An evaluation of the Nest Eigenvalue Sufficiency Test (NEST)

Pier-Olivier Caron[1]

[1]Université TÉLUQ

Presentation at the 10th European Congress of Methodology [EAM 2023] at Ghent University, Belgium

As part of the "Exploratory Factor Analysis" Symposium

Thursday July 13th

Preliminaries

Determining the number of factors to retain in an exploratory factor analysis is one of the methodological problems still open.

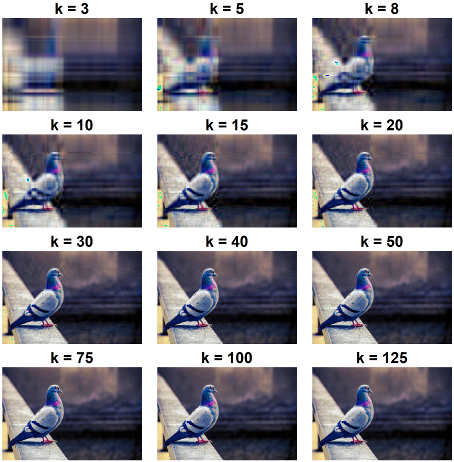
A plethora of stopping rules exist to answer this question. New ones are published every year.

One promising technique is the Next Eigenvalue Sufficiency Test (NEST; Achim (2017)), which shows excellent performance Brandenburg & Papenberg (in press).

However, it has not been systematically compared with its competitors.

Preliminaries

Purpose : get the most of the data, with the fewest factors.
Image taken from Caron (In preparation)



Objectives

To compare the performance of various recommended techniques (Auerswald & Moshagen, 2019) in addition to NEST.

- Next Eigenvalue Sufficiency Test (NEST) (Achim, 2017)
- Parallel Analysis (PA) (Horn, 1965)
- Sequential χ^2 model tests (SMT) (Lawley, 1940)
- Hull method (HULL) (Lorenzo-Seva et al., 2011)
- Empirical Kaiser Criterion (EKC) (Braeken & Assen, 2017)

Stopping rules

Parallel analysis (PA)

Resample the eigenvalues of a dataset with no factor (Identidy matrix) with the same characteristics as the target dataset (same number of variables and subjects).

The first empirical eigenvalues greater than those simulated are retained.

Next Eigenvalue Sufficiency Test (NEST)

Takes into account sampling error, like parallel analysis, but also the **sequence** of factor.

The test uses a correlation matrix containing the first k dimensions determined previously.

When $k = 0$, the test is equivalent to parallel analysis.

The k^{th} dimension is retained if the eigenvalue is greater than those simulated.

Empirical Kaiser criterion (EKC)

The distribution of eigenvalues asymptotically follows a Marchenko-Pastur distribution.

$\lambda_0 = (1 + \sqrt{p/n})^2$

for the 1^{st} and then corrected for next ones

$\lambda_j = \max\left(\frac{p \sum_{i=0}^{j} \lambda}{p-j-1}\left(1 + \sqrt{p/n}\right)^2, 1\right)$

The value of 1 is the minimum (like the Kaiser criterion).

The first k^{th} empirical eigenvalues above the criteria are retained.

Hull method (HULL)

Similar to Cattell's non-graphical variants, Hull's method attempts to find a kink in the eigenvalues.

Instead of using eigenvalues relative to the number of factors, Hull's method relies on goodness-of-fit indices (GFIs) relative to the degrees of freedom of the proposed model.

The last CFI is retained without improvement.

Sequential χ^2 model tests (SMT)

A sequential test of maximum likelihood (ML) estimation in which the covariance matrix of the model is equal to the sample covariance matrix.

We retain the first structure whose χ^2 is non-significant.

Method

Simulations

Simulations with synthetic factorial structures (Caron, 2016, 2019) in R (R Core Team, 2023).

The structures are

- 24 variables;
- 1 to 8 factors (`nf`);
- loadings ranging from .40 to .80 (`loadings`);
- inter-factor correlations from .00 to .30 (`Corr. Fact.`);
- sample sizes, 120, 240 and 480 (`n`);

In total, 360 scenarios are repeated 1000 times.

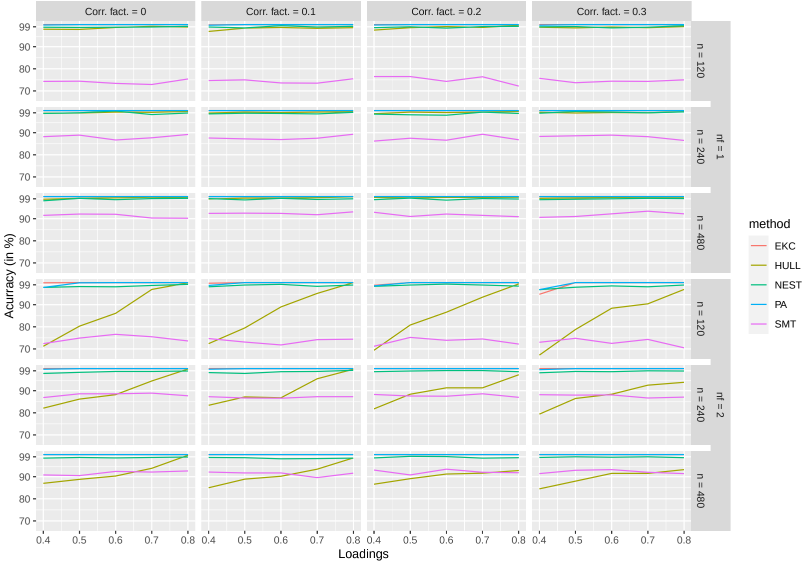
Performance

Performance is evaluated in terms of

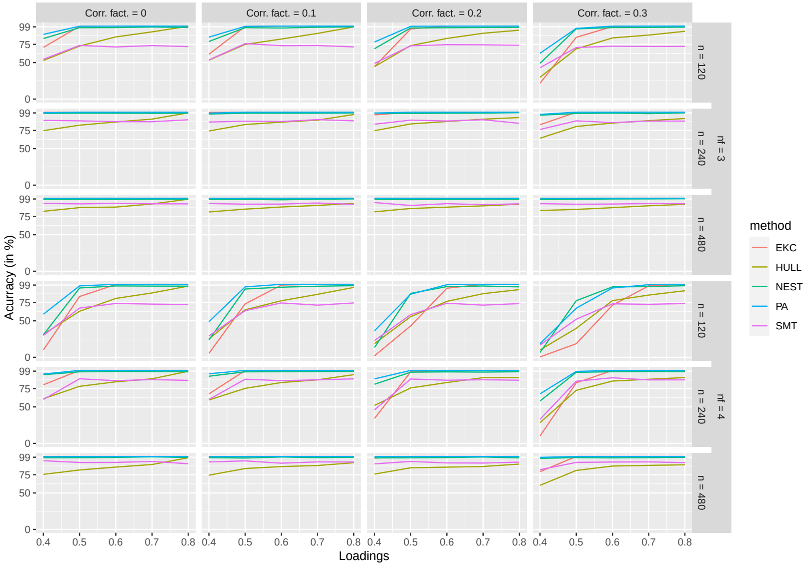
- accuracy (correct identification of dimensionality);
- bias (tendency to over- or under-estimate dimensionality).

Results

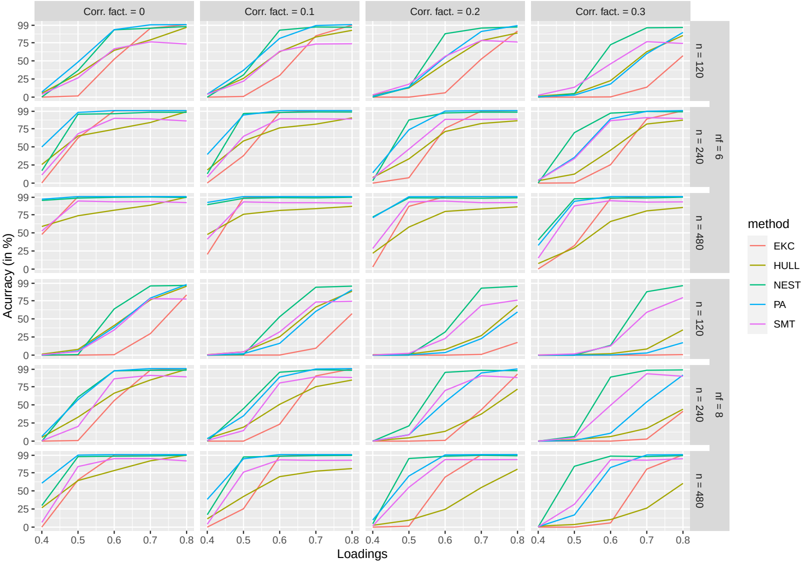
Very easy cases



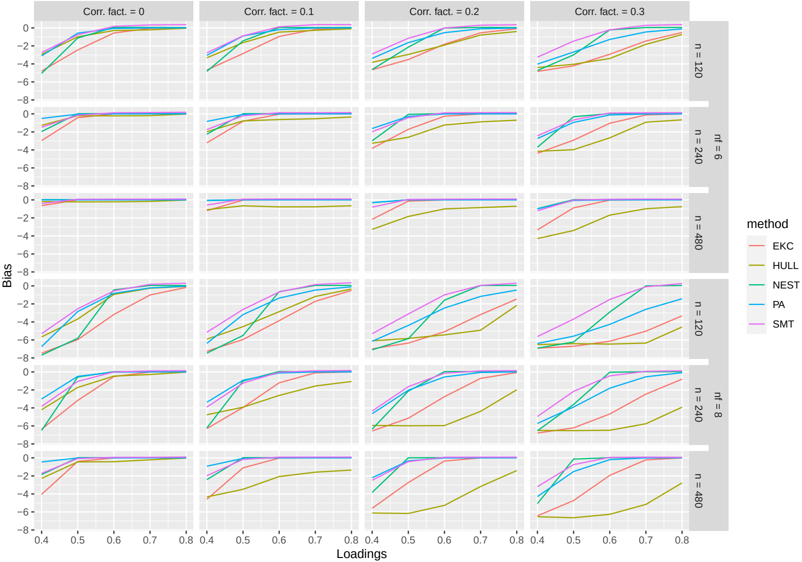
Easy and intermediate cases



Difficult and very difficult cases (power)



Difficult and very difficult cases (bias)



Limits

Limits of the simulations

- Factor structures with the same loadings.
- Equal eigenvalues for all factors.

Limits of NEST (and all techniques based on eigenvalues, such as PA and EKC)

- Suffers from the paralogism of the consequent assertion.

Conclusions

While most techniques do well in easy scenarios, NEST particularly stands out in difficult ones.

Test techniques with more realistic and varied factor structures.

Stay alert!

R Packages

Rnest (in development)

```
remotes::install_github(repo = "quantmeth/Rnest" )
```

Thanks to

Fonds d'aide à la recherche (FAR)

UNIVERSITÉ TÉLUQ

References I

Achim, A. (2017). Testing the number of required dimensions in exploratory factor analysis. *The Quantitative Methods for Psychology*, *13*(1), 64–74. https://doi.org/10.20982/tqmp.13.1.p064

Achim, A. (2020). Esprit et enjeux de l'analyse factorielle exploratoire. *The Quantitative Methods for Psychology*, *16*(4), 213–247. https://doi.org/10.20982/tqmp.16.4.p213

Achim, A. (2021). Determining the number of factors using parallel analysis and its recent variants: Comment on lim and jahng (2019). *Psychological Methods*, *26*(1), 69–73. https://doi.org/10.1037/met0000269

References II

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*(4), 468–491. https://doi.org/10.1037/met0000200

Braeken, J., & Assen, M. A. L. M. van. (2017). An empirical kaiser criterion. *Psychological Methods*, *22*(3), 450–466. https://doi.org/10.1037/met0000074

Brandenburg, N., & Papenberg, M. (in press). Reassessment of innovative methods to determine the number of factors: A simulation-based comparison of exploratory graph analysis and next eigenvalue sufficiency test. *Psychological Methods*. https://doi.org/10.1037/met0000527

References III

Caron, P.-O. (2016). A Monte Carlo examination of the broken-stick distribution to identify components to retain in principal component analysis. *Journal of Statistical Computation and Simulation*, *86*(12), 2405–2410. https://doi.org/10.1080/00949655.2015.1112390

Caron, P.-O. (2019). Minimum average partial correlation and parallel analysis : The influence of oblique structures. *Communications in Statistics - Simulation and Computation*, *48*(7), 2110–2117. https://doi.org/10.1080/03610918.2018.1433843

Caron, P.-O. (In preparation). *Méthodes quantitatives avec r*. mqr.teluq.ca

References IV

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Lawley, D. N. (1940). 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

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The hull method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*(2), 340–364. https://doi.org/10.1080/00273171.2011.564527

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/