THE MANIFEST ASSOCIATION STRUCTURE OF THE SINGLE-FACTOR MODEL: INSIGHTS FROM PARTIAL CORRELATIONS

MARIA DE FÁTIMA SALGUEIRO, PETER W. F. SMITH,
JOHN W. MCDONALD

ABSTRACT

The association structure between manifest variables arising from the single-factor model is investigated using partial correlations. The additional insights to the practitioner provided by partial correlations for detecting a single-factor model are discussed. The parameter space for the partial correlations is presented, as are the patterns of signs in a matrix containing the partial correlations that are not compatible with a single-factor model.

Southampton Statistical Sciences Research Institute
Methodology Working Paper M07/07

THE MANIFEST ASSOCIATION STRUCTURE OF THE SINGLE-FACTOR MODEL: INSIGHTS FROM PARTIAL CORRELATIONS

Maria de Fátima Salgueiro

departamento métodos quantitativos, iscte business school

Peter W.F. Smith

southampton statistical sciences research institute,

university of southampton

John W. McDonald

southampton statistical sciences research institute,

university of southampton

March 30, 2007

The research of the first author was supported by the Fundação para a Ciência e a Tecnologia and by the Southampton Statistical Sciences Research Institute.

The authors thank the Editor, the Associate Editor and three Reviewers for their constructive comments which helped to improve the paper.

Correspondence should be sent to Maria de Fátima Salgueiro

E-Mail:	fatima.salgueiro@iscte.pt
Phone:	+351-963077654
Fax:	+351-217903941

THE MANIFEST ASSOCIATION STRUCTURE OF THE SINGLE-FACTOR MODEL: INSIGHTS FROM PARTIAL CORRELATIONS

Abstract

The association structure between manifest variables arising from the single-factor model is investigated using partial correlations. The additional insights to the practitioner provided by partial correlations for detecting a single-factor model are discussed. The parameter space for the partial correlations is presented, as are the patterns of signs in a matrix containing the partial correlations that are not compatible with a single-factor model.

Key words: anti-image correlation matrix, elliptical tetrahedron, factor analysis, factor partial correlation, manifest partial correlation.

1. Introduction

Factor analysis is a classical approach to modeling multivariate data where all variables are treated on an equal footing. Traditionally, these are thought of as models for covariances or correlations, but partial correlations can be used.

Guttman (1953) advocated that all off-diagonal elements of the inverse correlation matrix close to zero was a criteria for ruling out a common-factor space for the population. Kaiser (1970), in a Presidential address to the Psychometric Society, stated that "among Guttman's landmark papers . . . one particular theme appears repeatedly: given a correlation matrix, we should always look carefully at its inverse, in order to assess the sampling adequacy of the data for factor-analytic purposes". The scaled, to have ones on the diagonal, inverse variance matrix has off-diagonal elements (i, j) equal to the negatives of the partial correlation between variables i and j, after conditioning on the remaining variables (Whittaker, 1990, page 143). This matrix is known as the anti-image correlation matrix and is available in standard statistical packages such as SPSS, SAS and STATA. For further details on the relationship between correlation matrices and anti-image correlation matrices, see Yanai and Mukherjee (1987).

This paper presents three main results and two corollaries that provide extra insights into single-factor models, based on partial correlations and on the pattern of their signs. Section 2 presents necessary, but not sufficient, constraints on the parameter space of the population partial correlations implied by a single-factor model. Section 3 defines factor partial correlations and derives the association structure between manifest variables after marginalizing over the latent variable in a single-factor model. Also presented in Section 3 are patterns of signs of partial correlations that are not compatible with a single-factor model. Section 4 re-expresses the classical tetrad conditions as a function of the manifest partial correlations. Section 5 contains a discussion.

2. The Parameter Space of the Single-Factor Model

The classical single-factor model can be written as $\boldsymbol{X_M} = \boldsymbol{\lambda} L + \boldsymbol{\delta}$, where $\boldsymbol{X_M}$ is the vector of the p manifest variables X_1, X_2, ..., X_p, L is the factor or latent variable, $\boldsymbol{\lambda}$ is a $p \times 1$ vector of factor loadings, and $\boldsymbol{\delta}$ is a vector of p variables representing random measurement error and indicator specificity. Variables are considered to be measured as deviations from their means, so $E[\boldsymbol{X_M}] = \mathbf{0}$ and $E[L] = 0$. The classical single-factor model assumes that $E[L\boldsymbol{\delta}] = \mathbf{0}$, $E[\boldsymbol{\delta}] = \mathbf{0}$, var$[\boldsymbol{\delta}]$ is diagonal and that L and $\boldsymbol{\delta}$ are multivariate normal, although the results presented in this paper do not require these normality assumptions. To identify the model, the latent variable L is scaled to have unit variance. The variance matrix for $\boldsymbol{X_M}$, with elements denoted by σ_{ij}, is $\boldsymbol{\Sigma_M} = \boldsymbol{\lambda}\boldsymbol{\lambda}^T + \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is the $p \times p$ diagonal variance matrix of $\boldsymbol{\delta}$, with non-negative elements. A necessary and sufficient condition for $\boldsymbol{\Sigma_M}$ to be a variance matrix of a single-factor model is that $p(p-1)/2 - p$ independent tetrad conditions are satisfied and $0 \leq \frac{\sigma_{ki}\sigma_{ij}}{\sigma_{kj}} \leq \sigma_{ii}$ for one pair $(j \neq k)$ for each i (Anderson and Rubin, 1956). The tetrad conditions are $\sigma_{ki}\sigma_{lj} - \sigma_{li}\sigma_{kj} = 0$, for all distinct i, j, k and l, from 1 to p.

Results in this paper are presented under the assumption that variances are positive and that all variance matrices are positive definite. Therefore, a necessary but, as will be seen later, not sufficient condition on the scaled inverse variance matrix is that it is positive definite.

Result 1. The $p \times p$ anti-image correlation matrix is positive definite if and only if its determinant is strictly positive and, for $p \geq 4$, all its symmetric $(p-1) \times (p-1)$ submatrices are positive definite. Hence, when $p = 3$, the positive definiteness constraint on the anti-image correlation matrix implies

$$1 - 2\,\rho_{12.3}\,\rho_{13.2}\,\rho_{23.1} - \rho_{12.3}^2 - \rho_{13.2}^2 - \rho_{23.1}^2 > 0, \tag{1}$$

where $\rho_{ij.k}$ is the partial correlation coefficient between X_i and X_j given X_k.

The proof of Result 1 follows from Hadley (1961), Equation 7-69. The different combinations of the three $\rho_{ij.k}$ form a convex set which is symmetric with respect to rotations corresponding to permutations of the components of $(\rho_{12.3},\ \rho_{13.2},\ \rho_{23.1})$.

Figure 1 displays the boundary of the convex set in a three-dimensional plot. These plots should be contrasted to those in Figure 1 of Rousseeuw and Molenberghs (1994): both are elliptical tetrahedrons, but with different orientations.

==========================

Insert Figure 1 about here

==========================

Ignoring for the moment the shaded area, Figure 2 supports the statement that any horizontal cross section of the surface defined by Figure 1 is an ellipse.

==========================

Insert Figure 2 about here

==========================

3. The Relationship Between the Manifest and the Factor Partial Correlations

Let $\boldsymbol{X}$ be the vector of the $p+1$ random variables, partitioned as $\left[\begin{array}{c|c} \boldsymbol{X_M}^T & L \end{array}\right]^T$, with positive definite scaled inverse variance matrix partitioned as

$$\boldsymbol{T} = \left[\begin{array}{c|c} \boldsymbol{I} & \boldsymbol{T_{ML}} \\ \hline \boldsymbol{T_{LM}} & 1 \end{array}\right],$$

where $\boldsymbol{I}$ is a $p \times p$ identity matrix since the single-factor model assumes that the manifest variables are conditionally independent, given the latent variable. The $p \times 1$ vector $\boldsymbol{T_{ML}} = \boldsymbol{T_{LM}}^T$ contains the non-zero elements $-\tau_{iL.(iL)}$, the negative of the partial correlation between manifest variable X_i and latent variable L, where (iL) denotes the remaining $p-1$ variables in $\boldsymbol{X}$, after removing X_i and L. Henceforth, the $\tau_{iL.(iL)}$ are called factor partial correlations.

The positive definiteness constraint on the matrix $\boldsymbol{T}$ implies that its determinant is positive, i.e., $|1 - \boldsymbol{T_{LM}}\boldsymbol{T_{ML}}| > 0$, and that the determinants of all its symmetric submatrices are also positive. Therefore,

$$\sum_{i=1}^{p} \tau^2_{iL.(iL)} < 1. \qquad (2)$$

Also, because the τ are partial correlations, $-1 < \tau_{iL.(iL)} < 1$. The parameter space for the factor partial correlations is defined by Equation 2, with the additional constraint that $\tau_{iL.(iL)} \neq 0$. Indeed, for the single-factor model to be of interest all factor partial correlations need to be of substantive interest, and not only non-zero mathematically. Note that Equation 2 defines an open unit hyper-sphere.

Result 2. The manifest partial correlations can be written as a function of the factor partial correlations as

$$\rho_{ij.(ij)} = \frac{\tau_{iL.(iL)}\,\tau_{jL.(jL)}}{\{(1-\tau^2_{iL.(iL)})(1-\tau^2_{jL.(jL)})\}^{1/2}}, \qquad i \neq j \in \{1, 2, \ldots, p\}, \tag{3}$$

where $\rho_{ij.(ij)}$ is the manifest partial correlation between X_i and X_j, given the remaining $p-2$ manifest variables.

The proof follows. Inverting $\boldsymbol{T}$, the scaled inverse variance matrix of $\boldsymbol{X}$, gives

$$\boldsymbol{T}^{-1} = \left[\begin{array}{c|c} \boldsymbol{I} & \boldsymbol{T_{ML}} \\ \hline \boldsymbol{T_{LM}} & 1 \end{array}\right]^{-1} = \begin{bmatrix} (\boldsymbol{I}-\boldsymbol{T_{ML}}\boldsymbol{T_{LM}})^{-1} & -(\boldsymbol{I}-\boldsymbol{T_{ML}}\boldsymbol{T_{LM}})^{-1}\boldsymbol{T_{ML}} \\ -\boldsymbol{T_{LM}}(\boldsymbol{I}-\boldsymbol{T_{ML}}\boldsymbol{T_{LM}})^{-1} & (1-\boldsymbol{T_{LM}}\boldsymbol{T_{ML}})^{-1} \end{bmatrix}. \tag{4}$$

The correlation matrix of $\boldsymbol{X_M}$ is given by scaling $(\boldsymbol{I}-\boldsymbol{T_{ML}}\boldsymbol{T_{LM}})^{-1}$. Hence, the anti-image correlation matrix is given by scaling $\boldsymbol{I}-\boldsymbol{T_{ML}}\boldsymbol{T_{LM}}$ to have ones on the diagonal. The off-diagonal element (i,j) is given by $-\frac{\tau_{iL.(iL)}\,\tau_{jL.(jL)}}{\{(1-\tau^2_{iL.(iL)})(1-\tau^2_{jL.(jL)})\}^{1/2}}$. Since the off-diagonal elements of the population anti-image correlation matrix are the negatives of the population manifest partial correlations, Equation 3 follows.

Two corollaries of Result 2 follow.

Result 2a. Marginalizing the single-factor model over the latent variable L yields a joint distribution for the manifest variables with no zero manifest partial correlations, that is, with no zero entries in the anti-image correlation matrix.

Since all factor partial correlations are assumed non-zero, and the manifest partial correlations are given by Equation 3, all $\tau_{iL.(iL)} \neq 0$ imply all $\rho_{ij.(ij)} \neq 0$. Therefore, there are no zero entries in the anti-image correlation matrix and the proof is complete.

From Equation 3, the sign of $\rho_{ij.(ij)}$ equals the product of the signs of $\tau_{iL.(iL)}$ and $\tau_{jL.(jL)}$. Let $\boldsymbol{\tau} = -\boldsymbol{T_{ML}}$ be a $p \times 1$ column vector of the factor partial correlations and let $\boldsymbol{\rho}$ be a $p \times p$ symmetric matrix containing the manifest partial correlations as off-diagonal elements. Recall that $\boldsymbol{\lambda}$ is the vector of factor loadings.

Result 2b. The off-diagonal elements of $sign(\boldsymbol{\rho})$ represent a pattern of signs for the manifest partial correlations that is compatible with a single-factor model if

$$sign(\boldsymbol{\rho}) \;=\; sign(\boldsymbol{\tau}\,\boldsymbol{\tau}^T) \;=\; sign(\boldsymbol{\lambda}\,\boldsymbol{\lambda}^T) \tag{5}$$

(for the off-diagonal elements). This result refers to population parameters and assumes the tetrad conditions are satisfied. The second equality follows from Salgueiro (2003, Equation 5.16): $\lambda_i = \frac{\tau_{iL.(iL)}}{1-\sum_{k=1}^{p}\tau^2_{kL.(kL)}}$.

Note that, when $p = 3$, Equation 3 defines a system of three equations of the form

$$\tau^2_{iL.jk} = \frac{1}{1 + \frac{\rho_{jk.i}}{\rho_{ij.k}\rho_{ik.j}}}. \tag{6}$$

Since $0 < \tau^2_{iL.jk} < 1$, then $\frac{\rho_{jk.i}}{\rho_{ij.k}\rho_{ik.j}} > 0$. Hence, three manifest variables can only define a single-factor model if their anti-image correlation matrix is positive definite and the product of the three manifest partial correlation coefficients is positive. Revisiting Figure 2, the shaded area represents the admissible region for the three manifest partial correlation coefficients in a single-factor model. The ellipse represents the positive definiteness constraint given by Equation 1. Additionally, there is now the constraint of a positive product of the three partial correlations. Indeed, it is not possible to have a single-factor model with three manifest variables in which just one or all three $\rho_{ij.k}$ are negative. Also, if $\tau_{iL.jk}$ is negative and the other two factor partial correlations are positive (or the opposite), then all manifest partial correlations in column (row) i of the anti-image correlation matrix will be negative, and all the remaining manifest partial correlations will be positive.

When $p \geq 4$, Equation 5 can be used to derive possible patterns of signs for the manifest partial correlations, given the hypothesized structure of signs for the factor partial correlations. Alternatively, by considering all possible patterns of signs for the factor partial correlations, Equation 5 can be used to obtain all possible patterns of signs for the manifest partial correlations.

4. An Extension of the Tetrad Conditions

The classical tetrad conditions for the single-factor model can be re-expressed in terms of partial correlations.

Result 3. When $p \geq 4$, the following $p(p-1)/2 - p$ tetrad conditions for the partial correlations have to be satisfied in a single-factor model:

$$\rho_{ki.(ki)}\rho_{lj.(lj)} - \rho_{li.(li)}\rho_{kj.(kj)} = 0, \tag{7}$$

with i, j, k and l distinct, from 1 to p, where $\rho_{ki.(ki)}$ is the partial correlation coefficient between X_k and X_i, given the remaining $p-2$ manifest variables.

The proof follows by repeated substitution of Equation 3 into the left hand side of Equation 7.

5. Discussion

This paper has provided some useful results for the practitioner. Inspection of sample partial correlations between manifest variables can provide evidence for and against the single-factor model. If the pattern of signs of the sample partial correlations do not obey Result 2b, then, provided this is not due to sampling error, the single-factor model can be rejected. Indeed, population partial correlations between manifest variables in a single-factor model should be non-zero. Therefore, provided there is adequate power for the tests that these are zero, from Result 2a, sample partial correlations not significantly different from zero also rule out the single-factor model. However, if power for these tests is small, the single-factor model should still be considered, even if there are non-significant sample partial correlations. Methods for estimating overall power for graphical Gaussian models were investigated by Salgueiro et al. (2005). As the single-factor model can be parameterized as a graphical Gaussian model, the results presented by Salgueiro et al. (2005) can be used for the single-factor model.

References

Anderson, T.W. and Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability, Vol. V*, 111-150, Berkeley: University of California Press.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, **18**, 277-296.

Hadley, G. (1961). *Linear Algebra*. Massachusetts: Addison-Wesley Publishing Company, Inc.

Kaiser, H.F. (1970). A second generation Little Jiffy. *Psychometrika*, **35**, 401-415.

Rousseeuw, P.J. and Molenberghs, G. (1994). The shape of correlation matrices. *The American Statistician*, **48 (4)**, 276-279.

Salgueiro, M.F. (2003). *Distributions of Test Statistics for Edge Exclusion for Graphical Models*. PhD thesis, University of Southampton.

Salgueiro, M.F., Smith, P.W.F. and McDonald, J.W. (2005). Power of edge exclusion tests in graphical Gaussian models. *Biometrika*, **92 (1)**, 173-182.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester: John Wiley & Sons.

Yanai, H. and Mukherjee, B.N. (1987). A generalized method of image analysis from an intercorrelation matrix which may be singular. *Psychometrika*, **52 (4)**, 555-564.

Figures

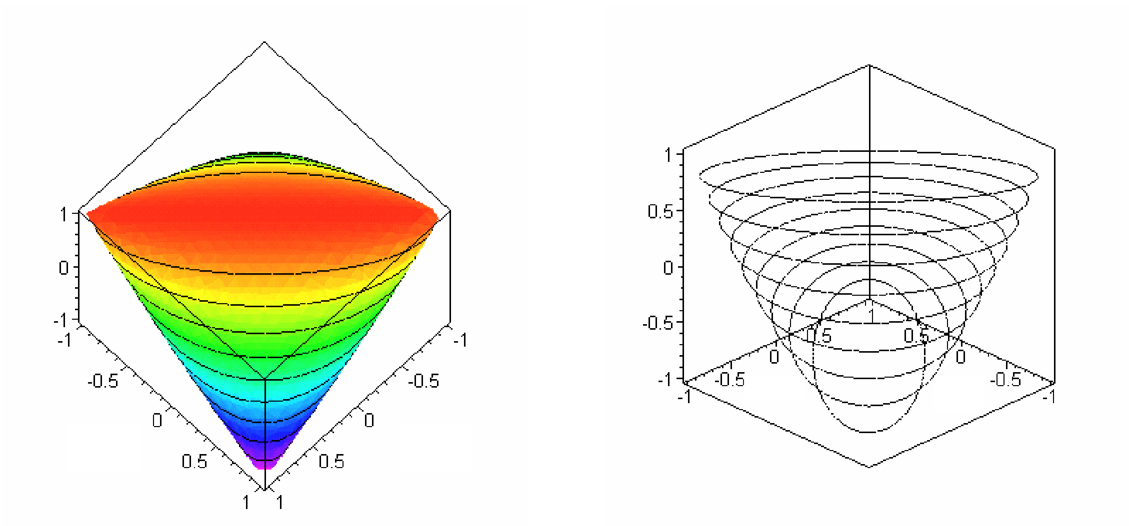

FIGURE 1.
Set of all possible partial correlation coefficients ($\rho_{12.3}$, $\rho_{13.2}$ and $\rho_{23.1}$) between three variables.

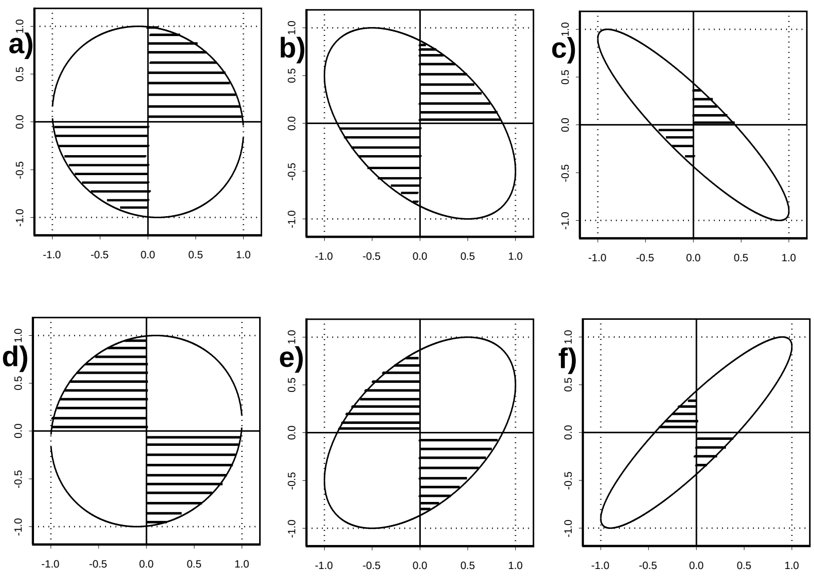

FIGURE 2.
The admissible region (shaded) for the three partial correlation coefficients compatible with a single-factor model: $\rho_{13.2}$ on the horizontal axis, $\rho_{23.1}$ on the vertical axis and $\rho_{12.3}$ taking values of a) 0.1, b) 0.5, c) 0.9, d) -0.1, e) -0.5, f) -0.9. The ellipse represents the positive definiteness constraint for partial correlations.