UNIVERSITAT ROVIRA I VIRGILI

# Not positive definite inter-item correlation matrix and expected improper results in Unweighted Least Squares in the Exploratory Factor Analysis

*Pere J. Ferrando*
*Urbano Lorenzo-Seva*

**Tarragona 2020**

Please reference this document as:

Ferrando, P. J., & Lorenzo-Seva, U. (2020). *Not positive definite inter-item correlation matrix and expected improper results in Unweighted Least Squares in the Exploratory Factor Analysis. Technical Report*. Department of Psychology, Universitat Rovira i Virgili, Tarragona.

Document available at: http://psico.fcep.urv.cat/utilitats/factor/

# Not positive definite inter-item correlation matrix and expected improper results in Unweighted Least Squares in the Exploratory Factor Analysis



Least-squares exploratory factor analysis based on tetrachoric/polychoric correlations is a robust, defensible and widely used approach for performing item analysis, especially in the first stages of scale development. A relatively common problem in this scenario, however, is that the inter-item correlation matrix fails to be positive definite. The text contained in this document is an appendix to the information published in Lorenzo-Seva and Ferrando (2020).

An inter-item correlation matrix is positive definite (PD) if all of its eigenvalues are positive. It is positive semidefinite (PSD) if some of its eigenvalues are zero and the rest are positive. Finally, it is indefinite if it has both positive and negative eigenvalues (e.g. Wothke, 1993). This last situation is also known as not positive definite (NPD).

In the first Unweighted Least Squares (ULS) approach, uniqueness values $\mathbf{\Psi}^2$ are first determined, and the loading estimates are obtained by solving the equation (e.g. Jöreskog, 2003):

$$(\mathbf{R}-\mathbf{\Psi}^2)\mathbf{\Lambda}=\mathbf{R}^*\mathbf{\Lambda}=\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Lambda}). \qquad (1)$$

With the usual ULS constraint that $\mathbf{\Lambda}'\mathbf{\Lambda}$ is diagonal (e.g. McDonald, 1985), it follows that the estimated columns of $\mathbf{\Lambda}$ are the eigenvectors of the reduced correlation matrix **R***. This means that, in order to obtain a solution for expression (1) in which the estimated elements of $\mathbf{\Lambda}$ are real and unique, **R*** must be, at least, PSD.

Now, a reduced correlation matrix (**R**-**D**) that is at least PSD can always be obtained, for example, via minimum trace FA (Bentler, 1972). If **R** is PD, then, the diagonal elements of **D** will generally all be positive, and can be properly interpreted as uniquenesses (i.e., **D** =

$\mathbf{\Psi}^2$). However, if **R** is indefinite, one or more diagonal entries of **D** will be necessarily negative (Bentler & Yuan, 2011, Lemma 1). This implies that some of the communality estimates (the diagonal elements of **R**-**D**) would be greater than 1 (i.e. strong Heywood cases).



Assume further that **D** is chosen so that **R*** has minimum rank (e.g., ten Berge & Kiers, 1991), we can then consider **R*** as the model-implied reduced matrix that would be obtained if an EFA model with all the factors that are needed to account for the common variance of the items was fitted to the data.

The developments discussed so far account for certain results repeatedly discussed in the EFA literature (e.g. McDonald, 1985). So, if a model with fewer factors than those needed to account for all the common variance is fitted to the data, it is possible that no Heywood cases will appear. However, they are more and more likely to appear as the number of factors is increased. And if a model with all the needed factors (in the minimum rank sense) is fitted, they will certainly appear.

We turn now to the second approach. In this approach (of which MINRES is the best-known procedure), the ULS criterion is applied only to the non-diagonal elements of **R**, and the uniquenesses (or commonalities) is obtained as a by-product. In this case **R*** is not constrained to be PSD and is, in fact, indefinite: For $n$ variables and $k$ specified factors, the smallest $n - k$ eigenvalues of **R*** sum to zero, which means that at least one must be negative. This result has been called *Harman's paradox* (ten Berge, 2000).

Now, to see how the improper results are also expected in this second approach, we shall first consider that the MINRES solution is also the solution of expression (1) when **R*** is obtained by using the MINRES estimated uniquenesses (Harman, 1976). And then we shall see that the reproduced correlation matrix:

$$\mathbf{R}^{rep} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}'. \qquad (2)$$

is always PSD (Gramian) by construction. Now, as the number of extracted factors increases, $\mathbf{R}^{rep}$ increasingly approaches **R*** (i.e. a PSD reduced matrix) so the expected improper estimates when **R** fails to be PD are also expected to appear here.



## References

Bentler, P. M. (1972). A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research*, *1*(4), 343-357. doi:10.1016/0049-089x(72)90082-8

Harman, H. H. (1976). *Modern factor analysis*. University of Chicago press.

Jöreskog, K. G. (2003). *Factor analysis by MINRES. Technical report*. Available on http://www.ssicentral.com/lisrel/techdocs/minres.pdf

Lorenzo-Seva, U. & Ferrando, P. J. (2020). Not Positive Definite Correlation Matrices in Exploratory Item Factor Analysis: Causes, Consequences and a Proposed Solution. *Structural Equation Modeling: A Multidisciplinary Journal*, DOI: 10.1080/10705511.2020.1735393

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale: LEA.

Ten Berge, J. M. (2000). Linking reliability and factor analysis: Recent developments in some classical psychometric problems. *Advances in personality psychology*, *1*, 138-156.

Ten Berge, J. M., & Kiers, H. A. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, *56*(2), 309-315. doi:10.1007/bf02294464

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K.A. Bollen y J.S. Long (Eds.), *Testing structural equation models* (pp. 256-293). Newbury Park: Sage.