Original Article

A Simulation-Based Scaled Test Statistic for Assessing Model-Data Fit in Least-Squares Unrestricted Factor-Analysis Solutions

Urbano Lorenzo-Seva [1], Pere J. Ferrando [1]

[1] *Faculty of Sciences of Education and Psychology, University Rovira i Virgili, Tarragona, Spain.*

Abstract

A shortcoming of least-squares unrestricted factor analysis (UFA) procedures, which are widely used in psychometric applications is that a test statistic for assessing model-data fit cannot be easily derived from the minimum fit function value. This paper proposes a chi-square type goodness-of-fit test statistic intended for the principal-axis, MINRES, and minimum-rank UFA procedures. The statistic is empirically obtained via intensive simulation based on a two-stage approach. First, a distribution of minimum fit function values is obtained from a scenario in which the null hypothesis of perfect model-data fit holds. Second, the obtained statistic is non-linearly transformed so that it has its first four moments equal to those of the theoretical reference chi-square distribution with the appropriate degrees of freedom. Extensions of the basic statistic are next proposed that include comparative and relative indexes based on it. Tests of close-fit and power assessment derived from the basic statistic are also proposed.

Keywords

Exploratory or unrestricted factor-analytic (UFA) solutions are those that impose the minimum constraints for identifiability and leave the common factor space unrestricted, which means that the initial solution can be (and usually is) further rotated. Even

though it is more indeterminate than a restricted solution, the unrestricted solution is much more flexible and does not require zero-loading constraints to be imposed for identifying the pattern, which means that the variables in a UFA solution are allowed to be factorially complex. This specification-related flexibility makes UFA a very valuable (possibly the best) tool in many psychometric applications, particularly, at the early stages of test development, when the aim is to assess the dimensionality of an item pool without imposing any particular relational structure to the data (Ferrando, 2021). And, at the latter stages, UFA continues to be a flexible and versatile tool for assessing the structure of an instrument when some of its items are complex.

The position taken in this paper is that in the scenario summarized above, not only is generally UFA the most appropriate model but also that, under rather common conditions, an UFA solution is expected to work better when fitted with simple procedures. To be more specific, the common conditions are: (a) a large number of items, (b) not too large samples, and (c) complex structures (e.g., Muñiz & Fonseca-Pedrero, 2019). And, with regards to the simple estimation procedures, we shall consider here the family of UFA procedures based on the unweighted least squares (ULS) criterion (see below). These methods have been considered (somewhat disparagingly) as approximate or second rate with respect to more statistically rigorous methods such as Maximum Likelihood (ML) or Generalized Least Squares (GLS). A literature review, however, (and also our experience) suggests that, when compared to ML or GLS solutions in the scenario considered here, ULS solutions are faster, computationally simpler, robust, stable (particularly for categorical variables), less prone to arrive at improper solutions, and less likely to be affected by minor irrelevant factors (Ferrando & Lorenzo-Seva, 2017). Possibly for these reasons, the ULS-based methods based on both continuous and discrete variables are commonly used in the type of applications considered here (Revelle, 2022).

Meaningful assessment of the appropriateness of an UFA solution in the scenario described above requires a complex and multifaceted approach to be undertaken. Goodness of model-data fit (GOF), as based on the chi-squared test statistic, is, in principle, a basic property that has to be assessed, and here lies the focus of the present proposal.

As for the intended use of the proposal, we do not consider the chi-squared test statistic as the final measure of fit, but as, (a) a useful measure when accompanied by power information and, (b) a necessary basis for obtaining GOF indices that might function well in psychometric applications of the UFA model. Also we do not regard these new indices as substitutes for those that currently exist and that do work, but rather as useful complements to them. Finally, the inherent difficulties of theoretically deriving a chi-squared statistic because of both the properties of the ULS estimator and the characteristics of the item scores (see below), suggest that intensive simulation is an appropriate approach for arriving at this type of statistic.

Aims of the Proposal

The present article proposes an empirical test statistic in chi-square metric for assessing goodness of model-data fit in UFA solutions based on the ULS criterion, specifically: Principal-Axis-Factoring (PAF), ULS-MINRES, and Minimum Rank Factor Analysis (MRFA). The UFA-ULS solutions, in turn can be based on the standard linear FA model in which the variables are treated as continuous or on the non-linear UVA-FA model in which they are treated as ordered-categorical.

Our proposal is fully empirical, and avoids theoretical developments at the cost of intensive simulation which, given the capabilities of modern computers, is perfectly affordable. The basic idea is to combine: (a) the rationale of previous empirical proposals based on the idea of sampling from a simulated scenario in which the model holds exactly with, (b) non-linear transformations that bring the distribution of the resulting fit statistic close to the expected chi-square distribution. We shall label the approach as LOSEFER, as an acronym of Lorenzo-Seva and Ferrando's approach.

Description of the Procedure

Consider a set of m observed variables related to p common factors, the population standardized variance-covariance (i.e., correlation) matrix $\mathbf{\Sigma}$ ($m \times m$) among the set of observed variables, and the corresponding estimate $\mathbf{R}$ ($m \times m$) obtained in a sample of N observations. When $\mathbf{R}$ is obtained from a large and representative sample from the population, $\mathbf{R}$ is expected to be a good estimate of $\mathbf{\Sigma}$.

The direct UFA model decomposes $\mathbf{\Sigma}$ as,

$$\mathbf{\Sigma} = \mathbf{\Gamma\Gamma'} + \mathbf{\Psi}, \tag{1}$$

where $\mathbf{\Gamma}$ is the loading matrix ($m \times p$), and $\mathbf{\Psi}$ is the diagonal matrix ($m \times m$) with the unique variances in the main diagonal. When an UFA solution is fitted to sample data, the aim is to estimate matrices $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ from the observed matrix $\mathbf{R}$. In terms of the sample estimate, matrix $\mathbf{R}$ is decomposed as,

$$\mathbf{R} = \mathbf{AA'} + \mathbf{U} + \mathbf{E}, \tag{2}$$

where $\mathbf{A}$ ($m \times k$) and $\mathbf{U}$ ($m \times m$) are the corresponding estimates of $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ in Expression (1), and $\mathbf{E}$ ($m \times m$) represents the amount of observed covariance in $\mathbf{R}$ that cannot be accounted for by the sample factor model. When k (the number of factors in the sample model) is chosen to equal p (the number of factors in the population model), the observed values in $\mathbf{E}$ will tend to be zero, and the estimated factor model is expected to attain an appropriate goodness-of-fit. However, the typical situation when fitting a UFA solution (especially when used with exploratory purposes) is that the value of p is not known, so that k is assigned a tentative value that aims to achieve an appropriate goodness-of-fit level for the sample factor model.

Goodness-of-Fit Assessment

In order to define a scalar goodness-of fit test statistic for a prescribed UFA solution obtained from the PAF, MINRES or MRFA approaches, the matrix $\mathbf{E}$ is derived from Expression (2) as,

$$\mathbf{E} = \mathbf{R} - \mathbf{AA'} - \mathbf{U}, \tag{3}$$

and the discrepancy function we shall consider here is:

$$\sum_{i=1}^{m-1} \sum_{j \neq i}^{m} e_{ij}^2, \tag{4}$$

where the e_{ij}^2 terms are the non-diagonal elements of $\mathbf{E}$. So, the discrepancy function (4) is the sum of non-duplicated squared residuals between the observed and reproduced correlation matrix. The test statistic we consider based on this discrepancy function is now:

$$c = (N-1) \sum_{i=1}^{m-1} \sum_{j \neq i}^{m} e_{ij}^2. \tag{5}$$

The discrepancy function in (4) is the ordinary or unweighted least squares (ULS) function, which is the simplest in covariance structure analysis. As mentioned above, all the methods considered here: PAF, ULS-MINRES and MRFA are essentially ULS methods, and so are based on the minimization of (4).

If the ULS estimates were asymptotically efficient, the distributional assumptions mentioned above were met, and the null hypothesis of exact fit in (1) would hold, then (5) would be asymptotically distributed as a chi-square variable with degrees of freedom,

$$df = \frac{1}{2}(m-k)(m-k+1) - m, \tag{6}$$

(see e.g., Lawley, 1940).

Because the ULS estimates are never asymptotically efficient, neither for continuous nor for ordered-categorical solutions, and the fulfillment of the distributional assumptions cannot be taken for granted, a *naïve* theoretical chi-square reference distribution cannot be claimed for (5). To address this problem, the proposal here is to non-linearly transform the test statistic (5) so that, when the fitted solution holds in the population, the transformed statistic closely approaches a central chi-square distribution with degrees of freedom (6). Next, the sample test statistic undergoes the same transformation, so that, when null hypothesis (H_0) holds, it is interpretable as a value sampled from a chi-square distribution. Overall, then, our proposal can be regarded as a correction of a

chi-square type test statistics that brings its distribution closer to that expected theoretically when the null hypothesis holds. The main basis difference with existing corrections of this type (i.e., robust statistics) is that the correction is not based on asymptotic theory but is empirical. Finally, as for requirements and assumptions for undertaking the transformation, the basic requirement is that both, the observed and reproduced covariance matrices are positive definite (see Lorenzo-Seva & Ferrando, 2021). And, as for the basic assumptions, being (4) a sum of squares that adopts only positive values, it is assumed that the distribution of (5) will be positively skewed, and will approach normality as the model becomes larger.

Empirically Obtaining the Scaled Test Statistic in Chi-Square Metric

Once an UFA solution has been fitted to an observed correlation matrix $\mathbf{R}$, the reproduced variance-covariance matrix is defined as,

$$\mathbf{R}^* = \mathbf{AA'} + \mathbf{U}. \tag{7}$$

Let's take this $\mathbf{R}^*$, as if it was a true population matrix, and decompose it using Cholesky's method,

$$\mathbf{L} = \text{chol}(\mathbf{R}^*). \tag{8}$$

If $\mathbf{Z}$ ($N \times m$) is a random matrix with columns normally distributed $N(0, 1)$, the product,

$$\mathbf{X} = \mathbf{ZL}, \tag{9}$$

produces a population matrix $\mathbf{X}$ ($N \times m$). From this population matrix $\mathbf{X}$, random samples $\mathbf{X}_i$ ($N_i \times m$) of observed scores are sampled for which the corresponding correlation matrix $\mathbf{R}_i$ is an estimate of the true population matrix $\mathbf{R}^*$. Matrix $\mathbf{R}_i$ is then factor analyzed in order to obtain estimates $\mathbf{A}_i$, $\mathbf{U}_i$, and

$$\mathbf{E}_i = \mathbf{R}_i - \mathbf{A}_i\mathbf{A}_i' - \mathbf{U}_i. \tag{10}$$

Finally, the test statistic c_i is obtained using expression (5) applied to matrix $\mathbf{E}_i$.

The process is repeated for an arbitrary number of times K ($i = 1...K$), in order to obtain a vector of $\mathbf{c}$ that contains the K values c_i. The distribution of the elements of this vector is then the distribution of the uncorrected test statistic when the null hypothesis holds. In the studies presented in this document, we used $K = 1,000$ and $N = 100,000$, and N_i equal to the size of the sample used to obtain observed matrix $\mathbf{R}$, and arrived at acceptable results. In addition, it must be pointed out that, if $\mathbf{X}$ is a set of ordinal variables, each $\mathbf{X}_i$ must be discretized using the empirical thresholds estimated from $\mathbf{X}$, and the computed correlation matrices must be based on polychoric correlations.

So far, our proposal is similar to previous existing developments, and particularly to Bollen and Stine's (1992) bootstrapped approach (see also Corrêa Ferraz et al., 2022). The basic idea, in effect, is to sample (or resample in the bootstrapping approach) from a population in which the null hypothesis exactly holds. Furthermore, this condition is obtained by using the parent sample matrix as a basis (i.e., $\mathbf{R}^*$ is taken as if it was a $\boldsymbol{\Sigma}$ in Equation 1). More specifically, the original data is transformed (according to (8) and (9) in our proposal) so that is forced to satisfy the null hypothesis.

From here on, however, our proposal differs from the previous related developments. Bollen and Stine (1992) considered only the ML-based scenario for continuous outcomes, in which the test statistic was expected to truly follow a central chi-square distribution under the null hypothesis. Second, Bollen and Stine's (1992) proposal was not intended to make the empirical bootstrap distribution closer to the theoretical chi square distribution, but rather, to obtain reference p values to which the untransformed sample test statistic could be compared.

Continuing with our proposal, once c is available, what we propose is nonlinearly transform it using a third degree polynomial,

$$y = a + b_1 c + b_2 c^2 + b_3 c^3, \tag{11}$$

so that the first moment, and the second, third, and fourth central moment estimates of the transformed c coincide with those of the reference chi-square distribution with degrees of freedom (df) in (6). In more detail, the coefficients of the polynomial (11) are obtained by solving the following system,

$$
\begin{aligned}
E(y) &= \bar{y} = df \\
Var(y) &= 2df \\
E(y - \bar{y})^3 &= 8df \\
E(y - \bar{y})^4 &= 48df + 12df^2
\end{aligned}
\tag{12}
$$

Where $E()$ is used for expectation. Technical details on how to determine the polynomial coefficients in (11) from the system (12) can be obtained from the authors. As a summary, of the different solving procedures we tried, the most effective was two step. First, the original c values were: (a) cube-root transformed, and (b) transformed to have the first four moments of a standard normal variable (i.e., the first four moments in the system being 0, 1, 0 and 3). Second, the normal-transformed variable was transformed again using Fleishman's (1978) procedure to obtain a chi-square distribution from a standard normal distribution, so that the final transformed variable have the first four moments as close as possible to those in (12).

Once the coefficients a, b_1, b_2, and b_3 have been obtained, we can now factor analyze the sample correlation matrix $\mathbf{R}$ and compute the sample-observed c statistic by using

expressions (4) and (5). Next, c is transformed to y using the fitted polynomial (11) and this transformed value is interpreted with relation to a chi-square distribution with degrees of freedom (6).

A crucial point in interpreting the transformed sample c value as a proper chi-square type test statistic is, indeed, that not only the first four moments, but its distribution in general would adhere to the theoretical distribution under the null hypothesis of model-data fit. Strictly speaking, this adherence cannot be guaranteed, so we assessed this point using intensive simulation. Full results are provided below. However, it can be advanced that our proposal shows a considerable viability in this respect.

Beyond the Scaled Test Statistic: GOF Indices Test of Close Fit and Power Analysis

Provided that y approaches closely enough the corresponding reference distribution, it can be further used as a basis for computing meaningful point estimates of selected GOF statistics, so that the proposed solution can be more thoroughly assessed. We shall propose to derive CFI and RMSEA point estimates directly from the transformed y statistic. However, it should be clear that this is only an initial proposal that is expected to be updated as more information about the performance of GOF indices in UFA will become available.

Let $\lambda_1 = y_1 - df_1$ the noncentrality parameter estimate for the solution under study, and $\lambda_o = y_o - df_o$ the corresponding estimate for the solution with zero common factors (i.e., the null or baseline model). In terms of y, the comparative fit index point estimate can then be obtained as:

$$CFI = 1 - \frac{\max(\lambda_{1,}, 0)}{\max(\lambda_{0,}, \lambda_{1,}, 0)} \qquad (13)$$

And the RMSEA point estimate as:

$$RMSEA = \max(\sqrt{\frac{\lambda_{1,}}{(N-1)df_1}}, \quad 0) \qquad (14)$$

Our view, however, is that meaningful information of a GOF statistic requires not only the point estimate to be reported, but also the corresponding confidence interval. In the implementation approach in which the indices proposed here are programed (see below), 90% confidence intervals are reported based on bootstrap resampling.

The choice of the RMSEA as an index directly derived from the test statistic allows two further pieces of important information to be obtained: the test of close fit and power analysis (Lee et al., 2012). In our view, this information is highly relevant for avoiding two common pitfalls when fitting UFA solutions. The first is to use too small samples

in order to achieve a better fit (at the cost of a gross loss of power). The second is to over-factor with the same aim, a practice that ends up in considering as relevant trivial, uninterpretable minor factors which are devoid of any substantive interest (Ferrando & Lorenzo-Seva, 2018).

The implementation is straightforward. In our proposal we have set the null (close fit) and alternative hypothesis as,

$$H_0: RMSEA \leq 0.05$$
$$H_1: RMSEA > 0.05 \tag{15}$$

With regards to power assessment, we have chosen the approach by Lee et al. (2012) in which the noncentrality parameter that expresses the lack of fit in the population is obtained by setting RMSEA values under H_0 and H_1. In particular, in our implementation, power is computed as the capacity for distinguishing between a close fit solution (H_0: RMSEA = 0.05) and a moderately misspecified solution (H_1: RMSEA = 0.08).

Simulation Studies

Two simulation studies have been carried out. The first aims to assess if statistic y in (11) actually: (a) has the expected chi square distribution when the true factor model is actually fitted in the sample data, and (b) leads to rejection rates close to those expected under the chi square distribution. The second study aims to assess the rejection rates when the sample factor model is misspecified.

First Simulation Study

A Monte Carlo simulation study was carried out using samples drawn from a true population model. Based on Expression (1), a population loading matrix was defined in which each observed variable had a salient loadings had a values of .70, and unicity equal to 1 minus communality of the variable. The number of factors and the number of salient variables per factor were manipulated in order to produce models with different degrees of freedom:

- 5 degrees of freedom: The population matrix was defined by a single factor and five variables.
- 64 degrees of freedom: The population matrix was defined by two factors and seven salient variables per factor.
- 207 degrees of freedom: The population matrix was defined by three factors and eight salient variables per factor.
- 492 degrees of freedom: The population matrix was defined by four factors and nine salient variables per factor.

The sample sizes were also manipulated: 200, 500, and 800. From each true population matrix, 500 samples were obtained, and a total of 6,000 were factor analyzed.

For each sample, the true number of factors was extracted using two extraction methods: ULS/Minres, and Principal Axes (PA); and the y statistic computed after each extraction. The statistic y was computed using population samples of 100,000 and the number of random samples extracted from the population were $K = 1,000$.

The four firsts moments of the empirical distribution of y were computed and compared to the expected ones for the theoretical distribution of chi square statistic with degrees of freedom df. Table 1 shows the outcomes related to ULS/MINRES extraction. Kurtosis values are printed as zero centered (i.e., kurtosis minus 3). Mean and variances were slightly overestimated, especially in small samples. The estimates of skewness and kurtosis in general do not differ from the values expected in the population, and only in small models (5 degrees of freedom) and large samples ($N = 800$) the estimates are overestimating the expected values.

Rejection rates after ULS/MINRES factor extraction are shown in Table 2. The worse rejection rates were observed when the sample size was small ($N = 200$). In addition, the worse rejections rates estimates were the ones expected to be .001. It means that is at the farthest tail of the distribution of statistic y is where less adherence to the chi-square distribution is observed. From a practical point of view, the rejection levels observed are reasonable.

Table 1

Distributional Statistics of Chi Square Estimates After ULS/MINRES Factor Extraction

df	*N*	Mean	Variance	Skewness	Kurtosis
5	200	5.225 (5.157, 5.293)	12.828 (12.482, 13.174)	1.309 (1.292, 1.325)	2.330 (2.243, 2.417)
	500	5.142 (5.099, 5.185)	12.564 (12.328, 12.799)	1.316 (1.301, 1.332)	2.320 (2.239, 2.401)
	800	5.113 (5.074, 5.151)	13.087 (11.864, 14.309)	1.325 (1.427, 1.486)	4.514 (4.348, 4.680)
Expected		5	10	1.265	2.400
64	200	64.437 (63.971, 64.904)	129.108 (127.232, 130.984)	0.349 (0.339, 0.359)	0.188 (0.160, 0.203)
	500	64.663 (64.359, 64.967)	131.821 (130.427, 133.216)	0.364 (0.353, 0.374)	0.195 (0.171, 0.218)
	800	64.398 (64.160, 64.635)	131.169 (129.860, 132.479)	0.351 (0.341, 0.362)	0.160 (0.154, 0.196)
Expected		64	128	0.354	0.188
207	200	210.899 (209.823, 211.974)	421.635 (417.337, 425.933)	0.186 (0.176, 0.197)	0.055 (0.039, 0.072)
	500	209.250 (208.517, 209.982)	422.762 (418.992, 426.532)	0.196 (0.185, 0.206)	0.065 (0.046, 0.084)
	800	207.627 (207.011, 208.243)	420.336 (416.793, 423.879)	0.197 (0.187, 0.208)	0.058 (0.041, 0.076)
Expected		207	414	0.197	0.058
492	200	504.421 (502.596, 506.245)	1016.597 (1007.436, 1025.758)	0.133 (0.124, 0.143)	0.027 (0.012, 0.042)
	500	497.292 (495.889, 498.696)	1006.444 (998.051, 1014.837)	0.124 (0.116, 0.136)	0.037 (0.022, 0.052)
	800	495.881 (494.735, 497.026)	1002.491 (994.828, 1010.154)	0.143 (0.113, 0.132)	0.026 (0.011, 0.041)
Expected		492	984	0.128	0.024

Note. 95th confidence intervals are shown in parenthesis.

Table 2

Rejection Rates After ULS/MINRES Factor Extraction

		Rejection rates		
df	*N*	**.100**	**.050**	**.001**
5	200	.131 (.126, .136)	.075 (.071, .078)	.021 (.019, .022)
	500	.125 (.122, .128)	.070 (.068, .073)	.018 (.017, .019)
	800	.105 (.103, .108)	.060 (.058, .062)	.019 (.018, .020)
64	200	.127 (.118, .135)	.046 (.043, .048)	.012 (.011, .013)
	500	.121 (.116, .127)	.066 (.062, .069)	.016 (.015, .017)
	800	.113 (.109, .118)	.060 (.057, .063)	.014 (.013, .015)
207	200	.171 (.159, .183)	.103 (.094, .112)	.032 (.028, .036)
	500	.138 (.130, .146)	.078 (.072, .083)	.021 (.019, .023)
	800	.183 (.112, .125)	.068 (.060, .068)	.016 (.014, .017)
492	200	.225 (.210, .240)	.145 (.133, .156)	.052 (.046, .058)
	500	.159 (.149, .169)	.093 (.086, .100)	.027 (.024, .030)
	800	.143 (.135, .151)	.080 (.075, .085)	.021 (.019, .023)

Note. 95th confidence intervals are shown in parenthesis.

Table 3

Distributional Statistics of Chi Square Estimates After PA Factor Extraction

df	*N*	Mean	Variance	Skewness	Kurtosis
5	200	4.617 (4.529, 4.705)	10.555 (10.358, 10.753)	0.988 (0.952, 1.025)	2.823 (2.710, 2.936)
	500	4.218 (4.141, 4.294)	10.938 (10.831, 11.044)	0.835 (0.802, 0.867)	2.139 (2.064, 2.213)
	800	3.951 (3.870, 4.031)	11.648 (11.543, 11.753)	1.025 (0.673, 0.741)	1.881 (1.830, 1.932)
Expected		5	10	1.265	2.400
64	200	63.047 (62.577, 63.517)	124.755 (123.029, 126.482)	0.358 (0.348, 0.368)	0.188 (0.168, 0.212)
	500	62.474 (62.151, 62.798)	124.491 (123.153, 125.830)	0.369 (0.358, 0.380)	0.196 (0.172, 0.220)
	800	61.850 (61.586, 62.113)	123.081 (121.884, 124.277)	0.360 (0.350, 0.369)	0.168 (0.155, 0.196)
Expected		64	128	0.354	0.188
207	200	210.253 (209.179, 211.326)	420.916 (416.594, 425.238)	0.186 (0.176, 0.196)	0.052 (0.037, 0.068)
	500	205.273 (204.534, 206.013)	410.350 (406.743, 413.958)	0.195 (0.185, 0.205)	0.060 (0.043, 0.077)
	800	202.637 (202.005, 203.269)	403.110 (399.714, 406.506)	0.199 (0.190, 0.209)	0.057 (0.040, 0.074)
Expected		207	414	0.197	0.058
492	200	509.195 (507.36, 511.029)	1024.019 (1014.824, 1033.214)	0.132 (0.122, 0.142)	0.029 (0.014, 0.045)
	500	493.335 (491.908, 494.762)	1004.453 (995.697, 1013.209)	0.122 (0.127, 0.148)	0.044 (0.029, 0.059)
	800	488.925 (487.757, 490.093)	980.619 (973.332, 987.906)	0.142 (0.116, 0.135)	0.025 (0.010, 0.039)
Expected		492	984	0.128	0.024

Note. 95th confidence intervals are shown in parenthesis.

PsychOpen GOLD

Table 3 shows the outcomes related to PA extraction. Mean and variances were slightly underestimated, especially in large samples. Again, the estimates of skewness and kurtosis in general do not differ from the values expected in the population, and only in small models (5 degrees of freedom) and large samples (N = 800) the estimates are underestimating the expected values.

Rejection rates after PA factor extraction are shown in Table 4. The outcomes are quite similar to those obtained after ULS/MINRES extraction pattern. However, PA extraction obtained slightly better rejection rates than ULS/MINRES. Again, rejection rates improve when sample sizes are large.

Table 4

Rejection Rates After PA Factor Extraction

		Rejection rates		
df	*N*	**.100**	**.050**	**.001**
5	200	.085 (.081, .090)	.046 (.043, .048)	.012 (.011, .013)
	500	.076 (.073, .079)	.040 (.038, .042)	.009 (.009, .010)
	800	.073 (.070, .076)	.039 (.037, .041)	.009 (.008, .009)
64	200	.106 (.098, .114)	.058 (.053, .063)	.015 (.013, .017)
	500	.089 (.084, .093)	.046 (.043, .049)	.010 (.009, .011)
	800	.078 (.074, .081)	.039 (.036, .041)	.008 (.007, .009)
207	200	.164 (.152, .176)	.099 (.090, .107)	.030 (.027, .034)
	500	.103 (.096, .109)	.055 (.051, .059)	.013 (.012, .015)
	800	.176 (.074, .084)	.040 (.037, .043)	.009 (.008, .010)
492	200	.265 (.249, .281)	.175 (.162, .188)	.066 (.059, .073)
	500	.135 (.126, .144)	.077 (.071, .083)	.022 (.019, .024)
	800	.102 (.096, .109)	.055 (.050, .059)	.013 (.012, .014)

Note. 95th confidence intervals are shown in parenthesis.

Finally, RMSEA, CFI and NNFI goodness-of-fit indices were computed. As the model that was fitted to the sample data systematically corresponded to the true population model, the values of these indices should indicate in all cases that an acceptable model fit had been attained. Table 5 shows the mean of goodness-of-fit indices after ULS/MINRES factor extraction. As can be observed, a good model fit was always reported. In addition, as the sample size became larger, the goodness-of-fit values improved for all the indices.

Table 5

Mean of Goodness-of-Fit Indices After ULS/MINRES Factor Extraction When the Model Proposed in the Sample is Correct in the Population

df	N	RMSEA	CFI	NNFI
5	200	.0227 (.0222, .0233)	.9960 (.9959, .9961)	.9926 (.9924, .9929)
	400	.0140 (.0137, .0142)	.9981 (.9981, .9981)	.9969 (.9969, .9970)
	800	.0140 (.0103, .0106)	.9960 (.9986, .9986)	.9979 (.9979, .9980)
64	200	.0128 (.0123, .0134)	.9967 (.9966, .9968)	.9955 (.9953, .9957)
	400	.0081 (.0079, .0083)	.9984 (.9983, .9984)	.9980 (.9979, .9980)
	800	.0062 (.0061, .0064)	.9987 (.9987, .9988)	.9985 (.9985, .9985)
207	200	.0114 (.0109, .0119)	.9965 (.9963, .9966)	.9955 (.9952, .9957)
	400	.0065 (.0063, .0068)	.9984 (.9984, .9984)	.9981 (.9981, .9981)
	800	.0047 (.0046, .0049)	.9988 (.9988, .9988)	.9986 (.9986, .9987)
492	200	.0109 (.0105, .0114)	.9962 (.9960, .9964)	.9953 (.9950, .9955)
	400	.0057 (.0054, .0059)	.9984 (.9984, .9985)	.9982 (.9981, .9982)
	800	.0042 (.0041, .0044)	.9988 (.9988, .9988)	.9987 (.9987, .9987)

Note. 95th confidence intervals are shown in parenthesis.

Table 6

Mean of Goodness-of-Fit Indices After PA Factor Extraction When the Model Proposed in the Sample is Correct in the Population

df	N	RMSEA	CFI	NNFI
5	200	.0180 (.0174, .0186)	.9969 (.9968, .9969)	.9944 (.9942, .9946)
	400	.0101 (.0098, .0104)	.9984 (.9984, .9985)	.9976 (.9976, .9977)
	800	.0101 (.0073, .0077)	.9969 (.9987, .9988)	.9983 (.9983, .9983)
64	200	.0114 (.0108, .0119)	.9970 (.9969, .9971)	.9960 (.9958, .9962)
	400	.0066 (.0064, .0068)	.9985 (.9985, .9985)	.9982 (.9982, .9982)
	800	.0049 (.0047, .005)	.9988 (.9988, .9988)	.9987 (.9986, .9987)
207	200	.0111 (.0106, .0116)	.9965 (.9964, .9967)	.9956 (.9954, .9958)
	400	.0054 (.0052, .0056)	.9986 (.9985, .9986)	.9983 (.9983, .9983)
	800	.0037 (.0035, .0038)	.9989 (.9989, .9989)	.9988 (.9987, .9988)
492	200	.0121 (.0117, .0126)	.9958 (.9956, .996)	.9947 (.9945, .9949)
	400	.0051 (.0049, .0053)	.9985 (.9985, .9986)	.9983 (.9983, .9983)
	800	.0034 (.0033, .0036)	.9989 (.9989, .9989)	.9988 (.9988, .9988)

Note. 95th confidence intervals are shown in parenthesis.

PsychOpen GOLD

Table 6 shows the mean of goodness-of-fit indices after PA factor extraction. Once again, a good model fit was always reported. It must be pointed out that the values obtained here suggested a better fit than those obtained after ULS/MINRES extraction. In addition, the estimates did not seem so much influenced by the sample size, as it happened after ULS/MINRES extraction.

The conclusion of the first simulation study is that, when the correct model is fitted to the sample data, the distribution of the y statistic shows a reasonable adherence to the expected chi-square distribution, and the related goodness-of-fit indices can be safely interpreted.

Second Simulation Study

The second study is mainly concerned with assessing the power and sensitivity of our proposed statistic. To assess so, we replicated the first simulation study with two variations. First, the number of factors retained in the sample data was one less than the true number in the population. Second, only three population models were considered: the models with 64, 207 and 492 degrees-of-freedom in the population.

ULS/MINRES and PA extraction methods were used to extract the incorrect number of factors. As the sample model was not the one that existed in the population, goodness-of-fit indices derived from y should suggest an improper model adjustment. In addition, the values should worsen when the discrepancy between the sample and the population models increases. Thus, if the population model had four factors, and the sample model is adjusted for three factors, the misspecification is not so strong as if the population model had two factors, and the sample model is adjusted for one single factor. The values of the goodness-of-fit indices should be sensitive to the different degrees of misspecification.

RMSEA, CFI and NNFI goodness-of-fit indices were again computed. Table 7 shows the mean of the goodness-of-fit indices after ULS/MINRES factor extraction. As can be observed, an unacceptable model fit was always reported for all the indices, and the worse goodness-of-fit values were related to the solutions with the lowest degrees of freedom. It must be noted that, when the sample size was large, the indices reported better goodness-of-fit values, but these values were still farther away from the usual threshold values used in applied research for judging the fit as acceptable.

Table 8 shows the mean of goodness-of-fit indices after PA factor extraction. As can be observed, the values of the goodness-of-fit are quite similar to the ones obtained after ULS/MINRES extraction.

The conclusion of the second simulation study is that, when the sample model is incorrectly specified, the goodness of fit indices derived from the proposed y statistic are expected to detect that the proposed model is wrong under most of the conditions expected to occur in practice.

Table 7

Mean of Goodness-of-Fit Indices After ULS/MINRES Factor Extraction When the Model Proposed in the Sample is Incorrect in the Population

df	N	RMSEA	CFI	NNFI
64	200	.4324 (.4214, .4435)	.7877 (.7706, .8048)	.7780 (.7593, .7966)
	400	.2458 (.2355, .2561)	.4279 (.4017, .4541)	.3257 (.2949, .3566)
	800	.1855 (.1768, .1942)	.5984 (.5780, .6188)	.5984 (.5780, .6188)
207	200	.3251 (.3157, .3346)	.6384 (.6258, .6511)	.6302 (.6167, .6437)
	400	.1760 (.1674, .1846)	.5348 (.5105, .5591)	.4393 (.4101, .4686)
	800	.1305 (.1232, .1379)	.7252 (.7093, .7412)	.6689 (.6497, .6881)
492	200	.2559 (.2474, .2644)	.5028 (.4854, .5201)	.4947 (.4772, .5122)
	400	.1376 (.1303, .1450)	.6025 (.5804, .6247)	.5231 (.4965, .5496)
	800	.1028 (.0966, .1089)	.7644 (.7501, .7787)	.7173 (.7001, .7345)

Note. 95th confidence intervals are shown in parenthesis.

Table 8

Mean of Goodness-of-Fit Indices After PA Factor Extraction When the Model Proposed in the Sample is Incorrect in the Population

df	N	RMSEA	CFI	NNFI
64	200	.4131 (.4005, .4257)	.7665 (.7467, .7863)	.7568 (.7358, .7779)
	400	.2402 (.2295, .2509)	.4404 (.4132, .4675)	.3404 (.3085, .3724)
	800	.1723 (.6824, .6247)	.1817 (.7008, .6464)	.1817 (.7008, .6464)
207	200	.3199 (.3102, .3295)	.6259 (.6116, .6401)	.6160 (.6009, .6311)
	400	.1761 (.1674, .1848)	.5323 (.508, .5566)	.4364 (.4071, .4657)
	800	.1328 (.1256, .1400)	.7213 (.7056, .737)	.6641 (.6452, .6830)
492	200	.2570 (.2488, .2653)	.5005 (.4835, .5175)	.4916 (.4745, .5086)
	400	.1368 (.1295, .1442)	.6044 (.5822, .6267)	.5254 (.4987, .5520)
	800	.1064 (.1004, .1124)	.7560 (.7420, .7700)	.7072 (.6904, .7240)

Note. 95th confidence intervals are shown in parenthesis.

Implementation

The code file developed is the R script "Losefer.r". This script uses only native functions in R, so no packages need to be downloaded. In order to use it, researchers have to store participants' responses in a text file, update the name of the input file, and execute the script. The script is implemented to allow different extraction procedures (Principal Component Analysis, Centroid, and Principal Axes). With this example script, research-

ers can easily adapt the code to use other extraction methods. We made it available via the PsychArchives repository.

In addition, we implemented the full item selection proposal in our program to compute factor analysis, that can be downloaded free from the site psico.fcep.urv.cat/util-itats/factor. The computing is offered as a LOSEFER chi-square adjustment method, and can be computed with Principal Component Analysis, ULS/MINRES, MRFA and ML. In this software, the response format can be linear variables and graded response variables.

Illustrative Example

A set of six items from the Statistical Anxiety Test (Vigil-Colet et al., 2008) was used for the illustrative example. All the items correspond to the *Anxiety to Examination* subscale, and are responded on a 5-point graded format. A sample of 459 undergraduate students from the first course of a degree on psychology answered the test.

A Robust Unweighted Least Squares (RULS) solution was first computed, and the chi-square statistic derived from this method was scaled using the mean and variance adjustment. In addition, the statistic was also corrected using the method proposed in this document. Finally, ULS/MINRES, PAF, and Minimum Rank Factor Analysis (MRFA) solutions were computed, and the corresponding y statistics were obtained as proposed in this article.

Initial information was obtained by assessing the existing fit statistics that are not derived from the chi-square test. The output of Parallel Analysis suggested that the unidimensional solution was the most appropriate. The percentage of explained common variance (ECV) was 0.77, below the 0.80 cut-off most commonly used. Finally, the RMSR and GFI estimates were 0.06 and 0.99 respectively. Overall, these results suggest that a single dominant factor underlies the responses to these 6 items, but that this factor is not yet able to fully account for the inter-item correlations.

The chi-squared derived GOF statistics here were RMSEA, and CFI. In addition, the Non-Normed Fit Index (NNFI) was also obtained. In the ULS/MINRES, PAF, and MRFA cases, these indices were computed based on the g values obtained after Losefer adjustment. Table 9 shows a summary of the goodness of fit results.

Table 9

Goodness-of-Fit Statistics for SAS Illustrative Example

| | Extraction method | | | |
| | Corrected statistic with mean and variance adjustment | Empirically obtained statistic (Losefer approach) | | |
Goodness-of-fit statistic	RULS	ULS/MINRES	PAF	MRFA
Chi square value	59.129	62.501	58.091	51.531
RMSEA	.110	.114	.109	.102
CFI	.983	.981	.982	.985
NNFI	.971	.968	.970	.974

Note. $df = 9$.

The four methods compared in the study (RULS, ULS/MINRES, PAF, and MRFA) arrived at similar estimates of both the chi square statistic, and goodness-of-fit indices. This outcome could be expected, because they were assessing exactly the same factor solution. However, MRFA reported a lower chi square index, and the goodness-of-fit indices also suggest that the factor solution fitted better than that fitted by the other three extraction methods. As a single factor was extracted in all cases, the only source that can explain this difference is that the loading values estimated by MRFA gave rise to a reproduced correlation matrix that was closer to the observed correlation matrix.

Finally, the close-fit and power results based on the PA-Losefer solution (which seems to be the most widely agreed) are reported in Table 10.

Table 10

Test of Close Fit, and Power Assessment Illustrative Example

Test/Analysis
Test of Close Fit
RMSEA Estimate = 0.109
$p < 0.001$ ($df = 9$) for RMSEA < 0.05
Power Analysis Results
H_0: RMSEA = 0.05
H_1: RMSEA = 0.08
Beta = .54 ($df = 9$)

The chi-square based outcomes agrees with the initial measures of fit above but, as expected, seem to be more sensitive to detect model misspecification, especially the test

statistic itself and the RMSEA. Clearly, it cannot be accepted now from the first outcome in Table 10 that a single factor closely fits the scale data in this example. Perhaps still more important, however, the power results suggest that, in such a small model (only 9 degrees of freedom), the power for distinguishing a moderate misspecification from a close fit is still unacceptably low, and a much larger sample should have been collected for this purpose.

Discussion

Adjudging the appropriateness of an UFA solution is a multi-faceted process that goes far beyond goodness of model-data fit (e.g., Ferrando & Lorenzo-Seva, 2018). At the same time, however, we believe that the basis test of fit statistic must necessarily be part of this process. It (a) provides relevant information on its own, especially when accompanied by power information, and (b) is the basis for computing goodness of fit indices that can provide additional information regarding the approximate or the relative fit of the proposed solution.

So far, the chi-squared test of fit statistic is available for certain statistical (in Lawley, 1940's, terms) UFA estimation procedures that either are fully efficient or for which theoretical corrections that compensate for its lack of efficiency are available. The test statistic, however, is not available for more humble UFA procedures that are often referred to, slightly derogatorily, as "approximate". In certain applied scenarios, however, it happens that these approximate procedures show advantages that more than compensate for their lack of statistical efficiency. So, to derive a basis chi-squared test of fit that can be used with these procedures is an issue of clear interest (e.g., Harman & Jones, 1966).

In this article we have proposed and implemented a test statistic of this type that can be used with ordinary least squares UFA solutions, solutions which, in turn, can be based on both the linear and the non-linear FA model. Overall, we consider that the proposed statistic works quite well under most of the conditions considered in the simulation. As a summary, it (a) closely adheres to the expected distribution under the null hypothesis, (b) demonstrates power and sensitivity for detecting a wrongly specified solution (if enough sample is available), and (c) allows for meaningful goodness of fit indices, tests of close fit, and power estimates to be derived. We acknowledge indeed that we (partly) make use of existing methodology for variable transformation, and also that, at its initial stages, our proposal is based on previous developments, particularly the bootstrapped approach by Bollen and Stine (1992). From here on, however, we consider our proposal to be mostly a new contribution.

As any initial proposal, this has its share of limitations and points that deserve further study. Regarding limitations, we have avoided complex theoretical developments at the cost of intensive simulation. So, the procedure places strong computational demands and can be time consuming. However, the computing power of informatic equipment

that is nowadays available for researchers should help to make the proposal rather feasible.

As for points that require further study, to start with further intensive simulation about its functioning in a variety of conditions beyond those considered here are clearly warranted. On the other hand, it is still not clear at present which are, (a) the chi-square-based fit indices, (b) the appropriate thresholds for these indices, and (c) the close-fit-test or power specifications that work best with UFA solutions. So, most of what is proposed here must be considered as tentative and is expected to be updated as more information will be available.

In spite of the shortcomings noted above, we believe that this proposal has great interest, wide applicability, and will be very useful for the FA practitioner, and more so taking into account that is implemented as a resource in a free, widely known and user-friendly UFA program, as well as in two of the best known statistical programs at present.

Competing Interests: The lead author is a member of the editorial board of Methodology but played no editorial role in this particular article nor intervened in any form in the peer review process.

Supplementary Materials

The supplementary materials provided are the R code, data examples, and script for different data extraction methods (see Lorenzo-Seva & Ferrando, 2023 in the Index of Supplementary Materials below).

Index of Supplementary Materials

Lorenzo-Seva, U., & Ferrando, P. J. (2023). *Supplementary materials to "A simulation-based scaled test statistic for assessing model-data fit in least-squares unrestricted factor-analysis solutions"* [R code]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.12951

Lorenzo-Seva, U., & Ferrando, P. J. (2023). *Supplementary materials to "A simulation-based scaled test statistic for assessing model-data fit in least-squares unrestricted factor-analysis solutions"* [Data extraction scripts, data examples]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.12950

PsychOpen GOLD

References

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research, 21*(2), 205–229. https://doi.org/10.1177/0049124192021002004

Corrêa Ferraz, R., Maydeu-Olivares, A., & Shi, D. (2022). Asymptotic is better than Bollen-Stine bootstrapping to assess model fit: The effect of model size on the chi-square statistic. *Structural Equation Modeling: A Multidisciplinary Journal, 29*(5), 731–743. https://doi.org/10.1080/10705511.2022.2053128

Ferrando, P. J. (2021). Seven decades of factor analysis: from Yela to the present day. *Psicothema, 33*(3), 378–386. https://doi.org/10.7334/psicothema2021.24

Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema, 29*(2), 236–240. https://doi.org/10.7334/psicothema2016.304

Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement, 78*(5), 762–780. https://doi.org/10.1177/0013164417719308

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*(4), 521–532. https://doi.org/10.1007/BF02293811

Harman, H. H., & Jones, W. H. (1966). Factor analysis by minimizing residuals (Minres). *Psychometrika, 31*(3), 351–369. https://doi.org/10.1007/bf02289468

Lawley, D. N. (1940). VI.—The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, 60*(1), 64–82. https://doi.org/10.1017/S037016460002006X

Lee, T., Cai, L., & MacCallum, R. (2012). Power analysis for test of structural equation models. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 181–194). Guilford Press.

Lorenzo-Seva, U., & Ferrando, P. J. (2021). Not positive definite correlation matrices in exploratory item factor analysis: causes, consequences and a proposed solution. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(1), 138–147. https://doi.org/10.1080/10705511.2020.1735393

Muñiz, J., & Fonseca-Pedrero, E. (2019). Ten steps for test development. *Psicothema, 31*(1), 7–16. https://doi.org/10.7334/psicothema2018.291

Revelle, W. (2022). *How to use the psych package for factor analysis and data reduction.* Department of Psychology Northwestern University.

Vigil-Colet, A., Lorenzo-Seva, U., & Condon, L. (2008). Development and validation of the statistical anxiety scale. *Psicothema, 20*(1), 174–180. https://doi.org/10.1037/t62688-000

PsychOpen GOLD

Methodology is the official journal
of the European Association of
Methodology (EAM).

leibniz-psychology.org

PsychOpen GOLD is a publishing
service by Leibniz Institute for
Psychology (ZPID), Germany.

PsychOpen GOLD