PSYCHOMETRIKA—VOL. 31, NO. 4
DECEMBER, 1966

RESOLUTION OF THE HEYWOOD CASE IN THE MINRES SOLUTION*

HARRY H. HARMAN

EDUCATIONAL TESTING SERVICE

AND

YOICHIRO FUKUDA

PLANNING RESEARCH CORPORATION

In the course of developing the minres method of factor analysis the troublesome situation of communalities greater than one arose. This problem—referred to as the generalized Heywood case—is resolved in this paper by means of a process of minimizing the sum of squares of off-diagonal residuals. The resulting solution is superior to the otherwise very efficient original minres method without requiring additional computing time.

From the earliest days in factor analysis a strange anomaly appeared now and then: while conditions for a solution might be satisfied, it sometimes happened that one of the resulting communalities was greater than unity. In other words, a factor solution might reproduce the observed correlations perfectly but lack the basic requirement that the communalities be positive numbers between 0 and 1. A solution that otherwise is satisfactory but produces a communality greater than one is known as a "Heywood case" [3]. This phenomenon was first discussed in connection with the Spearman two-factor solution, but even when a correlation matrix yields a suitable factor solution with several common factors for which one of the communalities exceeds unity, such a solution is referred to as a generalized Heywood case [1, p. 125].

1. *The Problem*

In a recent paper [2], the very efficient minres method of factor analysis was introduced. Unfortunately, this method exhibited the generalized Heywood case; though the instances were infrequent, they were disturbing nonetheless. For the factor model,

$$z_j = \sum_{p=1}^{m} a_{jp}F_p + d_jU_j\,, \qquad (j = 1, 2, \cdots, n) \tag{1}$$

*Both authors were with the System Development Corporation when this work was done.

or, in matrix notation,

$$z = Af + Du, \tag{1'}$$

the minres solution requires the minimization of off-diagonal residual correlations, the objective function being:

$$f(A) = \sum_{k=j+1}^{n} \sum_{j=1}^{n-1} \left(r_{jk} - \sum_{p=1}^{m} a_{jp} a_{kp} \right)^2. \tag{2}$$

This function is minimized for a specified m; from the resulting factor matrix A the diagonal matrix of communalities

$$H = (h_j^2) = \left(\sum_{p=1}^{m} a_{jp}^2 \right) \tag{3}$$

is obtained as a by-product.

In the original version of minres no further restraint is placed on the matrix A, and occasionally a solution is obtained for which one of the communalities exceeds unity. When a minres solution is obtained, satisfying (2) but with a Heywood case, the factor loadings for the unruly variable may be adjusted so that its communality is exactly one, without disturbing any of the other factor loadings. Forcing an excessive communality back to unity may be a solution of expediency, but it is not very elegant from a mathematical point of view.

The present paper develops a mathematical programming procedure, by introducing side conditions on the objective function (2), such that the resulting minres solution cannot involve a Heywood case. Specifically, the revised factor-loadings matrix A is obtained by minimizing (2) under the constraints

$$h_j^2 = \sum_{p=1}^{m} a_{jp}^2 \leqq 1 \qquad (j = 1, 2, \cdots, n). \tag{4}$$

The most efficient computing procedure for the original minres solution was found to be a variant of the Gauss–Seidel process [2, pp. 361–63]. In this method, the objective function is minimized by considering successive approximations of rows of factor loadings, making the changes or displacements in only one row of A at a time. Thus, for any row j in A an increment ϵ_p is added to each loading so that the new loadings may be written:

$$b_{jp} = a_{jp} + \epsilon_p \qquad (p = 1, 2, \cdots, m). \tag{5}$$

The impact on the objective function of replacing the a_{jp} by b_{jp} may be represented by

$$f_j = \sum_{\substack{k=1 \\ k \neq j}}^{n} \left(r_{jk} - \sum_{p=1}^{m} a_{kp} b_{jp} \right)^2 \qquad (j \text{ fixed}), \tag{6}$$

which is to be minimized subject to

$$\sum_{p=1}^{m} b_{ip}^2 \leqq 1, \tag{7}$$

i.e., the new values of the factor loadings must satisfy the constraints (4) as well. At this stage of the process, the r_{ik} and the a_{kp} are known, and only the b_{ip} may vary.

2. *Mathematical Solution to the Problem*

If the minimum of f_i is obtained at a point $(b_{i1}, b_{i2}, \cdots, b_{im})$ that belongs to the region defined by (7) there is no problem, and the original minres solution is satisfactory. If the point does not belong to the region, then the problem becomes complicated, primarily because of the inequality in the side condition. This inequality may be removed by means of the following:

THEOREM 1. *If the minimum of f_i is attained at a point outside of the region defined by* (7), *then a minimum of f_i under the constraint* (7) *will be attained at a boundary point of the region, so that the constraint may be replaced by*

$$\sum_{p=1}^{m} b_{ip}^2 = 1. \tag{8}$$

The proof can best be developed by rewriting the quadratic form in (6) as follows:

$$f_i = b'Wb + \sum_{p=1}^{m} 2v_p b_{ip} + K, \tag{9}$$

where b is a column vector of the unknown variables b_{ip} ; $W = (w_{pq})$, v_p , and K are constants determined from the known r_{ik} and a_{kp} . Since the first term in (9) is the square of a linear combination of the unknown b_{ip} , the symmetric matrix W is positive definite. Consequently, there exists an orthogonal matrix Q such that

$$Q'WQ = \Lambda \tag{10}$$

where Λ is a diagonal matrix whose diagonal elements, λ_p^2 $(p = 1, 2, \cdots, m)$, are the eigenvalues of W. Then the transformation

$$b = Q\hat{b} \tag{11}$$

diagonalizes the quadratic form $b'Wb$; (9) becomes

$$f_i = \sum_{p=1}^{m} \lambda_p^2 \hat{b}_{ip}^2 + \sum_{p=1}^{m} 2\hat{v}_p \hat{b}_{ip} + K \tag{12}$$

with the constraint (7) being replaced by a similar expression with carets over the b's.

The expression (12) may be simplified by completion of the square, namely

$$f_i = \sum_{p=1}^{m} (\lambda_p \hat{b}_{ip} - \lambda_p u_p)^2 + \hat{K}. \tag{13}$$

Then, by making the change of variables

$$x_p = \lambda_p \hat{b}_{ip} \qquad (p = 1, 2, \cdots, m), \tag{14}$$

and similarly by replacing the constants

$$\xi_p = \lambda_p u_p \qquad (p = 1, 2, \cdots, m), \tag{15}$$

the function to be minimized becomes

$$f_i = \sum_{p=1}^{m} (x_p - \xi_p)^2 + \hat{K}, \tag{16}$$

subject to the constraint

$$\sum_{p=1}^{m} \frac{x_p^2}{\lambda_p^2} \leqq 1. \tag{17}$$

In this simplified form it is evident that f_i is the sum of a constant ($\hat{K}$) and a square of the distance between a fixed point $(\xi_1, \xi_2, \cdots, \xi_m)$ and a variable point $(x_1, x_2, \cdots, x_m)$ belonging to the region defined by (17). Then, minimization of f_i is equivalent to locating a point satisfying (17) that is at the minimum distance from the given point $(\xi_1, \xi_2, \cdots, \xi_m)$.

If the given point belongs to the region, i.e.,

$$\frac{\xi_1^2}{\lambda_1^2} + \frac{\xi_2^2}{\lambda_2^2} + \cdots + \frac{\xi_m^2}{\lambda_m^2} \leqq 1, \tag{18}$$

then this point itself is the minimizing point, and the solution is

$$x_p = \xi_p \qquad (p = 1, 2, \cdots, m). \tag{19}$$

On the other hand, if the given point is *outside* of the region, i.e.,

$$\frac{\xi_1^2}{\lambda_1^2} + \frac{\xi_2^2}{\lambda_2^2} + \cdots + \frac{\xi_m^2}{\lambda_m^2} > 1, \tag{20}$$

then a point $(x_1, x_2, \cdots, x_m)$ belonging to the region must lie on its boundary in order to be at a minimum distance from the given point. In other words, since the $x_p/\lambda_m = \hat{b}_{ip}$ are obtained by the orthogonal transformation (11) from the original variables, distance is preserved; therefore the point on the boundary of the region can be expressed in terms of the b_{ip} as in (8). This completes the proof of Theorem 1.

Having reduced the side condition to an equality, conventional mathematical methods are applicable to the problem of minimizing the function under the constraint, when the minimum of f_j is attained outside the region (7). Furthermore, the preceding proof provides additional information that facilitates the solution of the problem. First, of course, it shows that when the minimum of f_j is attained in the region (7) its value is given by $\hat{K}$ in (16) and the minimizing point is (19). More important, for the case of the minimum of f_j being attained outside this region, the foregoing development suggests a much more tractable approach than that originally posed by the problem of minimizing (6) under the constraint (8). The simplified problem, which follows from (16) and (17), is to minimize

$$(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 + \cdots + (x_m - \xi_m)^2 \tag{21}$$

under the constraint

$$\frac{x_1^2}{\lambda_1^2} + \frac{x_2^2}{\lambda_2^2} + \cdots + \frac{x_m^2}{\lambda_m^2} = 1. \tag{22}$$

The method of Lagrange's multipliers is especially suitable to this problem. The essence of this method is the creation of a new function—the function (21) minus μ (the Lagrange multiplier) times the function in (22)—and the setting of the partial derivatives of this new function with respect to the m variables x_p equal to zero. This leads to the equations

$$\begin{aligned} x_1 - \xi_1 - \mu \frac{x_1}{\lambda_1^2} &= 0, \\ x_2 - \xi_2 - \mu \frac{x_2}{\lambda_2^2} &= 0, \\ &\cdots\cdots \\ x_m - \xi_m - \mu \frac{x_m}{\lambda_m^2} &= 0, \end{aligned} \tag{23}$$

which, together with (22), constitute a set of $(m + 1)$ equations in $(m + 1)$ unknowns $x_1, x_2, \cdots, x_m, \mu$.

The parameter μ can be determined from any one of the equations (23), namely:

$$\mu = \frac{\lambda_p^2(x_p - \xi_p)}{x_p} \qquad (p = 1, 2, \cdots, m), \tag{24}$$

and may be eliminated by setting any one of the m determinations equal to any other. Thus, each of the subsequent determinations (24) may be expressed in terms of the first, i.e.,

$$\lambda_p^2\left(1 - \frac{\xi_p}{x_p}\right) = \lambda_1^2\left(1 - \frac{\xi_1}{x_1}\right),$$

and, solving explicitly for the unknowns x_p in terms of x_1, produces

$$x_p = \frac{\lambda_p^2 \xi_p x_1}{(\lambda_p^2 - \lambda_1^2)x_1 + \lambda_1^2 \xi_1} \qquad (p = 2, 3, \cdots, m). \tag{25}$$

Before proceeding to the general solution to the problem of minimizing (21) under the constraint (22), some special situations should be noted. If $\xi_p = 0$ for any p, then $x_p = 0$ must be a solution in order to minimize the distance, and the terms corresponding to this p may be deleted. Furthermore, it may be assumed that $\xi_p > 0$ for every p. If an ξ_p were negative for any p, it could be replaced by $|\xi_p|$ and the resulting solution x_p replaced by $-x_p$. Therefore it may be assumed that every x_p is positive.

Substitution of the values (25) into (22) gives rise to a polynomial equation in x_1 of degree $2m$. The direct solution of such an equation can become quite cumbersome, so a numerical method of successive approximations is employed. The basis for it rests on the following:

THEOREM 2. *For a given* x_1 *between* 0 *and* min (ξ_1, λ_1), *with* x_p $(p = 2, 3, \cdots, m - 1)$ *determined by* (25) *and* x_m *by* (22), *if*

$$\lambda_m^2\left(1 - \frac{\xi_m}{x_m}\right) \gtrless \lambda_1^2\left(1 - \frac{\xi_1}{x_1}\right) \tag{26}$$

then

$$x_1 \lesseqgtr x_1^* , \tag{27}$$

where x_1^* *designates the solution for* x_1 .

The proof begins with the fact that x_p is an increasing function of x_1 (the ξ's being assumed positive). Then the two conclusions are reached by the following reasoning: If $x_1 < x_1^*$, then x_p is less than its solution x_p^*, and consequently x_m is larger than its solution x_m^*. Therefore,

$$\lambda_1^2\left(1 - \frac{\xi_1}{x_1}\right) < \lambda_1^2\left(1 - \frac{\xi_1}{x_1^*}\right) = \lambda_m^2\left(1 - \frac{\xi_m}{x_m^*}\right) < \lambda_m^2\left(1 - \frac{\xi_m}{x_m}\right).$$

If $x_1 > x_1^*$, then a similar argument leads to

$$\lambda_1^2\left(1 - \frac{\xi_1}{x_1}\right) > \lambda_m^2\left(1 - \frac{\xi_m}{x_m}\right),$$

completing the proof.

In the iterative scheme of Theorem 2, the following initial value for x_1 seems convenient:

$$x_1^0 = \xi_1 \Big/ \sqrt{\sum_{p=1}^{m} \left(\frac{\xi_p}{\lambda_p}\right)^2}. \tag{28}$$

The remaining x_p^0 are determined by (25) and (22). If, in the determination of x_m^0 by (22), the quantity for which the square root must be taken should be negative, then half this initial value may be tried.

3. *Numerical Illustration*

The mathematical programming of the preceding section has been incorporated into the original computer program for the minres solution. Specifically, the arbitrary adjustment in the factor loadings of a variable with communality greater than one, without disturbing any of the other factor loadings [2, p. 367], is replaced by the procedure developed in this paper. As a consequence, it is expected that the revised solution should provide a better fit to empirical data that exhibit a Heywood case.

The classical example (see [1], p. 125) used to illustrate the Heywood case consists of the hypothetical correlations in the upper triangle of Table 1. If the values 1.10, .81, .64, .49, .36 are placed in the diagonal of this correlation matrix, its rank will be found to be precisely one. Then, a single common factor with loadings (1.05, .90, .80, .70, .60) will reproduce the correlations with zero residuals. The only trouble is that the factor solution is not acceptable!

A proper minres solution with one factor was obtained by use of a desk calculator (for this simple instance) in about an hour, and on a Philco 2000 in seven seconds. The factor loadings, communalities, and residuals are shown in Table 1. The objective function for this solution is $f = .0044$. Of course, for the solution with the improper communality the objective function is precisely zero. On the other hand, for a solution with one principal component (the initial trial value in the minres program), the objective function is $f = .0728$. The principal-component solution, while extracting maximum variance, does not provide as good a fit to the off-diagonal correlations as the minres solution.

In the course of experimentation with the original minres method [2], eleven different problems were employed; in four instances Heywood cases appeared. These four problems provided the empirical data for testing the

TABLE 1

Minres Solution for Five Hypothetical Variables

Variable j	Correlations and Residuals*					Minres Solution	
	1	2	3	4	5	a_{j1}	h_j^2
1		.945	.840	.735	.630	1.000	1.000
2	.033		.720	.630	.540	.912	.832
3	.031	-.018		.560	.480	.809	.654
4	.028	-.015	-.012		.420	.707	.500
5	.025	-.012	-.009	-.008		.605	.366

* Correlations in upper triangle, residuals in lower triangle.

TABLE 2

Comparison of Original and Revised Minres

Example	Number of Factors	Objective Function			Time	
		Without Constraints	Arbitrary Adjustments	With Constraints	Original	Revised
5 Socio-economic	2	.00094	.00109	.00098	3.3 sec.	3.2 sec.
variables	3	.00000	.00012	.00000	4.6 sec.	16.2 sec.
24 Psychol. tests	5	.37771	.41376	.37811	4.7 min.	3.0 min.
36 MMPI items	12	.26971	.27851	.26595	8.4 min.	8.0 min.

efficacy of the revised procedures in the present paper. The results are shown in Table 2.

Comparison between the original and revised minres methods is made in the actual magnitude of the function (2) and the time required to attain a solution, for the same convergence standard. As a point of reference, the minimum value of the objective function, i.e., for a minres solution without the constraints (4), is also shown for each problem. In each instance, the objective function determined by the mathematical procedures of this paper is better than that obtained by the arbitrary adjustment for the factor loadings for the particular variable exhibiting the Heywood case without disturbing any of the other factor loadings. Furthermore, the fit is almost as good as the solution without constraints (in the last problem, the apparent anomaly of a "better fit" no doubt is due to rounding errors). Fortunately, the time required to obtain a proper minres solution appears to be no greater than the original method with the arbitrary adjustments for communalities greater than one. For the 24-variable problem the time was substantially less than originally. This was due to the fact that the in revised procedure convergence occurred in 600 iterations, while in the original method convergence did not occur within the maximum allowance of 1000 iterations.

4. *Conclusion*

While the original minres method [2] provides an excellent means of getting a "best" fit to the off-diagonal elements of a correlation matrix, unfortunately it sometimes leads to solutions with communalities greater than one. The aribtrary reduction in the factor loadings of the particular variable to force the communality back to one is not very satisfactory. The objective of the present paper was to resolve this problem by introducing constraints (viz., that communalities be numbers between zero and one) in the mathematical process of minimizing the sum of squares of off-diagonal residuals. Not only was this objective met, but the resulting solution was superior to that obtained without the mathematical constraints and required no more computing time.

REFERENCES

[1] Harman, H. H. *Modern factor analysis*. Chicago, Ill.: Univ. Chicago Press, 1960.
[2] Harman, H. H., and Jones, W. H. Factor analysis by minimizing residuals (Minres). *Psychometrika*, 1966, **31**, 351–68.
[3] Heywood, H. B. On finite sequences of real numbers. *Proc. roy. Soc. London*, 1931, **134**, 486–501.

Manuscript received 1/31/66