ordinal Categorical Data and Application to Poverty Data

Conference paper

Gisele Kamanou
United Nations Statistics Division, United States of America

EMBARGO

This text is on embargo until 29 August

Preliminary draft – August 2005

Methods of Factor Analysis for Ordinal Categorical Data and Application to Poverty Data

By Gisele Kamanou
Statistician, United Nations Statistics Division, New York

Abstract

The Pearson correlation has been the conventional measure of dispersion matrix used in empirical factor analysis. It implies assumptions of normality of variables that are unlikely to hold for multidimensional analysis of social phenomena such as poverty. This paper reviews the basic essentials of factor analysis for ordinal and categorical data and proposes a framework for the specification of a factor analysis model for poverty data. Key features of factor analysis are recalled to bring more clarity in the interpretation of the results of a few data examples.

1. Introduction

The study of multidimensional poverty involves the joint analysis of several variables. Making decisions based on more than one variable, however, poses some practical challenges such as those found in formulating policies to address an underlying characteristic that cannot be measured. Like most socio-economic manifestations, poverty is not directly measurable, though one would have little difficulty to recognize it. In studying multidimensional poverty in particular, one has to resort to several variables to attempt to describe and quantify its size and shape. It is therefore conceivable to wanting to know if there is/are one or more underlying (theoretical) variable(s) that would explain the measures taken on the multiple variables. The unobservable – but estimated variable(s) could reasonably be used to measure poverty. Furthermore, as the number of variables gets larger, simply descriptive analysis such as ranking, and cross tabulations or groupings become too cumbersome or have no clear practical meaning. In these situations and, provided a minimal loss of useful information, it is a practical need to reduce the data to a lower-dimensional space for convenience in the analysis and in the interpretation of the data. These two problems/objectives address two different philosophical issues whereby the first model of the unobserved variable is known as factor analysis and the data reduction model in achieved through the method of principal component analysis. However, factor analysis and principal component analysis are essentially the same in practice, whilst the difference between the two methods has been subjected to considerable debates.

Multidimensional scaling is a family of tools used for transforming a set of points in a high-dimensional space to a lower-dimensional one while minimizing the loss of information in the data. There exist a wide range of techniques used in Multidimensional scaling among which principal components analysis and factor analysis (to a lesser degree) are the most widely used by social scientist researchers. These two techniques are in principle applicable exclusively to real valued continuous data notably since the model specification of these techniques basically amounts to the Single Value Decomposition[1] (SVD) of the multiple-correlation matrix of the variables. Many variables used in the analysis of multidimensional poverty, however are non -normal. In many situations and for non-monetary variables in particular, variables are categorical or ordinal and often take values within a small range of discrete categories. In these situations, the contingency table of the variables is used in lieu of the correlation matrix and the assumptions of the factor analysis model based on the Pearson correlation matrix of the variables would be violated.

There is a vast literature of the factor analysis with non-normal data which can be grouped under two broad lines of reasoning. On the one hand, it is claimed that the parameter estimates of the standard model are robust enough to sustain non-normality

[1] SDV: Any real matrix, A, has a unique Single Value Decomposition (SVD), which consists of three matrices, U, $\sum$ and V, whose product is the original matrix. The first of these, U, is composed of orthogonal columns known as the Left Singular Vectors, and the last, V, is composed of orthogonal rows known as the Right Singular Vectors. $\sum$ is diagonal and contains the singular values. The singular vectors reflect principal components of A, and each pair has a corresponding value, the magnitude of which is a function of the variance accounted for by the vector. If A is symmetric and positive semi-definite, the left and right singular vectors will be identical and equivalent to its eigenvectors, and the singular values will be its eigenvalues.

(REF later). On the other hand, researchers have provided a large range of theoretical justifications for alternative model specifications in the presence of non-normality (REF later).

The remainder of the paper is organized in 3 sections. Section 2 highlights the generalities and conceptual differences between factor analysis and principal components analysis and presents a hypothetical formulation of factor analysis of multidimensional poverty. Section 3 discusses the limits of the conventional factor analysis for categorical and/or ordered data. Alternative methods to adapt factor analysis to non-normal data proposed in the recent literature are broadly discussed. To fix ideas and to show relevance to the analysis of multidimensional poverty, a solution proposed by Pr. Bartholomew (1980) for factor analysis when all the variables are measured on an ordered categorical scale is reviewed with some mathematical details. Section 4 gives a more specific account of the data restrictions of multidimensional poverty data. Two alternative methods for fitting non-normal data to a factor model are discussed. First, a model for ordinal poverty data is formulated within the logic of the logit model proposed by Bartholmew. Second, the data are transformed in an attempt to achieve normality and are used in a standard factor model. The two approaches are compared on the basis of few data examples. The paper concludes with some practical suggestions for empirical work.

2. Methods of factor analysis and principal component analysis: an overview

Factor analysis and principal component analysis are two commonly used methods to obtain a reduced-rank representation of a set of observed variables. Both methods attempt to approximate a set of points in a higher dimensional space with another set of points in a lower dimensional space while minimizing the loss of information. The information provided by a sample of observations can be viewed according to two broad categories of summary measured – the variances of the variables or the colinearity among the variables. Whether low-dimensional representation aims at retaining the maximum of variance or aims at fully accounting for the multicolinearity among the original variables is the basis of the difference between the two techniques.

In algebraic terms, both factor analysis and principal component analysis start with a set of n real valued vectors of dimensionality p and the output is a set of n vectors with dimensionality q where q is smaller than p. The goal is for the relative distance between the final vectors to reflect the relative distances between the initial vectors as closely as possible. The mathematical formulation of this problem is thus function of the distance metric defined for measuring the pairwise distances in the original and final spaces, which can be either in terms of variance of the variables or in terms of their covariances.

In addition to the instrumental choice of the distance measures to be compared between the initial space and its lower representation, the theoretical foundations underlying the two methods are very different (see for example: *Journal of Consumer* Psychology (2001), Velicer (1990), Widaman (1993)). Here we recall only the basic essentials of the two techniques to help frame the discussion in the subsequent sections on the use and interpretation of factor analysis in empirical analysis of multidimensional poverty.

One of the fundamental differences between principal component analysis and factor analysis is that principal component analysis is a purely analytical tool designed to reduce the dimension of the data without a prior supposition on the structure of the data or on the relationship among the variables. It is worth recalling the origin of principal component analysis to fit a subspace to a set of points in a higher dimensional space (Pearson 1901). Given n points in a p dimensional space, denoted by the random matrix $X = \begin{pmatrix} x_{11}\cdots & x_{1p} \\ x_{n1}\cdots & x_{np} \end{pmatrix}$, and a $(p \times p)$ dispersion matrix (covariance) Σ, assumed to be non-negative: $\Sigma = (\sigma_{ij})$, $\sigma_{ij} = \sum_{r=1}^{n}(x_{ir} - \bar{x}_i)(x_{jr} - \bar{x}_j) = \sum_{r=1}^{n} x_{ir}x_{jr} - n\bar{x}_i\bar{x}_j$, $\bar{x}_i = \frac{\sum_{r=1}^{n} x_{ir}}{n}$.

Principal component aimed at transforming a set of points to a lower dimensional space can be formulated by finding the linear transformation of $Y = T^{'}X$ where T is a $(p \times q)$ matrix so that Y is q-dimensional such that X can be reasonably approximated by Y. For practicality, and without loss of generality, the columns vectors $T_1, T_2, ..., T_q$ are assumed to independent, implying that the transformed variables are uncorrelated. The predictive efficiency of Y for X then depends on the residual dispersion matrix of X after subtracting its best linear predictor in terms of Y. It is given that the smaller the values of the elements of the residual dispersion matrix $\Sigma - \Sigma T(T^{'}\Sigma T)^{-1}T^{'}\Sigma$, the greater is the predictive efficiency. Using the Euclidean norm of the residual dispersion matrix

$$\left\| \Sigma - \Sigma T(T^{'} \Sigma T)^{-1} T^{'} \Sigma \right\| \qquad (2.1)$$

it follows from well established results of canonical reduction of matrices that the minimum of (2.1) is attained when $\Sigma T(T^{'} \Sigma T)^{-1} T^{'} \Sigma = \lambda_1 P_1 + \lambda_2 P_2 + ... + \lambda_q P_q$, where $\lambda_1, \lambda_2, ..., \lambda_q$ and $P_1, P_2, ..., P_q$ are the first q largest eigen values and associated eigen vectors of the matrix Σ respectively and that (2.1) holds when the i^th column vector of T is the vector associated with the i^th eigen values of Σ, that is $T_i^* = P_i$ (see for example in text books on multivariate analysis). The transformed variables $P_1^{'} X, P_2^{'}, ..., P_q^{'}$ are called the first q principal components of the random variables X.

In a more general form component analysis will be defined as any decomposition of the covariance matrix Σ. In principal component however, all variables are transformed to standard scores and the covariance matrix is now the correlation matrix R. The most basic interpretation of principal component follows from the following results: Under complete orthogonal transformation (e.g. when the original variables are orthogonal) from a p-dimensional space to another p-dimensional space, the sum of the variances of the variables, i.e. the trace Σ, is invariant. That is,

$$\text{trace } \Sigma = \text{trace } T^{'} \Sigma T = T_1^{'} \Sigma T_1 + T_2^{'} \Sigma T_2 + ... + T_p^{'} \Sigma T_p = \lambda_1 + \lambda_2 + ... + \lambda_p.$$

It then follows that under the q-dimensional transformation $T^{(q)}$ trace Σ is approximated by trace $T^{(q)'} \Sigma T^{(q)} = T_1^{'} \Sigma T_1 + T_2^{'} \Sigma T_2 + ... + T_q^{'} \Sigma T_q \leq trace \Sigma$. The first

q principal components are said to explain $100 \times (\lambda_1 + \lambda_2 + ... + \lambda_q)/(\lambda_1 + \lambda_2 + ... + \lambda_p)$ percent of the total variance.

In contrast to principal component analysis, factor analysis was developed to address a measurement issue (*Journal of Consumer Psychology* 2001). There has been considerable philosophical discussion on the essence of factor analysis variant theories in the literature. In most cases, and using the terms in a broad sense, a factor model posits the existence of one or more unobservable variables of interest (sometimes called the latent variables) that are empirically measured only through multiple, say p, manifest variables taken jointly. The measurements (on the p manifest variables) taken on a number of subjects are then used to make the best estimation of the variables of interest (the latent variables) via the methods of factor analysis. The basic assumption in factor model is that the co-variability that is common to all measures is attributed to the underlying latent variables, also called common factors. Although the number of latent variables (say q) conceivably can be of any value, the model is useful in practice only if q is smaller than p. It is often debated as to whether latent variables are real or not, but they are mathematical constructs which can be used to simplify nature of multidimensional poverty data.

The nature of the approximation is similar to the one for principal component but there are some important conceptual differences in factor analysis. The aim is to investigate whether p random variables $X_1, X_2, ...X_p$, with dispersion matrix Σ, can be approximated by linear functions of as few (dominant) factors as possible, say q factors. The basic foundation of the model is to conceive that there exist uncorrelated latent variables $F_1, F_2, \ldots$, presumably infinite in number, such that

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \ldots \\ X_p = a_{p1}F_p + a_{p2}F_2 + \ldots \end{cases}, \text{ in other words } \quad X = AF \qquad (2.2)$$

It is seen that the number of observed variables p is smaller than the number of factors, and unlike principal component analysis, there is no unique solution to the problem. This problem is called factor indeterminacy and has been widely discussed in the literature (Velicer F. and Jackson D (1990), more references later). An advantage of the factor model compared to the component model that follows from the indeterminacy problem is that factor analysis is assumed to yield the same factors whereas component analysis is characterized as unstable under sampling variables.

In the simplest form of factor analysis, equation (2.2) is truncated at the first q-th factors though to have the most dominant effect in explaining the collineariry among the observed variables. Thus, factor model will aim at approximating X by $X^{(q)} = A_q F^{(q)}$ where A_q is a matrix of the first q vectors of A in (2.2) and $F^{(q)}$ is the vector of the variables (factors) $F_1, F_2, \ldots, F_q$. Let us assume, without loss of generality, that the F_i $i = 1,2,\ldots$ have unit variance. Then the dispersion matrix of $X^{(q)}$ is $A_q A_q'$. Further, if the measure of the closeness between X and $X^{(q)}$ is defined by the Euclidean, i.e. $\left\| \Sigma - A_q A_q' \right\|$, it turns out that factor analysis leads to principal component analysis, and the parameters will be estimated by a similar method discussed above to give that $A_q^{'} = (\sqrt{\lambda_1}P_1, \sqrt{\lambda_2}P_2, ..., \sqrt{\lambda_q}P_q)$, i.e. the ith column of A_q is $\sqrt{\lambda_i}P_i$, where λ_i is the ith eigen value and P_i the associated eigen vector of $\sum$.

In a more general case of factor model, however, the measure of closeness is the amount of covariance between X_1, X_j explained by the q factors $F_1, F_2, ..., F_q$. Further the approximation X by $X^{(q)}$ assumes some error terms such that equation (2.2) can be written as:

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + ... + a_{1q}F_q + e_1 \\ X_p = a_{p1}F_p + a_{p2}F_2 + ... + a_{pq}F_q + e_p \end{cases} \tag{2.3}$$

$$\text{that is } X = AF^{(q)} + \mathrm{E} \tag{2.4}$$

A is called the matrix of factors loadings. It follows from (2.4) that $\sum = A\Phi A^{'} + \Psi$ or, $\sum - \Psi = A\Phi A^{'}$, where Φ and Ψ are the covariance matrix of $F^{(q)}$ and of residuals E respectively. The general formulation of the factor model is to approximate the covariance matrix in which a diagonal matrix of estimated unique variances (Ψ) is subtracted by a matrix of reproduced correlations, that is:

$$\sum - \hat{\Psi} \cong A\Phi A^{'} = \sum\nolimits_f^{*} \tag{2.5}$$

Because the rational in factor analysis is to express the variance shared among the p observed variables, the diagonal elements of $\sum$ now represent the communality between the variables, often measured (or estimated) by the squared correlation for each variable being predicted by all other (p-1) variables. The j[th] diagonal element of $\sum_f^*$ is thus the common variance of the j[th] variable that is presented by the factor solution.

The optimum choice for $\sum_f^*$ will be such that the off diagonal elements of correlation matrix of the residual $\sum - \sum_f^*$ are as small as possible and are comparable with the sampling errors of the original correlations. The estimates of the coefficients $A\Phi A^{'}$ are not trivial in practice and it is often assumed that $F_1, F_2, ..., F_q$ and $e_1, e_2, ..., e_q$ are independent from one another. It further can be assumed without loss of generality that the factors have unit variance and thus $\Phi = \mathrm{I}$ (the identity matrix). There are more unknowns than equations in the system of equation in (2.3) and the solution can only be derived by an iterative method. Jowett (1958) provides a simple illustrative example of the estimation involving one factor model. The procedure starts with a correlation matrix of the observed variables in which the 1's in the diagonal are replaced by the square coefficients of the factors loadings for the single factor. Equating $\sum$ with $AA^{'}$ thus gives the system of equations

$$X = \begin{pmatrix} \alpha_1^2 & \rho_{12} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & \alpha_2^2 & \rho_{23} & \cdots & \rho_{2p} \\ . & & & & \\ . & & & & \\ \rho_{p1} & \rho_{p2} & \rho_{p3} & \cdots & \alpha_p^2 \end{pmatrix} = \begin{pmatrix} \alpha_1^2 & \alpha_1\alpha_2 & \alpha_1\alpha_3 & \cdots & \alpha_1\alpha_p \\ \alpha_1\alpha_2 & \alpha_2^2 & \alpha_2\alpha_3 & \cdots & \alpha_2\alpha_p \\ . & & & & \\ . & & & & \\ \alpha_p\alpha_1 & \alpha_p\alpha_2 & \alpha_p\alpha_3 & \cdots & \alpha_p^2 \end{pmatrix}$$

Several procedures are used in practice to fit $\sum$ to A, Φ and Ψ, including the maximum likelihood (ML), the unweighted least squares (ULS) and the generalized least squares (GLS) techniques. Detailed description of these techniques and their relative advantages are documented in a wide range of text books or research work (e.g. Anderson 1959, Joreskog and Goldberg 1972). All these estimation techniques are based on a normality assumption of one sort or another. In the ML case for example, it is generally assumed that the residuals in (2.4) are normal distributed, i.e. $\mathrm{E} \sim MVN(0,\Psi)$ and that $F^{(q)} \sim MVN(0,\Phi)$ such that $X \sim MVN(0,\sum)$.

3. Fitting function and correlation matrix in factor analysis of categorical data: Pitfalls and alternatives

The most commonly used fitting method in (confirmatory) factor analysis is the Maximum Likelihood (ML) that uses the standard product-moment correlation. The limits of the Pearson correlation, however, have been discussed by many researchers. Babakus E., Ferguson C and Joreskog G. (1987) provide a critical review of the research work on the adequacy of Maximum Likelihood theory which is for the most part based on continuous data meeting normality distributional assumptions of the data. One of the themes of this work has been the (mis)use of the Pearson correlation matrix to analyze data that do not satisfy the distributional assumption. Recent research has been concerned with the ordinal data and in particular for instance when the observed variables are dichotomous.

To fix ideas, it is useful to start with a presentation of the model in a general case. Let Y be the underlying variables that we believe explain the observed variables X. The distributions f (x) and g(y) are related by

$$f(x) = \int_{\Re} \pi(x \mid y) g(y) dy \qquad (3.1)$$

where $f(x)$ is the probability mass of X and $g(y)$ is the density of Y. This is traditionally known as a latent structure model in which X in called the manifest variable and Y the latent variable thought to be continuous.

When the observations are all categorical x will identify a cell of the multi-way contingency table and $f(x)$ its multinomial probability. Under conditional independence (meaning that the observed dependence among the x's is wholly explained by their dependence on the y's), we can write:

$$f(x) = \int_{\Re} ... \int_{\Re} . \prod_{i=1}^{p} \pi_i(x_i \mid y) dy = \mathrm{E} \prod_{i=1}^{p} \pi_i(x_i \mid y) \qquad (3.2)$$

Following Bartelemew (1980), we let x's take value 0 or 1 and have that

$$\pi_i(x_i \mid y) = \{\pi_i(y)\}^{x_i} \{1 - \pi_i(y)\}^{1-x_i}, (x_i = 0,1) \qquad (3.3)$$

It is shown that π_i is the value of $\pi_i(y)$ when $y_j = \frac{1}{2}$ for all j. It therefore represents the "typical probability" of response in the i th dimension. In this case, equation (3.2) can be written as:

$$f(x) = \int_{\Re} ... \int_{\Re} \pi_i(y) ... \pi_j(y) dy \qquad (3.4)$$

In order to formulate the factor analysis model for the dichotomous case, we start by expressing the conditional distribution of y given x

$$p(y \mid x) = f(x \mid y) / f(x) = \prod_{i=1}^{p} \{\pi_i(y)\}^{x_i} \{1 - \pi_i(y)\}^{1-x_i} \qquad (3.5)$$

Bartlelemew gives an interpretation of (3.5) when y is uniform in (0,1). Under this condition, $\mathrm{E}(y_r \mid x)$ is the expected proportion of the population below an individual with observed value x, and, by definition, $\mathrm{E}(y_r \mid x)$ is the y-score of the individual on dimension r.

The y-score's can thus be estimated by evaluating the integrals

$$\int_0^1 \ldots \int_0^1 y_r . \pi_i(y) . \pi_j(y) dy \qquad (3.6)$$

The goodness of fit for this model is defined as a linear transformation of the log-likelihood $\Lambda = 2 \sum_i O_i \ln(O_i \mid E_i)$ (3.7)

where O_i and E_i are the observed and expected frequencies and the sum is over all cells of the contingency. Let Λ_0 be the value of (3.7) when the expected frequencies are calculated on the assumption of complete independence and Λ_q when they are obtained by fitting a model with q variables. The ratio $(\Lambda_o - \Lambda_q) / \Lambda_0$ represents how much departure from independence is accounted for by the model. If the method of fitting the model provides efficient parameters then Λ can be shown to have a χ^2-distribution with $(2^p - P'' - 1)$ where P' is the numbers of parameters and can be used to judge the goodness of fit of the model.

As an example, we review below the main inputs to the fitting functions based on the logit model developed by Barthelemew in Barthelemew (1980). The parameters (π_i) are estimated by an iterative approach on the basis of two core mathematical results:

First, $$\frac{R_{ij}-1}{\sigma^2}=\sum_{k=1}^{q}\alpha_{ik}\sigma_{jk}+\xi's \qquad (3.8)$$

Where $R_{ij}=\dfrac{\mathrm{E}\{\pi_i(y)\pi_j(y)\}\mathrm{E}(1-\pi_i(y))(1-\pi_j(y))}{\mathrm{E}\{\pi_{i(y)}(1-\pi_j(y))\}\mathrm{E}\{\pi_j(y)(1-\pi_i(y))\}}$ is the cross product ratio formed from the expected frequencies when the table is collapsed over all dimensions except i and j, $\sigma^2=\mathrm{E}(\log^2\{y/(1-y)\}=3.289$, and $\xi's$ are terms of 4th degree in α_{ik} and α_{jk},

Second, $$\frac{N_i}{N}=\int_0^1\frac{\pi_i y^{\alpha_i}}{\pi_i y^{\varepsilon_i}+(1-\pi_i)(1-y)^{\alpha_i}}dy,\ (i=1,2,...,p) \qquad (3.9)$$

where N_i is the number of positive responses on dimension i. The sample cross-product ration can be used to estimate R_{ij}. For the case of a one factor model $(q=1)$, expression (3.8) yields $\dfrac{R_{ij}-1}{\sigma^2\alpha_i\alpha_j}=1+\xi's$. Thus the aim in the approximation is for the ratio

$\dfrac{R_{ij}-1}{\sigma^2\alpha_i\alpha_j}$ to be close to 1.

[*Aside*: Using the logit first approximation, equation (3.9) is re-written using a normal approximation

$$\hat{\pi}_i\approx\Phi\{(1+\hat{\alpha}_i{}^2)^{\frac{1}{2}}\Phi^{-1}(N_i/N)\} \qquad (3.10)$$

The iteration procedure amounts to the following steps:

1. Find the vector of $\alpha's$ such that $\alpha_i\alpha_j$ is close as possible to the estimated value of

$c_{ij} = \frac{R_{ij} - 1}{\sigma^2}$ for $i, j = 1,2,..., p \quad i \neq j$

2. find π by equating $\mathrm{E}\pi_i(y)$, $(i = 1,2,..., p)$ to the corresponding marginal proportion using the vector α obtained in step1

3. improve the estimate of α by writing $c_{ij} = \alpha_i\alpha_j\theta_{ij}$ where θ_{ij} is assumed to have a weak link with π_i, π_j, α_i and α_j but will usually be close to 1. The new (improved) estimate of α will then be obtained by replacing the starting values $\hat{c}_{ij}$ by $\hat{c}_{ij} / \theta_{ij}$

4. repeat steps 2 and 3 until π and α (or Λ) converge.]

The issue as to which of the correlation matrix should be used as input to the factor analysis of discrete ordinal data in the general case has been widely researched subsequent to Barthelemew (1980). Several arguments consistent with those of the dichotomous case have been advanced against the Pearson correlation. More generally, discrete and ordinal data do not necessarily produce positive definite correlation matrices and factor analysis parameter estimates based on the Pearson’s correlation will be biased and model fit severely distorted (Johnson and Creech 1983 and others). In many situations in multidimensional poverty, and for non-monetary variables in particular, variables are categorical or ordinal and often take values within a small range of discrete categories. In these situations, the contingency table of the variables is used in lieu of the correlation matrix.

There are several methods for defining a correlation between two variables from a frequency table. The polychoric correlation introduced by Ritchie-Scott (1918) and Pearson and Pearson (1922), is computed from a polychoric table which is a table of more than 2X2 cells and fewer than 4X4 cells. It has been shown that the polychoric correlation coefficient, calculated from ordinal transformation of bivariate normal variables is an unbiaised estimate of the correlation between the original bivariate variables (Rigdon and Ferguson 1991, pp491). Rigdon and Ferguson confirmed the findings of the studies by (Joreskog and Sorbom1981) that the polychoric correlation coefficient is a better measure of correlation for ML factor analysis of ordinal data. They showed that, on the basis of factor loading bias and squared errors of the ML solution, the polychoric outperformed other alternative correlation measures including the Pearson's (product-moment) correlation, the Pearson's rho and the Kendall's tau.

A major drawback of the polychoric correlation, however, is that it produces poor fit statistics of the ML. Nonetheless, recent development in the work by Barbakus, Ferguson and Joreskog 1987, Joreskog and Sorbom in Joreskog and Sorbom1981 suggests that the use an unweighted least squares (ULS) solution will improve the fit statistics when the estimation is based on the polychoric.

4. Some data examples. [Under development]

In this Section we emulate the method proposed by Bartholmew to fit a logit model to dichotomous poverty data. We follow the line of more recent research and use the polychoric correlation with the ULS. We also fit the standard factor analysis based on the Pearson correlation and ML. In the last model however, rather than using the standard Z-scores of the observed variables, we use a U-score transformation introduced by Kamanou and Doksum 2002, which has been shown in the context of principal component analysis to attenuate the bias in the correlation coefficient when categorical variables are skewed . The data used in this analysis are the most recent household survey data from three West African countries including Sierra Leone, Gambia and Senegal. Since our primary objective in an exploratory factor analysis of multidimensional poverty data, we compare three methods (of combination of fitting function and alternative correlation matrix) on the basis of parameters estimates parameters rather than fit statistics.

[The remainder of this section is under development]

5. Conclusion [To be written]

References

Anderson T. W. (1959), *An introduction to multivariate statistical analysis*. New York: Wiley

Babakus, E. Ferguson, C and Joereskog, G. (1987), The *sensitivity of Confirmatory Maximum Likelihood Factor Analysis to Violations of Measurement Scale and Distributional Assumptions*, Journal of Marketing Research, Vol. 24, No. 2 (1987), 222-228

Bartholomew, D. J. (1980), *Factor Analysis for Categorical Data*, Journal of the Royal Statistical Society. Series B, Vol. 42, No 3, 293 – 321

Carroll, J. B. (1983), *The difficulty of a test and its factor composition revisited. In H. Wainer & S. Messick (Eds.),* Principle of modern psychological measurement. Hillsdale, NJ ournal of the Royal Statistical Society. Series B, Vol. 42, No 3, 293 – 321

Gilley William F. and Uhlig George E. (1980), *Factor Analysis and OrdinalData*, ???

Goodman, L.A. (1978), *Analysing Qualitative/Categorical Data Log-Linear Models and Latent Structure Analysis.* Reading, Mass: Addison-Wesley

Mislevy R. (1986), *Recent Development in Factor Analysis of Categorical Variables*. Journal of Educational Statistics, Vol. 11, No. 1, 1-131

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

Muthèn B. O. (1978), *Contribution to factor analysis of dichotomous variables.* Psychometrika, 43, 551-560

Muthèn B. (1985), *A comparison of some methodologies for factor analysis of non-normal Likert variables.* British Journal of Mathematical and Statistical Psychology, Vol. 38, 171-189

Johnson, R. and Creech C. (1983), *Ordinal Measures in Multiple Indicators Models: A Simulation Study of Categorization Error.* American Sociological Review, Vol. 48, 398-407

Jöreskog, K.G and Golberger, A. S. (1972), *Factor analysis by generalized least squares. Psychometrika. Vol*. 37, 243-259

Jöreskog K.G and Sörbom D. (1986), *PRELIS: A Preprocessor for Lisrel, Mooresville,* IN: Scientific Software, Inc

Jöreskog K.G and Sörbom D. (1981), *LISREL: Analysis of Linear Structure Relationships by the Method of Maximum Likelihood.* Chicago: National Educational Resources, INC

Journal of Consumer Psychology (2001), *Factor Analysis*. Vol. 10, (1&2), 75 – 82

Kamanou G. and Doksum K. (Unpublished manuscript - 2002), *A Comparison of Different Scaling Methods in Principal Component Analysis* (To request a copy of the paper please send an email to kamanou@un.org)

Pearson K. (1901), *On lines and planes of closest to systems of points in space.* Phil. Mag. 2, 559-572

Pearson K. and Pearson S. (1922), *On The Polychoric Coefficient of Correlation.* Biometrika Vol. 14, No 1/2, 127-156.

Rigdon E. and Ferguson E. (1991), *The Performance of the Polychoric Correlation Coefficient and Selected Fitting Function in Confirmatory Factor Analysis with Ordinal Data.* Journal of Marketing Research Vol. 38, 491-497

Ritchie-Scott. (1918), *The Correlation Coefficient of a Polychoric Table.* Biometrika Vol. 12, No 1/2, 93-133.

Velicer W. F. and Jackson D. N. (1990), *Component Analysis versus Common Factor Analysis: Some Issues in Selecting and Appropriate Procedure*, Multivariate Behavioral Research, 25 (1), 1 – 28

Wildaman K. F. (1993), *Common Factor Analysis Versus Principal Component Analysis Analysis: Differential Bias in Representing Model Parameters?* Multivariate Behavioral Research, 28 (3), 263 – 311

International Poverty Centre
SBS – Ed. BNDES, 10º andar
70076-900 Brasilia DF
Brazil

povertycentre@undp-povertycentre.org
www.undp.org/povertycentre
Telephone +55 61 2105 5000