Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. (3362 citations in Google Scholar as of 4/1/2016).

Who would have thought that a paper published over 50 years ago would still have an impact today. But this one still does! It must be due to the method being used — the simulation of the random sampling base in factor analysis. Over the years, the journal *Psychometrika* has been a haven for mathematical minutiae, but this paper was not one of them. It is to the credit of the current editor(s) that the recognition by the field is a primary criterion in making the journal so successful. This article has enjoyed almost 3,400 citations as of April, 2016, and the question of "why" comes to mind.

The solution to the burning question of the correct number of common factors in a matrix of correlations is easy. Henry Kaiser used to say that "it was so simple he solved it everyday before breakfast" (from J.L. Horn, Personal Communication, June 1977). Horn recognized that a solution was needed soon; he thought that he was proposing just a temporary solution, in place for others (Bartlett (1950) or Jöreskog (1967)) to develop a statistical procedure (which they soon did! For a full review, see Lawley & Maxwell (1963); Horn & Engstrom. 1979; Jöreskog & Sörbom (1979)). In Horn (1965) it was suggested that Guttman's latent-root-one lower bound estimate for the rank of a correlation matrix be accepted as a psychometric upper bound, following the proofs and arguments of Kaiser and Dickman. Kaiser shows that for a "principal component to have positive KR-20 internal consistency, it is necessary and sufficient that the associated eigenvalue be greater than one." (Kaiser, 1960, p. 6) and this was the source of the root-one criterion. But this was for a population, so Horn tried to add some sampling characteristics. The rank for a sample matrix should be estimated by subtracting out the component in the latent roots that can be attributed to sampling error, and least-squares "capitalization" on this error, in the statistical calculation of the correlations and the roots. He promoted a procedure, termed "Parallel Analysis" (or PA) by later authors, based on the generation of random variables for estimating the component that needs to be subtracted.

Horn became an expert in the classical techniques of common factor analysis. He generally asked why any elegant mathematical-statistical theory should be based on specific assumptions when we know these key assumptions are wrong and untestable. The important papers suggested that the number of common factors should not be determined simply using the well-known "eigenvalues greater than one" criterion defined by one of his favorite advisors, Henry Kaiser (1960; see Horn, 1965; and McArdle, 2007). But Horn suggested we make use of "computer simulation" techniques, mainly based on model assumptions that were random, and advocated their use whenever possible (see Horn & McArdle, 1980).

Horn (1965) determined the number of common factors by selecting the number of the eigenvalues of a correlation matrix that were greater than or equal to those provided by data computer-simulated with known characteristics. In this very simple idea, all that was needed was to generate "random data of similar size"; we could calculate the latent roots and vectors of these random data to provide a criterion tailored to the particular data set being analyzed. As an example, Horn (1965) implemented this procedure for $N = 297$ persons measured on $M = 65$ ability variables, from his doctoral dissertation research, where he thought he found evidence for $K = 16$ common factors by the Kaiser criterion but $K = 9$ by the PA criterion. Of course, we now recognize the need for statistical fluctuations in these latent roots, but these kinds of calculations were done almost 50 years ago! Horn's random variable approach has more recently been found to be the most accurate for determining the number of unrotated common factors (e.g., Ledesma & Valero-Mora, 2007; Montanelli & Humphreys, 1976; Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986; see also Dinno (2009) , Hayton (2009), and Courtney (2013) for recent evaluation, and hearty affirmation). This success also marks the beginning of Horn's fascination with the use of computer simulated data to solve the most complex problems in mathematical statistics (see Horn & McArdle, 1980).

## Extensions of the paper

Horn did not stop simulating factors. Other features of factor analysis that he investigated with simulation procedures included the well-known "rotation" problem (Horn, 1967). Here, he questioned the use of popular forms of factor rotation that were clearly founded on "substantive judgment" but that seemed to be considered "objective" largely because they were "blind" (p. 813). He knew that most substantive experiments were fairly small, and if the $N$ was reduced, factor loadings would average larger and appear to be more "respectable" (p. 819). Horn generated entirely random variables ($M = 74$) for this sized sample ($N = 300$), and found a clear willingness of famous faculty members to assign what seemed to be reasonable labels to several common factors that were no more than random rotations of simulated or "random" data. The reader was largely left to judge what this meant about the available factor analysis procedures, and what should be done next.

Horn wrote a great deal about his simulation studies of factor analysis, and focused considerable attention on "factor score estimates" (Horn & Miller, 1966; Wackwitz & Horn, 1971). He knew that factor analysis was based on observed scores and tried to provide a vehicle to calculate the unobserved (common factor) scores from these. Several factor score estimation approaches were examined using "exceedingly simple" (Wackwitz & Horn, 1971, p. 406) simulated data. These authors somewhat surprisingly concluded that "inexact procedures," such as the use of "unit-weighted salient variables," (but not just all unit weighted variables) led to the most replicable estimates for common factor scores. Although others have suggested further practical advances (e.g., Grice, 2001; Grice & Harris, 1998; DiStefano, Zhu & Mindnla, 2009), there is much less doubt about the procedures that can be used effectively. Indeed, the suggested use of a simple common factor score has become a part of our understanding of optimal factor scores today.

This computer simulation approach has certainly increased in popularity in psychometric and statistical research, and of course, Horn was not alone in understanding these topics (e.g., Ciesla, Cole, & Steiger, 2007; McArdle 2015; Muthén & Muthen, 2005). Simulation has been extended to estimation of standard errors (see Efron, 1979) and Bayesian estimation (e.g., Markov Chain Monte Carlo; see Ntzoufras, 2009), and even structural equation models (see Horn & McArdle, 1980; Muthén & Asparouhov, 2012). It appears that many scientists now agree that statistical analysis using simulations of otherwise highly complex systems is a viable approach to data analysis. I think this success is the reason for the recent citations.

To understand the popular move towards what Tucker and Lewis (1973) termed "confirmatory analysis," Horn first related his early work on PA to the development of the well-known chi-square test of the number of factors (Horn & Engstrom, 1979). Here he showed that his simulation approach (PA) matched the formal basis of Bartlett's (1950) chi-square test and Cattell's scree test (Cattell & Vogleman, 1977) — Horn was pleased. He next took on the arbitrary use of rotations in CFA; he was opposed to the term "confirmatory" for models that were "exploratory" at best (and see McArdle, 2012a). Horn suggested the idea of simulating complex systems that would not be possible to uniquely identify due to the selection of variables and persons (as in Horn & McArdle, 1980). Although he did not provide clear solutions to these problems, Horn was basically trying to point out that there were no available solutions!

According to Courtney (2013):

> In 2012 Ruscio and Roche introduced the comparative data (CD) technique in an attempt improve upon the PA method. In describing the method, the authors state that "rather than generating random datasets, which only take into account sampling error, multiple datasets with known factorial structures are analyzed to determine which best reproduces the profile of eigenvalues for the actual data" (p. 258). The authors explain that the strength of the technique is its ability to not only incorporate sampling error, but also the factorial structure and multivariate distribution of the items. Ruscio and Roche's (2012) simulation study determined that the CD technique outperformed all other methods aimed at determining the correct number of factors to retain. In their simulation study, the CD technique, utilizing Pearson correlations accurately predicted the correct number of factors 87% of the time. Although, it should be noted that simulated data did not involve more than five factors. Therefore, the applicability of the procedure to estimate factorial structures beyond five factors is yet to be tested. (p. 4).

The CD method was used in Horn and McArdle (1980).

## Notes on the Author

John Horn (1928–2006) was a pioneer in multivariate thinking and the application of multivariate methods to research on intelligence and personality. His key works on individual differences in the methodological areas of factor analysis and the substantive areas of cognition are reviewed here. John was also my mentor, teacher, colleague, and friend. It is tempting now to review John Horn's main contributions to the field of intelligence by highlighting his methods of factor analysis and his substantive debates about intelligence, but this is done elsewhere (in McArdle & Hofer, 2014).

As a leader in multivariate methodology, Horn tried to reach the incredible heights of his well-known mentor, Raymond Cattell. As illustrated here, John believed strongly in a multivariate scientific approach, and questioned the typical use of un-weighted sum scores as if they represented the best scores of the psychological constructs of interest. On a substantive basis, John believed that there were important individual differences among adults within the domains of cognition and personality (see Horn & Knapp, 1974; Horn & Donaldson, 1977). Although aspects of these debates linger on, for the most part, much can be said about John as an atypical person with an atypical background, but we will not emphasize this here (see McArdle, 2007, 2012b).

Some of John Horn's early comments on the methods of factor analysis are worth repeating especially the central concept of a "functional unity" (Horn, 1972, p. 161-162). He applied the ideas about a multivariate meta-theory to data on cognitive abilities, to create various testable hypotheses. This type of reasoning provides the basis for arriving at several substantive results on cognitive abilities, perhaps the most important being that Horn expanded on this initial work of his primary advisor, Raymond Cattell (see Horn, 1965; Horn & Cattell, 1966; 1967) to identify additional functional unities of primary mental abilities (see Woodcock,1989).

For these reasons I think that John Horn's major contributions to psychology, only some of which have been discussed here, continue to be ahead of his time, and have a profound influence on our thinking and critical approach to answering complex questions. His contributions to factor analysis and the structure of intelligence, the important methodological debates of the 1970s and 1980s regarding age and cohort effects and related issues of sample selectivity, the innovative ideas underlying his approach to evaluating state, trait and trait-change (1972), and his willingness and encouragement to engage in critical evaluation of fundamental ideas and accepted scientific approaches (i.e., the g-theory; see McArdle, 2012b) are and will remain important contributions. Through his research and teaching he forced people to question popular assumptions, evaluate all the data available, and consistently challenged us to think longer, harder, and better. His work will continue to inspire important research in the fields of multivariate analysis and human cognitive abilities for decades to come. The interested reader can see McArdle and Hofer (2014).

## References


Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, *3*, 77–85.

Cattell, R. B. (1957). Personality and motivation structure and measurement. World Books: New York.

Cattell, R. B. (1966). Psychological theory and scientific method. In R. B. Cattell (Ed.), Handbook of multivariate experimental psychology (pp. 1–18). Chicago: Rand McNally & Co.

Cattell, R.B. (1971). Abilities: Their structure, growth and action. Boston: Houghton-Mifflin.

Cattell, R. B., & Vogelman, S. (1977). A comprehensive trial of the scree and kg criteria for determining the number of factors. *Multivariate Behavioral Research*, *12*. 289–325.

Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. *Practical Assessment, Research & Evaluation*, *18(8)*, 1–14.

Dinno, A. (2009). Exploring the sensitivity of Horn's parallel analysis to the distributional form of random data. *Multivariate Behavioral Research*, *44*, 362–388.

DiStefano, C., Zhu, M., & Mindnla, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, & Evaluation*, *14(20)*, 1–11

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Mathematical Statistics*, *7*, 1–26.

Grice, J. W. (2001). A comparison of factor scores under conditions of factor obliquity. *Psychological Bulletin*, *6*, 67–83.

Grice, J. W., & Harris, R. J. (1998). A comparison of regression and loading weights for the computation of factor scores. *Multivariate Behavioral Research*, *33*, 221–247.

Hayton, J. C. (2009). Commentary on "Exploring the Sensitivity of Horn's Parallel Analysis to the Distributional Form of Random Data." *Multivariate Behavioral Research*, *44*, 389–395.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185.

Horn, J. L. (1967). On subjectivity in factor analysis. *Educational and Psychological Measurement*, *27*, 811–820.

Horn, J. (1972). State, trait and change dimensions of intelligence. *British Journal of Educational Psychology*, *42*, 159–185.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*, 253–270.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26*, 107–129.

Horn, J. L., & Donaldson, G. (1977). Faith is not enough: A response to the Baltes-Schaie claim that intelligence does not wane. *American Psychologist*, *32*, 369–373.

Horn, J. L., & Engstrom, R. (1979). Cattell's scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. *Multivariate Behavioral Research*, *14*, 283–300.

Horn, J. L., & Knapp, J. R. (1974). Thirty wrongs do not make a right: Reply to Guilford. *Psychological Bulletin*, *81*, 502–504.

Horn, J.L. & McArdle, J.J. (1980). Perspectives on Mathematical and Statistical Model Building (MASMOB) in Research on Aging. In L. Poon (Ed.), Aging in the 1980's: Psychological Issues. Washington, D. C.: American Psychological Association, 503–541.

Horn, J. L. & McArdle, J.J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R. MacCallum, (Eds.). Factor Analysis at 100 years (pp. 205–247). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Horn, J. L., & Miller, W. C. (1996). Evidence on the estimation of factor scores. *Educational and Psychological Measurement*, *26*, 617–622.

Jöreskog, K. G., & Sörbom, D. (1979). Advances in factor analysis and structural equation models. Cambridge, MA: Abt.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151.

Lawley, D.N. & Maxwell, A.E. (1963). Factor analysis as a statistical method. London: Butterworth

Ledesma, R. & Valero-Mora, P. (2007) Determining the Number of Factors to Retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis. *Practical Assessment, Research & Evaluation*, *12(1)*, 1–11.

McArdle, J. J. (2007). John L. Horn (1928–2006). *American Psychologist*, *62*, 596–597.

McArdle, J. J. (2012a). Ethical issues in factor analysis. In A. T. Panter & S. K. Sterba (Eds.), Handbook of ethics in quantitative methodology (pp. 313–340). New York: Taylor & Francis.

McArdle, J.J. (2012b). Testing the Idea of General Intelligence. *F&M Scientist*, *1*, 27–66.

McArdle, J.J (2014). Loehlin's original models and model contributions. *Behavior Genetics*, *44*, 5–11.

McArdle, J.J., & Hofer, S.M. (2014). Fighting for Intelligence: A Review of the Academic Research of John L. Horn. *Multivariate Behavioral Research*, *49*, 1–16. Winner of the Tanaka Award as the best MBR Paper, Oct. 2015.

Montanelli, R. G., Jr. & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika*, *41*, 341–348.

Muthén, B. & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. Psychological Methods, 17, 313-335.

Muthén, L. K. & Muthén, B. O. (2005). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 599–620.

Ntzoufras, I. (2009). Bayesian modeling using WinBUGS. New York: Wiley.

Spearman, C. E. (1904). 'General intelligence,' objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In Goffen, R. D. and Helms, E., (Eds.), Problems and Solutions in Human Assessment: Honoring Douglas N. Jackson at Seventy (pp. 41-71). New York: Springer.

Wackwitz, J. H., & Horn, J. L. (1971). On obtaining the best estimates of factor scores within an ideal simple structure. *Multivariate Behavioral Research*, *6*, 389–408.

Woodcock, R.W. (1998). Extending Gf-Gc Theory Into Practice. In McArdle, J.J. & Woodcock, R. W. (Eds.). Human Abilities in Theory and Practice (pp. 137–156). Mahwah, NJ: Erlbaum.

Zwick, W.R. & Velicer, W.F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432–442.

John J. McArdle, Department of Psychology, University of Southern California

Acknowledgements: We thank the National Institute on Aging (AG0713720 to the author) for support of this research and to Carol Prescott (University of Southern California) for helpful comments on the draft. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.