DOCUMENT RESUME

ED 079 412 TM 003 016

AUTHOR Jennerich, Robert I.
TITLE On the Stability of Rotated Factor Loadings: The Wexler Phenomenon.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RB-73-43
PUB DATE Jun 73
NOTE 27p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Factor Analysis; *Factor Structure; Orthogonal Rotation; *Standard Error of Measurement; Statistical Analysis; Technical Reports

ABSTRACT

The formulas which give the standard errors of factor loading estimates while available and computable are complicated, and our understanding of them is limited. A nontechnical description of their behavior under favorable and unfavorable conditions is given. Of particular interest is their behavior in the presence of singularities arising from equal eigenvalues and undefined rotation. (Author)

RB-73-43



RESEARCH BULLETIN



# ON THE STABILITY OF ROTATED FACTOR LOADINGS: THE WEXLER PHENOMENON

Robert I. Jennrich

University of California at Los Angeles
and
Educational Testing Service

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
June 1973

# ON THE STABILITY OF ROTATED FACTOR LOADINGS: THE WEXLER PHENOMENON


## ABSTRACT

The formulas which give the standard errors of factor loading estimates while available and computable are complicated and our understanding of them is limited. A nontechnical description of their behavior under favorable and unfavorable conditions is given. Of particular interest is their behavior in the presence of singularities arising from equal eigenvalues and undefined rotation.

ON THE STABILITY OF ROTATED FACTOR LOADINGS: THE WEXLER PHENOMENON

## 1. INTRODUCTION

Numerous authors have looked at the sampling stability of loading estimates which arise in factor analysis. While analytic results for the standard errors of unrotated loadings have been available for some time (Lawley, 1953, 1967; Lawley & Maxwell, 1971) those for analytically rotated loadings are fairly recent (Archer & Jennrich, 1973; Jennrich, 1973a,b). Consequently much of the early work on stability was based on simulation studies. After reviewing several studies, Cliff & Hamberger (1966) found that the standard errors of factor loading estimates were about the same as those for correlations, that is about $1/\sqrt{n}$ in magnitude for a sample of size $n$. This is a crude but useful summary of some rather complicated results. It is useful because it is simple and reasonably accurate when everything is going right and crude because it can be fairly wide of the mark when this is not the case. Its usefulness may be considerably enhanced by understanding the mechanisms which cause it to be inaccurate. To identify some of these we begin with a result from an interesting unpublished dissertation by Wexler (1968).

Wexler investigated the finite sample variances of maximum likelihood factor loading estimates comparing them with those obtained using the asymptotic formulas of Lawley (1953). These were Lawley's early results derived under the assumption of known unique variances $\psi$, a restriction which was later removed (Lawley, 1967). Let $A$ be the unrotated (i.e., canonical, Rao, 1955) factor loading matrix for a population satisfying the usual assumptions in maximum likelihood factor analysis (Lawley &

Maxwell, 1971). Let $\Lambda$ be the varimax rotation of $A$ and let $T$ be the transformation which takes $A$ into $\Lambda$ so that

$$\Lambda = AT \quad . \tag{1}$$

Wexler looked at two types of estimates for the rotated loadings $\Lambda$. The first was the estimate

$$\Lambda^* = \hat{A}T \tag{2}$$

computed from the maximum likelihood estimate $\hat{A}$ of $A$ using the population value $T$. The second was the maximum likelihood estimate $\hat{\Lambda}$ of $\Lambda$ computed without assuming $T$ is known so that

$$\hat{\Lambda} = \hat{A}\hat{T} \tag{3}$$

where $\hat{T}$ is the matrix which takes $\hat{A}$ into its varimax rotation $\hat{\Lambda}$. Lawley's formulas give the asymptotic standard errors for the components of $\hat{A}$ and by a simple transformation the asymptotic standard errors for the components of $\Lambda^*$. Because $\hat{T}$ is a function of the data, the asymptotic standard errors for the components of $\hat{\Lambda}$ are a little more complicated (Archer & Jennrich, 1973) and were not available at the time of Wexler's study.

In general Wexler's simulation studies showed reasonably good agreement with Lawley's formulas. Figure 1 is reproduced from his thesis. It shows loading variances computed using Lawley's formulas plotted against actual simulated variances for the components $\lambda^*_{ir}$ of $\Lambda^*$. There is

one point for each loading in $\Lambda$ . The agreement is not perfect, but taking into account the fact that only 100 simulations were used, there are no statistically surprising departures from the asymptotic results nor do there appear to be any systematic ones.

---

Insert Figure 1 about here

---

Because the asymptotic variances for the $\hat{\Lambda}$ loadings were not available at the time of Wexler's thesis, it had been suggested that the asymptotic variances of the $\Lambda^*$ loadings be used as an approximation for the variances of the $\hat{\Lambda}$ loadings. To test this suggestion Wexler plotted the asymptotic variances of the $\Lambda^*$ loadings against actual simulated variances for the $\hat{\Lambda}$ loadings. His plot is given in Figure 2. Many who have seen this figure find it somewhat surprising. First the suggested approximation does not seem to be satisfactory. But of greater interest to us is the fact that for the most part the loadings computed using $\hat{T}$ are considerably more stable than those using the true population value T . We would like to understand why this is so.

---

Insert Figure 2 about here

---

The results in Figure 2 were obtained using a population with a very good varimax loading matrix, i.e., one with nice simple structure. Wexler repeated his entire analysis using a population with only fair

simple structure. The results corresponding to Figure 2 are shown in Figure 3. The phenomenon displayed in Figure 2 is far less pronounced here. It is still abundantly clear, however, that it would be unwise to use the asymptotic variances of the components of $\Lambda^*$ to approximate those of $\hat{\Lambda}$ .

---

Insert Figure 3 about here

---

These examples suggest what we shall call the Wexler phenomenon:

When good simple structure exists rotated loadings may be surprisingly stable.

One manifestation of this phenomenon is that rotated loadings may be considerably more stable than unrotated loadings. The opposite can also happen and this suggests the anti-Wexler phenomenon:

When good simple structure does not exist rotated loadings may be surprisingly unstable.

We intend to investigate these somewhat vague statements in greater detail.

## 2. FORMS OF DEGENERACY

We believe that the Wexler phenomena are associated with forms of degeneracy in the specification of a factor analysis model. Of particular interest are those which arise from:

(i) equal eigenvalues

(ii) undefined rotation.

For the purpose of thinking about these we specialize to a simple case. We shall in fact leave factor analysis entirely and consider the case of principal components analysis with quartimax rotation. This simplification allows us to understand the details clearly without sacrificing the essential issues at hand.

It is easy to give an example which displays both forms (i) and (ii) of degeneracy. Let the 3 by 2 matrix A given in Figure 4 represent the first two principal components of a 3 by 3 population covariance matrix. In practice it is common to plot the rows of A . As displayed in Figure 4 they constitute three equally spaced points on the unit circle. The first two eigenvalues here, being the column sums of squares for A , are equal. Thus A has the first form of degeneracy. On the other hand it is easy to show that the quartimax criterion (Harman, 1967, p. 298) is constant here over all orthogonal rotations of A . As a consequence, the quartimax rotation of A is undefined and A also displays the second form of degeneracy.

---

Insert Figure 4 about here

---

To understand their effect we shall look at examples showing these forms of degeneracy separately. Let the 4 by 2 matrix A given in Figure 5 represent the first two principal components of a 4 by 4 population covariance matrix. The first two eigenvalues here are clearly

equal. On the other hand, as can be seen from Figure 5, A has an independent cluster structure so its quartimax rotation is well defined. The matrix A is in fact its own quartimax rotation. We want to consider the statistical stability of an estimator $\hat{A}$ of A obtained by factoring a sample covariance matrix, and that of $\hat{\Lambda}$ the quartimax rotation of $\hat{A}$. When eigenvalues are equal as they are here we know from the results of Anderson (1963) that as the sample size $n \to \infty$,

$$\hat{A} \to AX \tag{4}$$

in distribution where X is a random orthogonal matrix. This means that for large n with high probability a plot of the rows of $\hat{A}$ will look like a random rotation of a slight perturbation of the points displayed in Figure 5. Because of the random rotation

$$\hat{A} \not\to A \tag{5}$$

as $n \to \infty$ so that $\hat{A}$ is not a consistent estimator of A. On the other hand quartimax rotation of $\hat{A}$ will undo the random rotation so that when n is large $\hat{\Lambda}$ will with high probability look like a slight perturbation of A. In more precise terms (4) and the fact that A is its own quartimax rotation imply that

$$\hat{\Lambda} \to A \tag{6}$$

in probability as $n \to \infty$. Thus for large n the rotated loadings $\hat{\Lambda}$, which converge, will have a greater stability than the unrotated loadings $\hat{A}$, which do not, giving rise to the Wexler phenomenon.

---

Insert Figure 5 about here

---

Consider next the example given in Figure 6. As before, let A represent the first two principal components of a 3 by 3 population covariance matrix. The eigenvalue ratio here is a comfortable 2.71. On the other hand the constant c in A has been carefully chosen so that the quartimax criterion is constant over every orthogonal rotation of A making the quartimax rotation of A undefined.

---

Insert Figure 6 about nere

---

As before let $\hat{A}$ be an estimate of A obtained by factoring a sample covariance matrix and let $\hat{\Lambda}$ be the quartimax rotation of $\hat{A}$. Because of the eigenvalue ratio (and assuming the third eigenvalue is not equal or nearly equal to the second) we expect $\hat{A}$ to be near A when n is large. Since quartimax rotation is undefined at A, however, we expect small changes of $\hat{A}$ in a neighborhood of A to produce large changes in the rotated loadings $\hat{\Lambda}$. That is we expect the stability of $\hat{\Lambda}$ to be poor compared to that of $\hat{A}$ giving rise to the anti-Wexler phenomenon. As we shall see in the next section, this in fact happens.

The first form of degeneracy discussed here was considered by Jöreskog (1963, p. 86) using a form of estimation proposed by him together with target rotation. He observed that when the appropriate eigenvalues are nearly equal his loading estimates could be expected to have large variances before rotation but moderate variances after. This, in a slightly different context, is a manifestation of the Wexler phenomenon. Jöreskog did not consider cases corresponding to Figures 4 and 6 possibly because they were not particularly interesting in the context of target rotation.

## 3. TWO CONFIRMATORY EXAMPLES

By considering the specific case of principal components analysis and orthomax rotation we have set forth rationales for the Wexler and anti-Wexler phenomena. In this section by looking at two specific examples and computing exact asymptotic variances we will verify that the two forms of degeneracy considered do in fact produce the Wexler and anti-Wexler phenomena. To demonstrate the generality of the arguments given, we choose examples which are technically quite different but clearly analogous to those of the last section. They are based on standardized (i.e., computed from a sample correlation matrix) maximum likelihood loading estimation and varimax rotation. Clearly maximum likelihood factor analysis is not the same as the principal components analysis nor is varimax rotation the same as quartimax.

Let A be the matrix on the left in Table 1 and let it represent the standardized loadings in a canonical (Rao, 1955) factor analysis model

for a normal population. The model involves 2 factors and 6 score variables. This is the smallest number of variables a two factor model with perfect structure can have and still have identifiable unique variances (see the identifiability conditions summarized by Anderson & Rubin, 1956).

The appropriate eigenvalues here are the diagonal elements of $A'\Psi^{-1}A$ where $\Psi = \mathrm{diag}(I - AA')$ is the matrix of standardized unique variances. (Estimates of these eigenvalues are provided by standard maximum likelihood factor analysis programs.) Because the eigenvalue ratio here is nearly one (1.07 to two decimal places) analogy with the example of Figure 5 suggests that the standard errors for the maximum likelihood estimates of the components of $A$ will be quite large. On the other hand the perfect structure of $A$ suggests that varimax rotation may produce estimates with relatively small standard errors. Using $A$ and the formulas of Jennrich (1973b) the asymptotic standard errors of both the unrotated and rotated estimates were computed and are recorded in Table 1. Clearly at least some of the unrotated loading estimates are highly unstable while all of the rotated loading estimates are quite stable. In the worst cases the standard errors of the unrotated loadings are about 20 times as large as the corresponding rotated loadings. It is easy to believe from this example that the Wexler phenomenon may be made arbitrarily pronounced by choosing population values sufficiently close to the appropriate form of degeneracy.

------------------------

Insert Table 1 about here

------------------------

Turning to the anti-Wexler phenomenon let A be the 6 by 2 matrix on the left in Table 2 and as before let it represent the standardized loadings in a canonical factor analysis model for a normal population. The eigenvalue ratio here is 4.25 so that arguing by analogy with the last example in the previous section we expect the maximum likelihood estimates of the unrotated loadings to have moderate standard errors. On the other hand the value .81 in the upper left-hand corner of A was carefully chosen so that the varimax rotation of A is nearly undefined (the precise value which renders it undefined is $(.43)\sqrt[4]{12} \doteq .8003$ ). Thus it is reasonable to expect large standard errors for the maximum likelihood estimates of the rotated loadings.

Using the formulas of Jennrich (1973b) again, the actual asymptotic standard errors are given in Table 2. As expected the standard errors for the unrotated loadings have moderate values while at least some of those for the rotated loadings are quite large. In the worst cases the latter are about 26 times as large as the former and it is easy to believe that the anti-Wexler phenomenon may be made arbitrarily pronounced by choosing an appropriate sufficiently singular example.

-------------------------

Insert Table 2 about here

-------------------------

## 4. DISCUSSION

Standard error formulas for analytically rotated factor loadings have only recently become available. While we can now compute standard errors at will the computations which lead to them are complicated and our understanding of them is quite limited. A crude but simple summary asserts that for a factor loading estimate (of any kind) the

$$\text{standard error} = 1/\sqrt{n} \quad . \tag{7}$$

Both unrotated and rotated loading estimates, however, can be made to have arbitrarily large standard errors by choosing examples with the appropriate form of singularity. Interestingly, rotated loadings need not have large standard errors simply because the unrotated loadings from which they are computed do. And conversely very stable unrotated loadings can lead to very unstable rotated loadings. We have called these observations the Wexler and anti-Wexler phenomena and we know in some detail why the summary (7) must be crude.

An approximation proposed by C. W. Harris and reported by Cattell (1966, p. 235) asserts that for a factor loading estimate $\hat{\lambda}_{ir}$ the

$$\text{standard error} = \left(\frac{(1 - h_i^2)\phi^{rr}}{n - k - 1}\right)^{\frac{1}{2}} \tag{8}$$

where $h_i^2$ is the communality of the $i$ -th variable, $\phi^{rr}$ is the $r$ -th diagonal element in the inverse of the matrix of factor correlations, and $k$ is the number of factors. Since there is nothing in this formula which

allows for the effect of nearly equal eigenvalues or poorly defined rotation we know from consideration of the Wexler phenomena that this too must, in some cases at least, be a very crude approximation.

On the other hand a summary as simple as (7) can be quite useful when it is necessary to make simple inferences without the aid of a computer. Lawley & Maxwell (1971, p. 43) present a maximum likelihood factor analysis involving 211 observations, 9 variables, and 3 factors. The eigenvalues involved are quite distinct and direct quartimin rotation (Jennrich & Sampson, 1966) gives loadings with good simple structure so that one might be tempted to use

$$1/\sqrt{211} = .069 \tag{9}$$

as a standard error for the rotated loadings. The actual asymptotic standard errors computed by Jennrich (1973a) are reproduced in Table 3. Considering the simplicity of the formula which led to the value .069, its agreement with the computed values is rather pleasing and good enough for rough inferential purposes (cf Jennrich, 1973a). Because of the Wexler phenomenon, however, one must use a good deal of caution with such an approximation when eigenvalues are not clearly distinct or rotations are not well defined.

------------------------

Insert Table 3 about here

------------------------

The author is indebted to Norman Wexler for making his unpublished results available and for reviewing an earlier draft of this manuscript. He is also grateful to Harry Harman for reviewing the manuscript and for pointing out years ago the need for results on the statistical stability of factor loading estimates.

REFERENCES

Anderson, T. W. (1963). Asymptotic theory for principal-component analysis. Ann. math. Statist. 34, 122-148.

Anderson, T. W. & Rubin, H. (1956). Statistical inference in factor analysis. Proc. Third Berkeley Symp. math. Statist. Prob. 5, 111-150.

Archer, C. O. & Jennrich, R. I. (1973). Standard errors for rotated factor loadings. Research Bulletin 73-17. Princeton, N.J.: Educational Testing Service.

Cattell, R. B. (1966). Handbook of Multivariate Experimental Psychology. Chicago: Rand McNally.

Cliff, N. & Hamberger, C. N. (1966). The study of sampling errors in factor analysis by means of artificial experiments. Mimeographed report. Los Angeles, Calif.: University of Southern California.

Harman, H. H. (1967). Modern Factor Analysis. Chicago: University of Chicago Press.

Jennrich, R. I. (1973a). Standard errors for obliquely rotated factor loadings. Research Bulletin 73-28. Princeton, N.J.: Educational Testing Service.

Jennrich, R. I. (1973b). Simplified formulas for standard errors in maximum likelihood factor analysis. Research Bulletin 73-40. Princeton, N.J.: Educational Testing Service.

Jennrich, R. I. & Sampson, P. F. (1966). Rotation for simple loadings. Psychometrika 31, 213-223.

Jöreskog, K. G. (1963). Statistical Estimation in Factor Analysis: A New Technique and Its Foundation. Stockholm: Almqvist & Wiksell.

Lawley, D. N. (1953). A modified method of estimation in factor analysis and some large sample results. In Uppsala Symposium on Psychological Factor Analysis, Nordisk Psychologe's Monograph Series No. 3. Stockholm: Almqvist & Wiksell. Pp. 35-42.

Lawley, D. N. (1967). Some new results in maximum likelihood factor analysis. Proc. Roy. Soc. Edinburgh A67, 256-264.

Lawley, D. N. & Maxwell, A. E. (1971). Factor Analysis as a Statistical Method. New York: American Elsevier.

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

Wexler, N. (1968). An evaluation of an asymptotic formula for factor loading variance by random methods. Unpublished doctoral dissertation, Rutgers University, New Brunswick, N.J.

Table 1.--Asymptotic Results in the Case of Nearly Equal Eigenvalues and Well Defined Rotation: Wexler Phenomenon

| Population Loadings* | | Asymptotic Standard Errors+ | | | |
|---|---|---|---|---|---|
| | | Unrotated | | Rotated | |
| .81 | .00 | .05 | 1.31 | .05 | .07 |
| .81 | .00 | .05 | 1.31 | .05 | .07 |
| .81 | .00 | .05 | 1.31 | .05 | .07 |
| .00 | .80 | 1.39 | .05 | .07 | .05 |
| .00 | .80 | 1.39 | .05 | .07 | .05 |
| .00 | .80 | 1.39 | .05 | .07 | .05 |

*The rotated and unrotated loadings are identical here.

+Standard errors are scaled to correspond to a sample size of 100.

Table 2.--Asymptotic Results in the Case of Nearly Undefined Loadings and a Good Eigenvalue Ratio: Anti-Wexler Phenomenon

| Population Loadings* | | Asymptotic Standard Errors+ Unrotated | | Rotated | |
|---|---|---|---|---|---|
| .81 | .00 | .06 | .09 | .06 | 2.33 |
| .81 | .00 | .06 | .09 | .06 | 2.33 |
| -.43 | .43 | .11 | .12 | 1.27 | 1.31 |
| -.43 | .43 | .11 | .12 | 1.27 | 1.31 |
| -.43 | -.43 | .11 | .12 | 1.27 | 1.31 |
| -.43 | -.43 | .11 | .12 | 1.27 | 1.31 |

*The rotated and unrotated loadings are identical here.

+Standard errors are scaled to correspond to a sample size of 100.

Table 3.--Asymptotic Standard Errors for a Direct Quartimin Rotation of Maximum Likelihood Loading Estimates: Reproduced from Jennrich (1973a)

| | Factor | | |
|---|---|---|---|
| Variate | I | II | III |
| 1 | .074 | .082 | .096 |
| 2 | .069 | .082 | .073 |
| 3 | .072 | .053 | .067 |
| 4 | .054 | .084 | .075 |
| 5 | .065 | .069 | .041 |
| 6 | .046 | .058 | .056 |
| 7 | .064 | .036 | .141 |
| 8 | .064 | .081 | .116 |
| 9 | .046 | .050 | .059 |

FIGURE CAPTIONS

FIG. 1--Scatter plot of asymptotic variances versus empirical variances of loadings estimating the elements of a good simple structure population factor matrix. The population transformation matrix was used. Multiplicity of plots is indicated by encircled points.

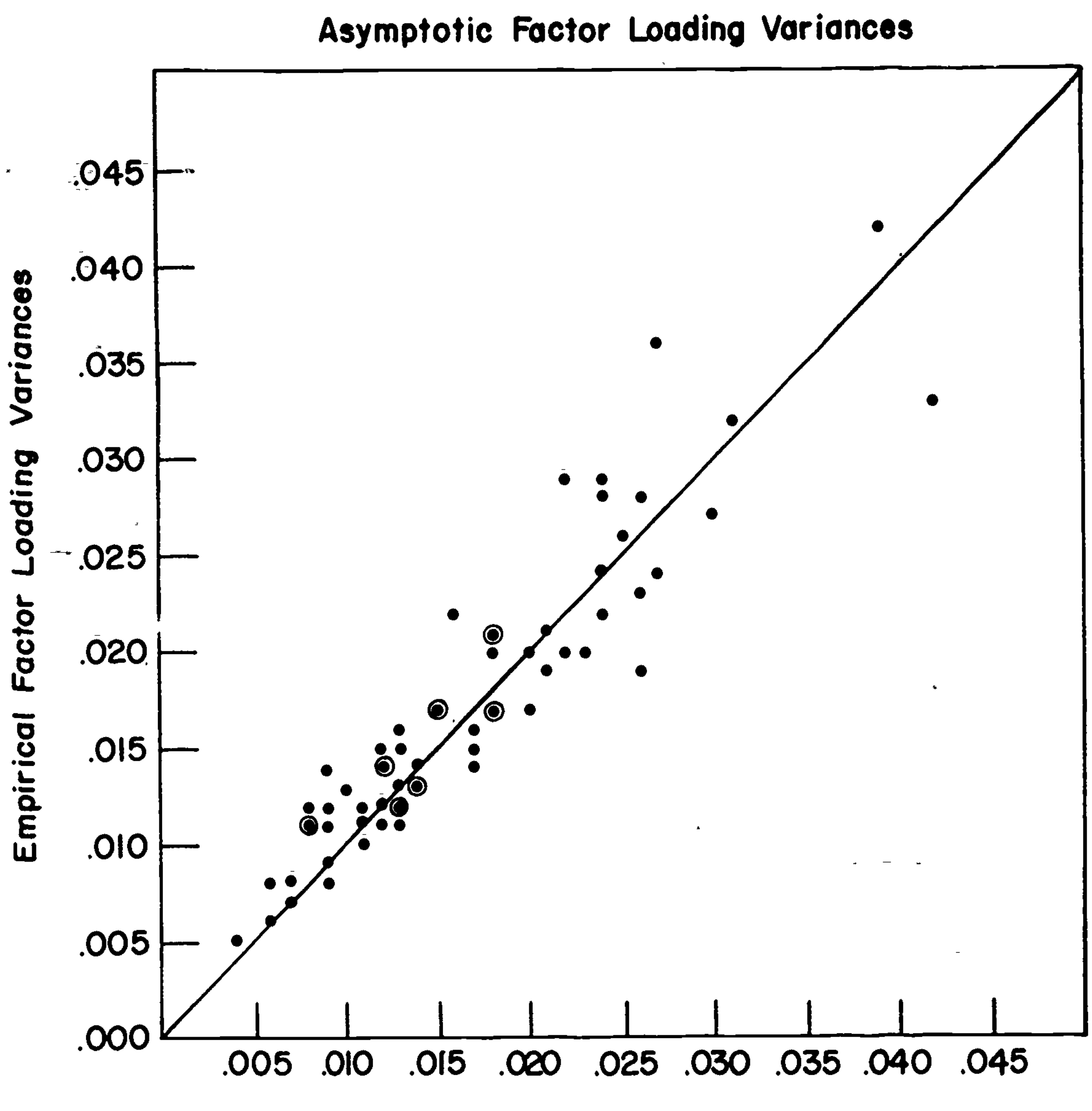
FIG. 2--Scatter plot of asymptotic variances versus empirical variances of loadings estimating the elements of a good simple structure population factor matrix. The sample transformation matrix was used. Multiplicity of plots is indicated by encircled points.

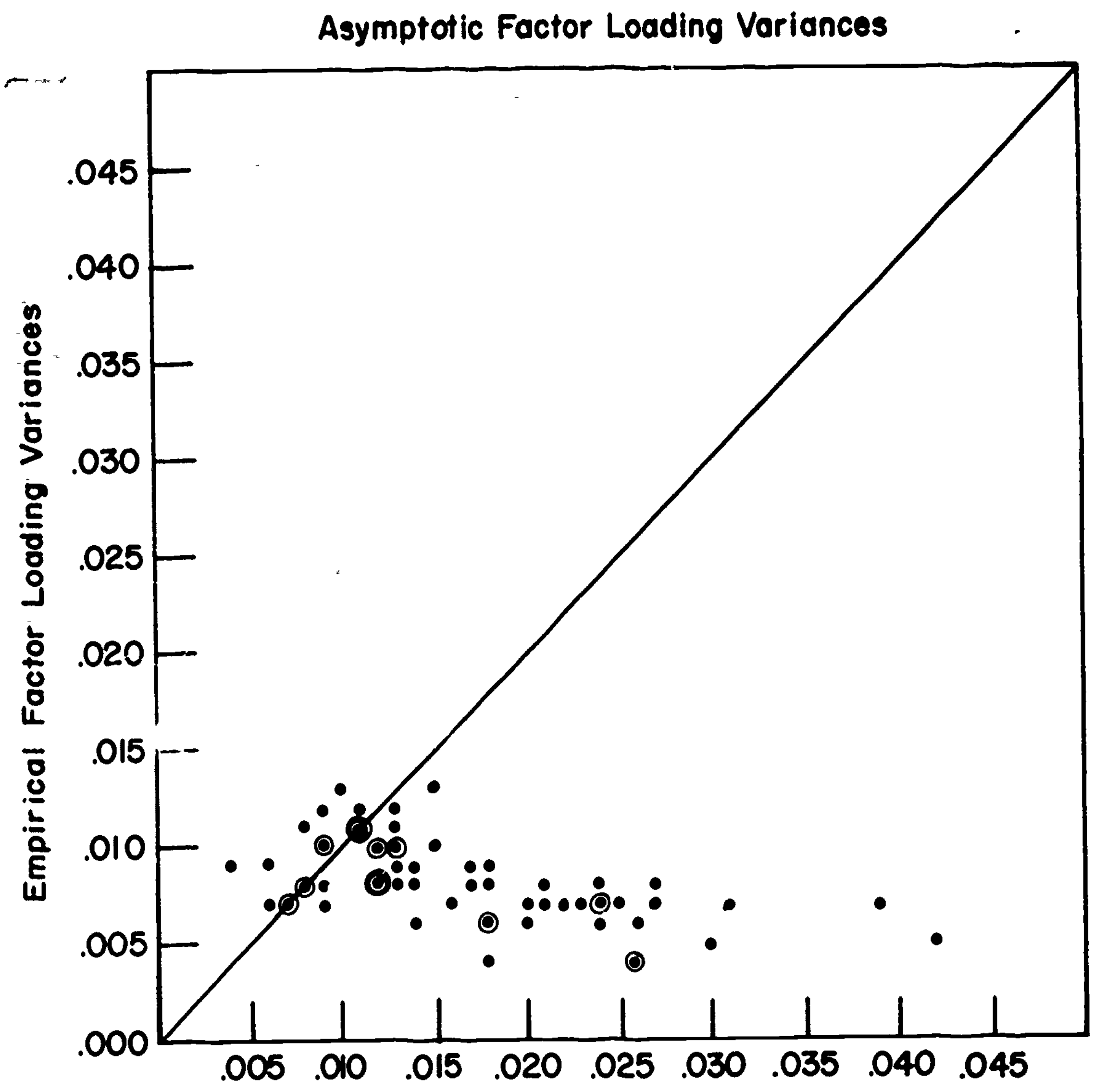
FIG. 3--Scatter plot of asymptotic variances versus empirical variances of loadings estimating the elements of a population factor matrix with only fair simple structure. The sample transformation matrix was used. Multiplicity of plots is indicated by encircled points.

FIG. 4--A principal components example displaying both forms of degeneracy: Equal eigenvalues and undefined rotation.

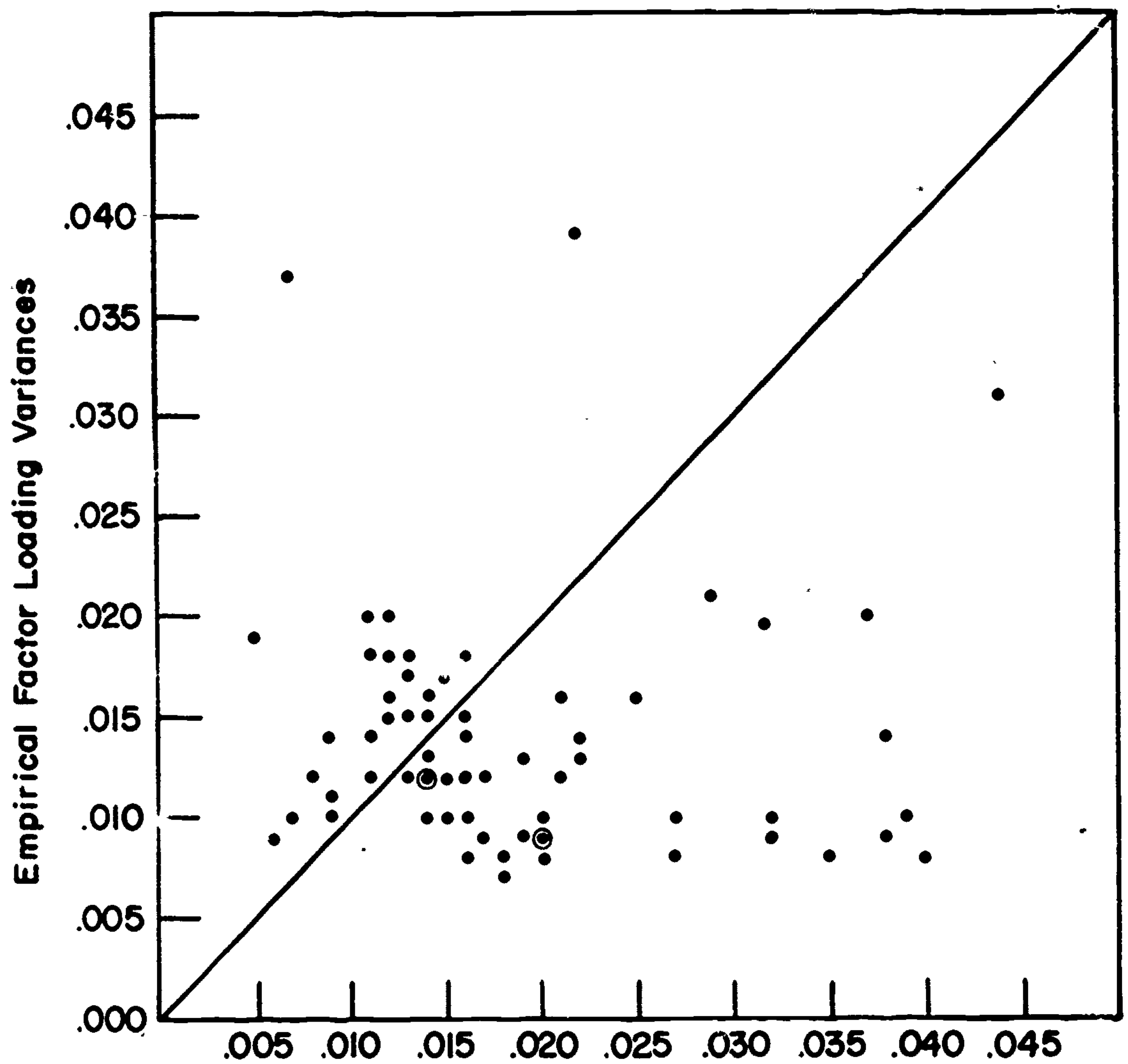
FIG. 5--A principal components example displaying the degeneracy which leads to the Wexler phenomenon: Equal eigenvalues with well-defined rotation.

FIG. 6--A principal components exqmple displaying the degeneracy which leads to the anti-Wexler phenomenon: Undefined rotation with distinct eigenvalues.

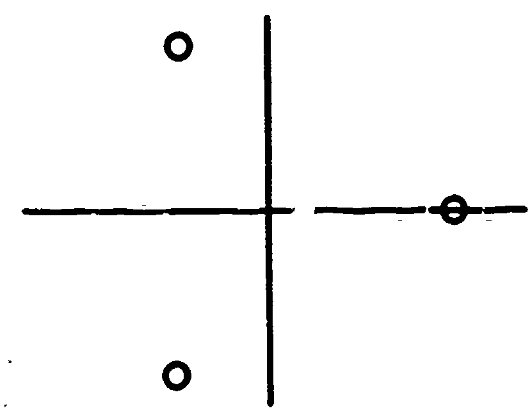

Asymptotic Factor Loading Variances
Empirical Factor Loading Variances
.045
.040
.035
.030
.025
.020
.015
.010
.005
.000
.005 .010 .015 .020 .025 .030 .035 .040 .045


Asymptotic Factor Loading Variances
Empirical Factor Loading Variances
.045
.040
.035
.030
.025
.020
.015
.010
.005
.000
.005 .010 .015 .020 .025 .030 .035 .040 .045

Asymptotic Factor Loading Variances

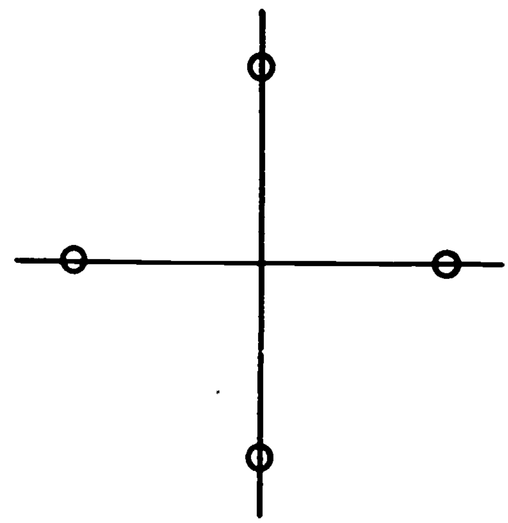


$$A = \begin{pmatrix} 1 & 0 \\ a & b \\ a & -b \end{pmatrix}$$



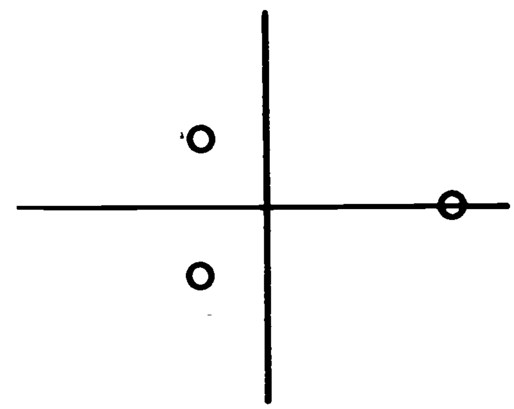
$a = \cos(2\pi/3)$ $\quad b = \sin(2\pi/3)$ $\quad$ eigenvalue ratio = 1

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}$$



eigenvalue ratio = 1

$$A = \begin{pmatrix} 1 & 0 \\ -c & c \\ -c & -c \end{pmatrix}$$



$$c = 2^{-3/4}$$

eigenvalue ratio = 2.71