Some Contrasts Between Maximum Likelihood Factor Analysis and Alpha Factor Analysis

Henry F. Kaiser
University of California, Berkeley

Gerhard Derflinger
University of Economics and Business Administration, Vienna

The fundamental mathematical model of Thurstone's common factor analysis is reviewed. The basic covariance matrices of maximum likelihood factor analysis (MLFA) and alpha factor analysis (AFA) are presented. Putting aside the principles on which they are based, these two methods are compared in terms of a number of computational and scaling contrasts following from the application of their respective developments. The paper concludes with a discussion of the number-of-factors problem, the weighting problem in MLFA and AFA, and possible bases for a choice between the two. *Index terms: alpha factor analysis, common factor analysis, maximum likelihood factor analysis, number of common factors, scaling and weighting in common factor analysis.*

Thurstone's common factor analysis postulates p attributes (random variables) z_j, where each z_j is partitioned into two parts, c_j (the common part) and y_j (the unique part):

$$z_j = c_j + y_j \qquad (j = 1, 2, \ldots, p) \quad . \tag{1}$$

Means for all variables are 0:

$$\mu(\mathbf{z}) = \mu(\mathbf{c}) = \mu(\mathbf{y}) = 0 \quad , \tag{2}$$

where the μ operator takes the mean of the parenthetical expression and $\mathbf{z}$, $\mathbf{c}$, and $\mathbf{y}$ are p-vector attributes of the z_j, c_j, and y_j, respectively, and $\mathbf{0}$ is the p null vector. Traditionally, the variances of all the z_j are assumed to be 1, that is, the z_j are standardized. This standardization is only a convenience and is not necessary for the methods under consideration here; they are scale free.

Next it is assumed that

$$\mu(\mathbf{cy}') = \mu(\mathbf{yc}') = \mathbf{0} \quad , \tag{3}$$

where $\mathbf{0}$ is the $p \times p$ null matrix. That is, all common parts are uncorrelated with all unique parts. Additionally, it is assumed that

$$\mu(\mathbf{yy}') = \mathbf{U}^2 \quad , \tag{4}$$

where $\mathbf{U}^2$ is the diagonal matrix of the uniquenesses (unique part variances); different unique parts do not covary.

From the above, it follows that if the z_j are standardized,

$$\mu(\mathbf{cc}') = \mathbf{R} - \mathbf{U}^2 \quad , \tag{5}$$

where $\mathbf{R} = \mu(\mathbf{zz}')$ is the intercorrelation matrix of the z_js. The matrix $\mathbf{R} - \mathbf{U}^2$ is the covariance matrix of the common parts, often called the "reduced" correlation matrix. Let

$$\text{diag}(\mathbf{R} - \mathbf{U}^2) = \mathbf{H}^2 \quad , \tag{6}$$

where $\mathbf{H}^2$ is the diagonal matrix of the communalities (common part variances). Clearly, $\mathbf{H}^2 + \mathbf{U}^2 = \mathbf{I}$.

The c are then factored into q common factors $\mathbf{x}$:

$$\mathbf{c} = \mathbf{Fx} \quad , \tag{7}$$

where $\mathbf{F}$ is a $p \times q$ matrix of factor loadings, and

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 14, No. 1, March 1990, pp. 29–32

0146-6216/90/010029-04$1.45

q ($q < p$) is the rank of $\mathbf{R} - \mathbf{U}^2$. (As will be seen, q is only the "approximate" rank of $\mathbf{R} - \mathbf{U}^2$.)

Lawley (1940), in his classic paper on maximum likelihood factor analysis (MLFA), required a principal-axes decomposition of $\mathbf{U}^{-1}(\mathbf{R} - \mathbf{U}^2)\mathbf{U}^{-1}$, implying that the y_j, the unique parts, are standardized. This is easily seen by writing $\mathbf{U}^{-1}(\mathbf{R} - \mathbf{U}^2)\mathbf{U}^{-1} = \mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1} - \mathbf{I}$ (see also Bentler, 1968; Rao, 1955). This development is scale free: The same solution is obtained regardless of the scaling of the z_j. This is a characteristic of substantial theoretical significance.

Kaiser and Caffrey (1965), in their paper on alpha factor analysis (AFA), required a principal-axes decomposition of $\mathbf{H}^{-1}(\mathbf{R} - \mathbf{U}^2)\mathbf{H}^{-1}$, implying that the c_j, the common parts, are standardized. This development is also scale free. Alpha factor analysis is derived from finding uncorrelated common factors x_s ($s = 1, 2, \ldots, q$), each of which successively has maximum generalizability in the coefficient-alpha sense, a psychometric measure of reliability in the generalized Kuder-Richardson sense. McDonald (1970) had some theoretical reservations regarding the Kaiser-Caffrey development in terms of coefficient alpha; Kaiser and Caffrey might have been better advised to have presented their development of AFA less pretentiously with regard to the scale-free property of AFA, which is not inconsequential.

But Kaiser and Caffrey, in comparing MLFA and AFA, stated:

> . . . if one is devoted to the principle of maximum likelihood, MLFA is an answer, and that is that; while if one is devoted to the principle of maximum generalizability in the sense of coefficient alpha, AFA is an answer, and that is that. On the other hand, it may be argued that no one should accept either of these principles as a virtue in itself, and one may properly inquire as to whether either of these principles leads to results which consistently appear to have real virtues. (p. 9)

Rather than discussing the relative merits of the principle of maximum likelihood and the principle of maximum generalizability, this paper examines some contrasts which follow from the application of these principles.

Computational Considerations

Five contrasts involving computational considerations are:

1. If $\mathbf{R}$ is singular (as is often true in practice), MLFA incurs a Heywood Case, whereas there is no problem with AFA. (In MLFA the Heywood Case forces a division by 0.)
2. Even with a positive definite sample $\mathbf{R}$, the Heywood Case arises about 50% of the time in MLFA (Bentler & Lee, 1975) whereas there is no problem with AFA. The complementary problem with AFA (dividing by 0) occurs when there is a zero communality. This attribute must be removed prior to the analysis. But this is as it should be—zero communality occurs if and only if the offending attribute correlates exactly 0 with all the $p - 1$ other attributes—and it is clear that this attribute should be removed. (Of course, in practice this will occur with probability 0.)
3. In AFA, problems with convergence arise only if one or more communalities are very small; the corresponding attributes should be removed. On the other hand, MLFA will be numerically unstable with a communality near 1, forcing the use of a special algorithm in this case (Jöreskog, 1977).
4. Both methods are computationally feasible (Derflinger, 1968, 1979, 1984; Jöreskog, 1977). Extensive numerical experience suggests that for a given number of factors, AFA computes about twice as fast as MLFA.
5. In MLFA, successive approximations to the communalities often seem to wander aimlessly, whereas in AFA, successive approximations to the communalities apparently converge on the correct values with little meandering.

Some of the above computational contrasts are based only on the present authors' experience. This experience is extensive but unsystematic; the reader should beware of the interpretation given unsystematic experience, no matter how extensive. Unfortunately, a large search of the literature does not reveal explicit numerical contrasts of MLFA and AFA such as those described above.

It is well known that squared multiple correlations (SMCs) of a given attribute on the remaining $p - 1$ attributes are lower bounds to the communalities (Guttman, 1940). MLFA communalities often bear little apparent relationship to their corresponding SMCs; however, AFA communalities are highly correlated with their corresponding SMCs, a psychometric result. As Kaiser and Caffrey repeatedly pointed out, AFA is a psychometric method whereas MLFA is not; hence this result is not surprising.

Scaling Considerations

Two contrasts involving scaling considerations are:

1. The natural scaling for AFA (in the metric of the common parts) is correct for rotation and transformation problems (Kaiser, 1958; Kaiser & Dickman, 1959). This is not true for MLFA's scaling in the metric of the unique parts.
2. Glass (1966) proved that AFA is invariant under corrections for attenuation (unreliability). This does not occur for MLFA; for it to occur, it is necessary to scale in the metric of the common parts (AFA). At least to the psychometrician, this is a property of substantial importance.

Additionally, MLFA assumes multivariate normality. There are no distribution assumptions for AFA.

Appropriate Number of Common Factors

With AFA, the Kaiser-Guttman Rule for the number of common factors to retain—eigenvalues of $\mathbf{R}$ or $\mathbf{H}^{-1}(\mathbf{R} - \mathbf{U}^2)\mathbf{H}^{-1}$ greater than 1.0—is typically used. This is often appropriate or nearly appropriate if the interpretation of "appropriate number of common factors to retain" is the number—less than the total number—which responds optimally to the rotation problem. For the number of common factors, MLFA sometimes may use its significance test, which is dubious at best (Hotelling, 1933; Rao, 1955) and is preposterous at worst (Kaiser, 1960, 1976). But in MLFA it is not necessary to determine the number of common factors with MLFA's significance test (Kaiser, 1974).

For a thoughtful scientist, for any universe of scientific content the *total* number of common factors is arbitrarily large (in contrast to the *appropriate* number of factors to retain, the approximate rank of $\mathbf{R} - \mathbf{U}^2$). Surely, anyone will avoid testing for the statistical "significance" of a common factor when this significance is a function of irrelevancies such as the sample size n of entities, the arbitrary level of significance adopted, and the power of the statistical test. This does not imply that AFA, with the Kaiser-Guttman Rule, determines the appropriate number of common factors exactly, whereas MLFA with its significance test does not; rather, for this crucial unsolved problem, AFA is usually closer to the appropriate number of common factors, whereas MLFA usually is not.

Weighting

In MLFA the original attributes z_j, their common parts c_j, and their unique parts y_j are weighted by $\mathbf{U}^{-1}$. This seems appropriate because an original attribute with small residuals will be weighted more heavily. However, in AFA the basic attributes z_j, c_j, and y_j are weighted by $\mathbf{H}^{-1}$, indicating that an original attribute z_j with small communalities will be weighted more heavily, which seems counterintuitive. This observation was first made by McDonald (1968) and is espoused quite generally throughout the psychometric community as a principal justification for preferring MLFA to AFA.

Let it be suggested diffidently that this distinction may miss a point of possibly greater importance. Consider the question: What attributes are being studied in Thurstone's common factor analysis? One answer is that it is not the original attributes z_j, but rather their common parts c_j, that are being studied. Strictly speaking, the conclusion of such a study would not be expressed as, for example, "the loading of the (original) attribute Numerical Judgment on the Spatial factor is .26"; rather, it would be expressed as "the loading of *the common part* of the attribute Numerical Judgment on the Spatial factor is .26." Common factor analysis is primarily concerned with the c_j (the common parts), not the original z_j. Therefore, how the z_j are weighted is only of secondary concern. Primarily, the common parts c_j are studied, and because their "true" metric is unknown, they are assigned equal variance—hence AFA.

The distinction of the preceding paragraph is not a matter of fact; it is one of opinion, or point of view. The viewpoint defended above is a minority view, but not one of insignificance (personal communications: Tukey, 1962; Guttman, 1964; Novick, 1984).

Conclusions

Fundamentally, the choice between MLFA and AFA rests on whether the investigator considers Thurstone's common factor analysis as a statistical method or a psychometric method; if statistical then MLFA, if psychometric then AFA. A review of the material above suggests that common factor analysis is first, at least, a psychometric procedure. Furthermore, AFA treats the number-of-factors question more sensibly, perhaps weights the basic variables of common factor analysis more in accordance with that which is important, and is numerically better behaved.

References

Bentler, P. M. (1968). Alpha-maximized factor analysis (alphamax): Its relation to alpha and canonical factor analysis. *Psychometrika, 33,* 335–345.

Bentler, P. M., & Lee, S. Y. (1975). *Maximum likelihood factor analysis with unique solutions.* Unpublished manuscript, Department of Psychology, University of California, Los Angeles.

Derflinger, G. (1968). Neue Iterationsverfahren in der Faktorenanalyse. *Biometrische Zeitschrift, 10,* 58–75.

Derflinger, G. (1979). A general computing algorithm for factor analysis. *Biometrische Zeitschrift, 21,* 25–38.

Derflinger, G. (1984). A loss function for alpha factor analysis. *Psychometrika, 49,* 325–330.

Glass, G. V. (1966). Alpha factor analysis of infallible variables. *Psychometrika, 31,* 545–561.

Guttman, L. (1940). Multiple rectilinear prediction and the resolution into components. *Psychometrika, 5,* 75–99.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24,* 417–441, 498–520.

Jöreskog, K. G. (1977). Factor analysis by least squares and maximum likelihood methods. In K. Enslein, H. Ralston, & H. S. Wilf (Eds.), *Statistical methods for digital computers.* New York: Wiley.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23,* 187–200.

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

Kaiser, H. F. (1974). A computational starting point for Rao's canonical factor analysis: Implications for computerized procedures. *Educational and Psychological Measurement, 34,* 691–692.

Kaiser, H. F. (1976). Review of Lawley and Maxwell's *Factor analysis as a statistical method* (2nd ed.). *Educational and Psychological Measurement, 36,* 586–589.

Kaiser, H. F., & Caffrey, J. (1965). Alpha factor analysis. *Psychometrika, 30,* 1–14.

Kaiser, H. F., & Dickman, K. W. (1959). Analytic determination of common factors. *American Psychologist, 14,* 425.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, 60,* 64–82.

McDonald, R. P. (1968). A unified treatment of the weighting problem. *Psychometrika, 33,* 351–381.

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 23,* 1–21.

Rao, C. R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika, 20,* 93–111.

Thurstone, L. L. (1938). *Primary mental abilities* (Psychometric Monograph No. 1). Chicago: University of Chicago Press.

Acknowledgments

The research reported in this paper was supported in part by the Office of the Dean, Graduate School of Education, University of California, Berkeley, in part by the Committee on Research, University of California, Berkeley, and in part by the Wiener Hochschuljubiläumsstiftung, Vienna. The penultimate draft of this paper was prepared while the first author was a guest professor at the University of Economics and Business Administration, Vienna.

Author's Address

Send requests for reprints or further information to Henry F. Kaiser, Graduate School of Education, University of California, Berkeley CA 94720, U.S.A.