Investigating the Accuracy of Parallel Analysis in Underextraction Conditions: A Monte Carlo Study

Justin M. Jones M.S. Candidate

Reagan D. Brown Ph.D.

Department of Psychological Sciences, Western Kentucky University

Introduction

- Identifying the correct number of factors when conducting a factor analysis is arguably the most important decision a researcher must make[9]
- Identifying the correct number of factors has important implications for structural validity, construct validity, content meaning, and the psychometric properties of scales and subscales[2]
- There are numerous methods to identify the number of factors from a factor analysis such as eigenvalue greater than one, Bartlett's test, scree test, and minimum average partial correlation (MAP), however many of these methods are either overly subjective or are inaccurate under a variety of conditions [3,7,9]
- Research has suggested that parallel analysis[8] is the most accurate method for identifying the number of factors from a factor analysis[1,5,6]
- Under certain conditions, parallel analysis has still been shown to overextract or underextract the number of factors[1]
- Parallel analysis may be inaccurate in the presence of high correlations between factors, low sample size, a high number of variables per factor, low factor loadings, or poorly defined factors[1,6]
- A suggested improvement for parallel analysis is to use a stricter criterion other than the 50th percentile eigenvalue, such as the 95th percentile eigenvalue[4]
- While this generally performs better than the 50th percentile eigenvalue, it is still prone to underextraction and overextraction[1]
- An additional margin criterion may further serve to improve the accuracy of parallel analysis

Research Question

- What effect will a margin criterion combined with a percentile criterion have on the accuracy of parallel analysis when determining the number of factors using a principal axis factoring method on a correlation matrix which engenders underextraction?

Hypothesis

- the 90th percentile criteria, in conjunction with an absolute margin, will prove to be more accurate in identifying the correct number of factors than all other criteria

Method

Population Generation

- A dataset consisting of 1,000,000 cases with scores on 12 variables will be generated to create a population matrix that engenders underextraction
- The population matrix will have a two factor structure whereby each factor will be defined by six variables
- Each variable will have a loading of .5 on its respective factor
- The correlation between factors will be .7

Procedure

- A sample of 240 or 480 cases will be randomly drawn from the population
- An exploratory factor analysis (common factor model) will be performed on the sample data
- A parallel analysis of 200 or 500 replications will be conducted on a sample of random data of the same size and with the same number of variables as the sample data
- The eigenvalues from the 50th, 90th, 95th, or 99th percentile from the parallel analysis will be compared to the eigenvalues obtained from the factor analysis of the sample data
- The number of factors will then be determined by applying the margin criterion (e.g., absolute or 10%). The number of factors will be defined as the highest factor in the sample data which has a positive eigenvalue greater than the corresponding random parallel analysis eigenvalue
- The process will be repeated 1000 times at which point the results will be compared against the known population values

Proposed Analysis

- The percentage of iterations which identified the correct number of factors will be calculated for each condition
- These percentages will then be compared across conditions

Proposed Results

Table 1

Expected Results

Eigenvalue Criterion	Margin Criterion	Expected Accuracy
50th Percentile	Absolute	≅ (+) 70%
50th Percentile	10%	≅ (+) 75%
90th Percentile	Absolute	≅ (+) 85%
90th Percentile	10%	≅ (+) 90%
95th Percentile	Absolute	≅ (-) 80%
95th Percentile	10%	≅ (-) 80%
99th Percentile	Absolute	≅ (-) 70%
99th Percentile	10%	≅ (-) 65%

Note: (+) signifies that a method is expected to overextract, (-) signifies that a method is expected to underextract. Percentages indicate expected accuracy of the criterion on identifying the correct number of factors in terms of number of iterations (out of 1000) where the correct number of factors were identified.

References

Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement*, *70*, 885–901. http://doi.org/10.1177/0013164410379332[1]

Conway, J. M., & Huffcutt, A. I. (2003). A Review and Evaluation of Exploratory Factor Analysis Practices in Organizational Research. Organizational Research Methods, 6(2), 147–168. http://doi.org/10.1177/1094428103251541[2]

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299. http://doi.org/10.1037/1082-989X.4.3.272[3]

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, *55*, 377–393. http://doi.org/10.1177/0013164495055003002[4]

Green, S. B., Thompson, M. S., Levy, R., & Lo, W.-J. (2015). Type I and type II error rates and overall accuracy of the revised parallel analysis method for determining the number of factors. *Educational and Psychological Measurement*, *75*, 428–457. http://doi.org/10.1177/0013164414546566[5]

Green, S. B., Xu, Y., & Thompson, M. S. (2017). Relative accuracy of two modified parallel analysis methods that use the proper reference distribution. *Educational and Psychological Measurement*,[6]

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*, 191–205. http://doi.org/10.1177/1094428104263675[7]

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. http://doi.org/10.1007/BF02289447[8]

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432. http://doi.org/10.1037/0033-2909.99.3.432[9]