6

Exploratory Common Factor Analysis

In this chapter, we study the *algebraic* properties of exploratory factor analysis, an extremely popular data analytic technique that dates back to the beginning of the 20th century. Exploratory factor analysis is the historical precursor to confirmatory factor analysis and structural equation modeling. Major books have been written about factor analysis, and focus of this chapter is on the key algebraic properties of the factor analysis *model*, rather than statistical or practical aspects of the method, which will be discussed in subsequent chapters.

6.1 SPEARMAN'S SINGLE-FACTOR THEORY OF GENERAL INTELLIGENCE

In 1904, Charles Spearman, a British psychologist, proposed his "single factor" theory of intelligence. Spearman sought to explain the relationships among various measures of mental ability by means of a single (underlying) ability, which he called general intelligence, or "g." Spearman was an empiricist, and also mathematically talented, and so he became, in a sense, one of the most important early "mathematical psychologists" by proposing a sophisticated (and falsifiable) mathematical model. This model came to be called the "common factor model."

Spearman's g was a "latent" variable, in the sense that there did not exist independent operations and criteria for measuring it. Rather, it was defined only in terms of the equations of the factor analysis model. Spearman postulated that g existed, even though it was only evidenced indirectly by the battery of mental ability tests. The existence of a g could be tested, however, because, if a g exists, and if, using linear regression, it is partialled out of the observed variables, their partial covariances should all become zero. Suppose the observed variables are gathered in a random vector $\mathbf{x}$. Then, recalling Equation 5.22 and realizing that the variance of the latent variable may be arbitrarily set to 1, Spearman deduced that the possible existence of a g could be verified by showing that a vector of regression weights $\mathbf{b}$ exists such that $\mathbf{\Sigma_{XX}} - \mathbf{bb}'$ is diagonal.

Clearly, for a given $p \times p$ covariance matrix $\mathbf{\Sigma_{XX}}$ with $p > 2$, there may not be any $\mathbf{b}$ such that $\mathbf{\Sigma} - \mathbf{bb}'$ is diagonal, and so Spearman's model was falsifiable. Spearman spent a number of years gathering data on mental ability tests in the hope that it would verify his model. He hoped that a number of benefits would ensue from fitting the common factor model (with a single common factor) to a set of mental ability tests. First, by fitting the common factor model and determining $\mathbf{b}$, the factor loadings, he hoped to discover which ability tests loaded on general intelligence. Second, by obtaining the sample equivalent of $\boldsymbol{\xi}$, the vector of observed intelligence factor scores, he hoped to be able to obtain a pure measure of intelligence for each individual. This intelligence score could, ultimately, be registered for each person, and help determine that person's position in the society. There were a number of complications that sidetracked him.

1. There was the nasty issue of sampling variability. Even if the single factor model held in the population, it would almost certainly not hold in a sample of size N from that population. And, unfortunately, the statistical theory to deal with this problem was not available.

2. Other researchers believed there was more than one fundamental factor of mental ability. Soon, the common factor model was extended so that it would allow more than one common factor. Garnett (1919) introduced the notion of a multiple common factor model, which was subsequently popularized by Thurstone. It is easy to show that, for any data set, a factor model with $m+1$ factors will always fit a covariance matrix "better" (in the primitive sense of leaving smaller residuals) than a model with m factors, so long as the fit with m factors is not perfect. Hence, there were many debates, based on prejudice as well as fact, about just how many mental ability factors were needed to explain performance on mental tests. Indeed, Sir Godfrey Thomson (Thomson, 1916) demonstrated a rather striking phenomenon, i.e., that a single factor model could be mimicked by a model with a very large number of factors.

3. E. B. Wilson, a well known Harvard statistician and mathematician, discovered the problem of *factor indetermiancy* (see discussion below), and a lively debate between Wilson, Spearman, and others ensued. The debate lasted a decade, and, by the time it was over, Spearman's golden

moment of intellectual opportunity had passed, in the sense that others (most notably L. L. Thurstone) had seized the momentum and distracted from the extraordinarily fundamental nature of his contributions.

Ultimately, of course, multiple factor analysis and related methods outgrew the domain of mental ability research, and became widely employed in an astonishing diversity of areas. Steiger (1994) identified 4 common, somewhat related rationales underlying the factor model:

1. *The Partial Correlation Rationale.* This rationale, developed and championed by Spearman, viewed common factors as the explanatory concepts underlying a set of observed correlations. If these concepts were to be measured, and the partial correlations among the observed variables were to drop to zero after the "common factors" were partialled out, then the factors "explain the correlations among the observed variables" in the partial regression sense.

2. *The Random Noise Rationale.* In this view of factor analysis, the observed variables represent our best attempt to measure some physical process. Unfortunately, our measurements are noisy — there is random noise polluting the measurement. A classic example might be EEG responses to carefully timed standardized auditory signals, recorded at several sensors. It may be that each sensor will pick up output from several unified, consistent sources within the brain, but that these signals will also include random, uncorrelated electrical noise. In this case, the underlying sources are the "common factors" $\boldsymbol{\xi}$ underlying the measured signals in $\mathbf{x}$.

3. *The True Score Rationale.* In psychometrics, we commonly measure attributes with devices that are assumed to be degraded by random error. In particular, classical true score theory postulates measurements that involve an underlying true score component, and a random error component. If we measure the same ability with several items, this turns out to be a special case of the common factor model. What we are really interested in is the underlying true scores on the variables of interest. The distinction between the observed scores on measures of a trait, and the underlying trait itself, can be especially crucial when we seek to establish linear regression relations among variables that have varying amounts of error variance. Observed correlations can be attenuated by unreliability, and so the regression relations among the unreliable measures of a set of traits can mislead one about the relations among the traits themselves. Because of this problem, it is common to try to estimate regression relationships between the common factors underlying a group of measures, rather than the measures themselves.

4. *The Data Reduction Rationale.* In many situations, it is computationally inconvenient to operate with a large number of measures. We seek to reduce the number of measures, while simultaneously classifying them into groups, and increasing the reliability of what they measure. This data reduction rationale for factor analysis is a major use for factor analytic technology. We factor analyze a group of items to discover the major sources of variation underlying them, and to find out which items are related to which sources. The resulting information allows us to parcel items into groups, to gain a better understanding of the structure underlying our items, and refine our measures of the sources of variation.

Before we continue discussing the conceptual foundations of factor analysis, I shall introduce a notation which will be specialized for our discussion of common factor analysis and principal component analysis.

6.2 THE POPULATION COMMON FACTOR MODEL

Let $\mathbf{x}$ be a $p \times 1$ random vector of observed variables. Let $\boldsymbol{\xi}$ be an $m \times 1$ vector of "common factors." Let $\boldsymbol{\Sigma}$ be the variance-covariance matrix of the observed variables, and let $\boldsymbol{\Psi}$ be the variance-covariance matrix of the common factors. Define $\boldsymbol{\Lambda}$ to be a $p \times m$ matrix of least squares multiple regression weights for predicting the variables in $\mathbf{x}$ from those in $\boldsymbol{\xi}$. Further, assume that all random variables are in deviation score form. Then the multiple common factor model states that

$$\mathbf{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta} \tag{6.1}$$

The residual variables in the $p \times 1$ random vector $\boldsymbol{\delta}$ are referred to by a number of names, depending on the historical and conceptual context. They have often been called "specific factors," "unique factors," or "unique variables." They can be viewed conceptually in a number of ways. For example, they might be viewed as that part of an observed test that is unique to that particular test. Or, they might be viewed as random error or noise, superimposed on a group of signals.

In the multiple common factor model, we stipulate that $\mathcal{E}(\boldsymbol{\xi}\boldsymbol{\delta}') = \mathbf{0}$, and it then immediately follows from expected value theory that

$$\begin{aligned}
\boldsymbol{\Sigma} &= \mathcal{E}(\mathbf{x}\mathbf{x}') \\
&= \mathcal{E}\left((\boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta})(\boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta})'\right) \\
&= \mathcal{E}\left(\boldsymbol{\Lambda}\boldsymbol{\xi}\boldsymbol{\xi}'\boldsymbol{\Lambda}'\right) + \mathcal{E}\left(\boldsymbol{\Lambda}\boldsymbol{\xi}\boldsymbol{\delta}'\right) + \mathcal{E}\left(\boldsymbol{\delta}\boldsymbol{\xi}'\boldsymbol{\Lambda}'\right) + \mathcal{E}\left(\boldsymbol{\delta}\boldsymbol{\delta}'\right) \\
&= \boldsymbol{\Lambda}\mathcal{E}\left(\boldsymbol{\xi}\boldsymbol{\xi}'\right)\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\mathcal{E}\left(\boldsymbol{\xi}\boldsymbol{\delta}'\right) + \mathcal{E}\left(\boldsymbol{\delta}\boldsymbol{\xi}'\right)\boldsymbol{\Lambda}' + \mathcal{E}\left(\boldsymbol{\delta}\boldsymbol{\delta}'\right) \\
&= \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\mathbf{0} + \mathbf{0}'\boldsymbol{\Lambda}' + \mathcal{E}\left(\boldsymbol{\delta}\boldsymbol{\delta}'\right) \\
&= \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \mathcal{E}\left(\boldsymbol{\delta}\boldsymbol{\delta}'\right)
\end{aligned} \tag{6.2}$$

In the common factor model we usually stipulate that $\mathcal{E}(\boldsymbol{\delta}\boldsymbol{\delta}') = \mathbf{U}^2$, where $\mathbf{U}^2$ is a diagonal, positive definite matrix with diagonal entries greater than zero and less than one. The resulting equation,

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \mathbf{U}^2, \tag{6.3}$$

became known as the "Fundamental Theorem of Factor Analysis." In the "orthogonal common factor model," the common factors are "orthogonal," or uncorrelated, and of unit variance, so that $\boldsymbol{\Psi} = \mathbf{I}$. It is also common to restrict the common factors to have unit variances (i.e., variances of 1) in the more general case, since these variances are essentially arbitrary.

A random vector $\boldsymbol{\xi}$ "fits the common factor model" if, when partialled from $\mathbf{x}$ in the multiple regression system, it leaves a residual which has a diagonal, positive definite variance-covariance matrix $\mathbf{U}^2$). It can be shown that a $\boldsymbol{\xi}$ (of order $m \times 1$) which fits the m-factor common factor model exists if and only if there exists an $\boldsymbol{\Lambda}$ such that $\boldsymbol{\Sigma} - \boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \mathbf{U}^2$, a diagonal positive definite matrix. Hence, we may "fit the common factor model" by finding whether there exists a matrix $\boldsymbol{\Lambda}$ such that $\boldsymbol{\Sigma} - \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ is of the desired form, or, alternatively, finding whether there exists a diagonal matrix $\mathbf{U}^2$ with diagonal entries greater than zero and less than one which, when subtracted from $\boldsymbol{\Sigma}$, leaves a matrix which is "Gramian and of rank m" (i.e., may be written as $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ for some matrix $\boldsymbol{\Lambda}$ of rank m). In practice, we do the latter.

If the observed variables are in standard score form, then $\boldsymbol{\Sigma}$ will be a correlation matrix. In this case, recalling the results from the previous chapter, the squared multiple correlation of the observed variables in $\mathbf{x}$ with the common factors in $\boldsymbol{\xi}$ is given by the diagonal elements of the matrix $\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}'$. These diagonal entries are frequently referred to as the "communalities" of the observed variables.

6.3 FACTOR ANALYSIS IN THE SAMPLE

It is, *a priori*, extremely unlikely that the common factor model would fit a population covariance matrix perfectly. And, even if it did, any sample covariance matrix based on N independent observations taken from that population would almost certainly not fit a sample equivalent of the factor model perfectly, due to sampling error. Consider a set of N observed scores on p variables in the data matrix $\mathbf{X}$. In this case, we specify

$$\mathbf{X} = \mathbf{Y}\hat{\boldsymbol{\Lambda}}' + \mathbf{Z} + \mathbf{E} \tag{6.4}$$

with the appropriate side condition that all variables are in deviation score form, that $\mathbf{Y}'\mathbf{Z} = \mathbf{0}$, and that $\mathbf{Z}'\mathbf{Z} = \hat{\mathbf{U}}^2$ is diagonal and positive definite. In this case the scores in $\mathbf{Y}$ are called "common factor scores."

Once the sample estimate factor pattern $\hat{\boldsymbol{\Lambda}}$, factor correlation matrix $\hat{\boldsymbol{\Psi}}$, and unique variance matrix $\hat{\mathbf{U}}^2$ have been obtained, one may attempt to reproduce the sample covariance matrix $\mathbf{S}$ via the sample equivalent of the "fundamental theorem of factor analysis." (Equation 6.3) Specificially, one obtains

$$\hat{\Sigma} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Psi}}\hat{\mathbf{\Lambda}}' + \hat{\mathbf{U}}^2 \tag{6.5}$$

Of course, as mentioned above, it is *extremely* unlikely that, in practice, one will have $\hat{\mathbf{\Sigma}} = \mathbf{S}$, and so the fit of the common factor model is assessed, in practice, by examining the "residual covariance matrix" $\mathbf{S} - \hat{\mathbf{\Sigma}}$. Large elements in this matrix indicate that something has gone wrong in the fitting of the factor model.

We will have much more to say about the residual covariance matrix in subsequent sections dealing with the practical and statistical aspects of factor analysis. Keep in mind that, since the exploratory factor model is often fit to standardized data, the distinction between the sample covariance matrix $\mathbf{S}$ and the sample correlation matrix $\mathbf{R}$ is eliminated, and so often the residual covariance matrix is also a residual correlation matrix.

6.4 NON-UNIQUENESS PROBLEMS IN THE MULTIPLE COMMON FACTOR MODEL

The unequivocal support that Spearman sought for his "theory of g" fueled his enthusiasm for the common factor model. The model of Equation 6.1, along with the appropriate side conditions, is sometimes referred to as the "factor model at the random variable level." If this model fits the data, then a simple consequence Equation 6.3. This "fundamental theorem of factor analysis," allows one to test whether the m-factor model is tenable by examining whether a diagonal positive definite $\mathbf{U}^2$ can be found so that $\mathbf{\Sigma} - \mathbf{U}^2$ is Gramian and of rank m. The early factor analysts, especially Spearman, found this notion almost magical. You can test whether a (hopefully small) set of m variables explaining the variation in $\mathbf{x}$could exist, without ever observing such variables directly. Moreover, you could examine the linear regression relations between $\mathbf{x}$ and the unobserved, hypothetical $\boldsymbol{\xi}$ by matrix factorization of $\mathbf{\Sigma} - \mathbf{U}^2$. The idea is indeed fascinating, and it is easy to understand why Spearman and Thurstone found variants of it so compelling.

There were two elements of the factor model that, if identified, could provide substantial practical benefits. The "factor pattern," $\mathbf{\Lambda}$, by revealing the regression relationships between the observed variables and the more fundamental factors that generate them, could provide information about the structure of the variables being investigated. The sample equivalent of $\boldsymbol{\xi}$ would provide scores on the factors. So, for example, if the factor model fit a set of mental ability tests, one could determine a small set of underlying mental abilities that explain a larger number of tests, and the ratings of the test takers on these fundamental abilities. Indeed, Hart and Spearman (1912) envisioned a virtual factor analytic utopia:

> Indeed, so many possibilities suggest themselves that it is difficult to speak freely without seeming too extravagant ... It seems even possible to anticipate the day when there will be yearly official registration of the "intellectual index," as we will call it, of every child throughout the kingdom ... The present difficulties of picking out the abler children for more advanced education, and the "mentally defective" children for less advanced, would vanish in the solution of the more general problem of adapting education to all ... Citizens, instead of choosing their career at almost blind hazard, will undertake just the professions really suited to their capacities. One can even conceive the establishment of a minimum index to qualify for parliamentary vote, and above all for the right to have offspring. [Hart & Spearman, 1912, pp. 78–79]

Unfortunately, it turned out that there was a hierarchy of indeterminacy problems associated with the factor analysis algebra presented above. Rather than discuss the problems in the clear, systematic way that simple accuracy would seem to demand, authors committed to the common factor model have generally omitted at least one, or described them in obscure, misleading cliches. I describe them here, and urge the reader to compare my description with treatments of the factor model found in many other texts and references.

1. *Identification of* $\mathbf{U}^2$. There may be more than one $\mathbf{U}^2$ that, when subtracted from $\mathbf{\Sigma}$, leaves it Gramian and of rank m. This fact, well known to econometricians, and described with considerable clarity and care by Anderson and Rubin (1956), is not described clearly in several factor analysis texts. One reason for the confusion may be that necessary and sufficient conditions for identification of $\mathbf{U}^2$ have never been established, and there are a number of incorrect statements and theorems in the literature. There are some known conditions when $\mathbf{U}^2$ is not identified (described by Anderson and Rubin). For example, $\mathbf{U}^2$ is never identified if either $p = 2$ and $m = 1$, or if $p = 4$ and $m = 2$. On the other hand, if the number of variables is sufficiently large relative to the number of factors so that $(p - m)^2 > (p + m)$, then $\mathbf{U}^2$ will almost certainly be identified. *However*, if any column of $\mathbf{\Lambda}$ can be rotated into a position where it has only 2 non-zero elements (see discussion of rotation below), then $\mathbf{U}^2$ will not be identified. This means that the identification of $\mathbf{U}^2$ can never be determined simply by counting the number of observed variables and the number of factors.

2. *Rotational Indeterminacy of* $\mathbf{\Lambda}$. Even if $\mathbf{U}^2$ is identified, $\mathbf{\Lambda}$ will not be if $m > 1$. Suppose, for example, we require m orthogonal factors. If such a model fits, then infinitely many $\mathbf{\Lambda}$ matrices will satisfy $\mathbf{\Sigma} - \mathbf{U}^2 = \mathbf{\Lambda}\mathbf{\Lambda}'$, since $\mathbf{\Lambda}\mathbf{\Lambda}' = \mathbf{\Lambda_1}\mathbf{\Lambda_1}'$ so long as $\mathbf{\Lambda_1} = \mathbf{\Lambda}\mathbf{T}$, for any orthogonal $\mathbf{T}$. If one allows correlated common factors, then even more solutions are possible. Starting from a given $\mathbf{\Lambda}$, such that $\mathbf{x} = \mathbf{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta}$, we see that it is also true that $\mathbf{x} = \mathbf{\Lambda}_1\boldsymbol{\xi}_1 + \boldsymbol{\delta}$, where $\mathbf{\Lambda}_1 = \mathbf{\Lambda}\mathbf{T}$ (for any nonsingular $\mathbf{T}$) and $\boldsymbol{\xi}_1 = \mathbf{T}^{-1}\boldsymbol{\xi}$. Thurstone "solved" this very significant problem with his

"simple structure criterion," which was essentially a parsimony principle for choosing a $\mathbf{\Lambda}$ that made the resulting factors easy to interpret. Thurstone concluded that the common factor model was

> most appropriately applied when, for any given observed variable, the model used only the smallest number of parameters (factors) to account for the variance of the variable. Thus, if in a factor analysis of n variables r common factors were obtained, Thurstone deemed the factor solution ideal when each variable required fewer than r factors to account for its common variance. By the same token, when it came to interpreting the common factors by noting the observed variables associated with each respective factor, parsimony of interpretation could be obtained when each factor was associated with only a few of the observed variables. [Mulaik, 1972, p. 218]

Translated into an operational definition, simple structure meant that a "good $\mathbf{\Lambda}$" should satisfy the following (in an m-factor orthogonal solution):

(a) Each row of $\mathbf{\Lambda}$ should have at least 1 zero.

(b) Each column of $\mathbf{\Lambda}$ should have at least m zeros.

(c) For every pair of columns of $\mathbf{\Lambda}$, there should be several "nonmatching" zeros, i.e., zeros in different rows.

(d) When 4 or more factors are obtained, each pair of columns should have a large proportion of corresponding zero entries.

In the early days of factor analysis, rotation of the initial $\mathbf{\Lambda}$ to a "best simple structure" $\mathbf{\Lambda_1} = \mathbf{\Lambda T}$ was an art, requiring careful calculation and substantial patience. Development of "machine rotation" methods and digital computers elevated factor analysis from the status of an esoteric technique understood and practiced by a gifted elite, to a technique accessible (for use and misuse) to virtually anyone. Perhaps lost in the shuffle was the important question of why one would expect to find "simple structure" in many variable systems.

3. *Factor Indeterminacy.* If the first two problems are overcome, a third one remains. Specifically, the common and unique factors $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ are not uniquely defined, even if $\mathbf{\Lambda}$ and $\mathbf{U}^2$ are. To see this, suppose the factors are orthogonal, and so $\mathbf{\Psi} = \mathbf{I}$. Then consider *any* $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ constructed via the formulas

$$\boldsymbol{\xi} = \mathbf{\Lambda}'\mathbf{\Sigma}^{-1}\mathbf{x} + \mathbf{Ps} \tag{6.6}$$

and

$$\boldsymbol{\delta} = \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{x} - \mathbf{U}^{-1}\mathbf{\Lambda}\mathbf{Ps} \tag{6.7}$$

where $\mathbf{s}$ is any arbitrary random vector satisfying

$$\mathcal{E}(\mathbf{s}\mathbf{s}') = \mathbf{I} \tag{6.8}$$

and

$$\mathcal{E}(\mathbf{s}\mathbf{x}') = \mathbf{0} \tag{6.9}$$

$\mathbf{P}$ is an arbitrary Gram-factor satisfying

$$\mathbf{P}\mathbf{P}' = \mathbf{I} - \mathbf{\Lambda}'\mathbf{\Sigma}^{-1}\mathbf{\Lambda} \tag{6.10}$$

It is easy to verify, using matrix expected value algebra, that any $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ satisfying Equations 6.6–6.10 will fit the common factor model. Once $\mathbf{\Lambda}$ is known, $\mathbf{P}$ can be constructed easily via matrix factorization methods. $\mathbf{s}$ is a completely arbitrary random vector in the space orthogonal to that occupied by $\mathbf{x}$.

Equation 6.6 shows that common factors are not determinate from the variables in the current analysis. There is an infinity of possible candidates for $\boldsymbol{\xi}$. Each has the same "determinate" component $\mathbf{\Lambda}'\mathbf{\Sigma}^{-1}\mathbf{x}$, but different "arbitrary component" $\mathbf{Ps}$. These candidates for $\boldsymbol{\xi}$ each have the same covariance relationship with $\mathbf{x}$, but possibly differ substantially from each other.

Since dozens of papers have been written on the topic of factor indeterminacy, it is obviously difficult to summarize briefly. There are a number of misconceptions about factor indeterminacy. For example, though some writers seem to believe factor indeterminacy is a "sampling" problem, and refer to it as the "factor score indeterminacy" problem, the above equations demonstrate it exists at the population level. Moreover, it is not without practical implications. For historical reviews of the factor indeterminacy issue, consult Steiger and Schönemann (1978) and Steiger (1979). These papers press the view that, since factor analysis was, in its early days, a "methodology with a mission," (i.e., to promote the views of Spearman and Thurstone on the structure of human abilities), the tendency was for its proponents to ignore its indeterminacy problems. Certainly, there was a surprising absence of discussion of factor indeterminacy in texts by Spearman (1927), Thurstone (1935, 1947), and Harman (1960, 1967).

Lovie and Lovie (1995) present some very interesting and valuable historical details on the relationship between Charles Spearman and E. B. Wilson. Wilson was a professor at Harvard in 1927 when Spearman arrived on a speaking tour to promote his new book, "The Abilities of Man." Wilson had already developed some preliminary notions about factor indeterminacy, but Spearman did not seem inclined to listen to them. His initial interaction with Spearman so galvanized Wilson that he spent much of his Christmas holiday working on a commentary on

Spearman's work. The position Lovie and Lovie (1995) take on the events was that Spearman and Wilson were cooperative collaborators, and that the historical account of Steiger and Schönemann (1978) was overly partisan. Steiger (1996a) presses the view that Lovie and Lovie (1995) offer a rather sugar-coated view of Spearman's motivations (and Wilson's as well). The simple fact was, "The Abilities of Man" represented, for Spearman, the culmination of two decades of mathematical and empirical work. Wilson's discovery of factor indeterminacy came at a time that, for Spearman, was, to say the least, inopportune.

For a modern discussion of factor indeterminacy, see the special issue of *Multivariate Behavioral Research* devoted to a provocative "target" paper by Maraun (1996c) and a series of responses by various authors, and Maraun's rebuttal and commentary (Maraun, 1996b, 1996a).

The following examples illustrate each of the 3 non-uniqueness problems discussed above.

Example 6.1 (Unidentified $\mathbf{U}^2$) Consider the following correlation matrix:

$$\mathbf{R} = \begin{bmatrix} 1.00 & 0.25 \\ 0.25 & 1.00 \end{bmatrix}$$

Suppose we wish to fit a single common factor model to these data. The model will be of the form

$$\mathbf{R} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{U}^2$$

In this case, the model is so simple, we can solve it as a system of simultaneous equations. Specifically, you can show that, for

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix},$$

any λ_1 and λ_2 satisfying

$$\lambda_1 \lambda_2 = r,$$

and also satisfying the side conditions that

$$0 < \lambda_i^2 < 1, \quad i = 1, 2$$

will yield an acceptable solution, with diagonal elements of $\mathbf{U}^2$ given by

$$u_i^2 = 1 - \lambda_i^2$$

So, for example, two acceptable solutions, as you may verify, are

$$\boldsymbol{\lambda} = \begin{bmatrix} .5 \\ .5 \end{bmatrix}, \quad \mathbf{U}^2 = \begin{bmatrix} .75 & 0 \\ 0 & .75 \end{bmatrix}$$

and

$$\boldsymbol{\lambda} = \begin{bmatrix} 3/4 \\ 1/3 \end{bmatrix}, \quad \mathbf{U}^2 = \begin{bmatrix} 7/16 & 0 \\ 0 & 8/9 \end{bmatrix}$$

Example 6.2 (Rotational Indeterminacy) Suppose you factor analyze 6 tests, 3 of which are supposed to be measures of verbal ability, and 3 of which are supposed to be measures of mathematical ability. You factor analyze the data, and are given an "unrotated factor pattern" that looks like the following.

$$\mathbf{\Lambda} = \begin{bmatrix} .424 & .424 \\ .354 & .354 \\ .283 & .283 \\ .424 & -.424 \\ .354 & -.354 \\ .283 & -.283 \end{bmatrix}$$

It looks like all 6 of the tests load on the first factor, which we might think of as a "general intelligence factor," while the 3 verbal tests (in the first 3 rows of the factor pattern) load negatively on the second factor, while the 3 mathematical tests load positively. It seems that the second factor is some kind of "mathematically and not verbally inclined" factor!

Of course, there are, as we mentioned above, infinitely many *other* factor patterns that fit the data as well as this one, i.e., produce the identical product $\mathbf{\Lambda\Lambda}'$. Simply postmultiply $\mathbf{\Lambda}$ by any 2×2 orthogonal matrix $\mathbf{T}$, for example, and you will obtain an alternative $\mathbf{\Lambda}_1 = \mathbf{\Lambda T}$.

The family of 2×2 orthogonal matrices is of the form

$$\mathbf{T} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

where θ is the "angle of rotation."

To see where the term "rotation" comes from, suppose we draw a plot of the 6 variables in "common factor space" by using the factors as our (orthogonal) axes, and the factor loadings as coordinates. We obtain a picture as in Figure 6.1. Note that, in this picture, you can read the factor loadings for any variable by simply reading its coordinates on the two axes.

It is fairly easy to see in the picture that, if the Factors labeled "Factor I" and "Factor II" were simply rotated -45 degrees (to the positions represented by the dotted lines in the drawing), that the points representing the lower right cluster would fall directly on the new Factor II' axis, and would have zero projections onto the new Factor I' axis. Similarly, the points representing the upper right cluster would now fall directly on the revised Factor I' axis, and would have zero projection on the revised Factor II' axis.

To rotate the two factors -45 degrees, we recall that $\cos(45) = .7071$, and $\sin(45) = -.7071$. Hence the rotation matrix is

$$\mathbf{T} = \begin{bmatrix} .7071 & .7071 \\ -.7071 & .7071 \end{bmatrix}$$

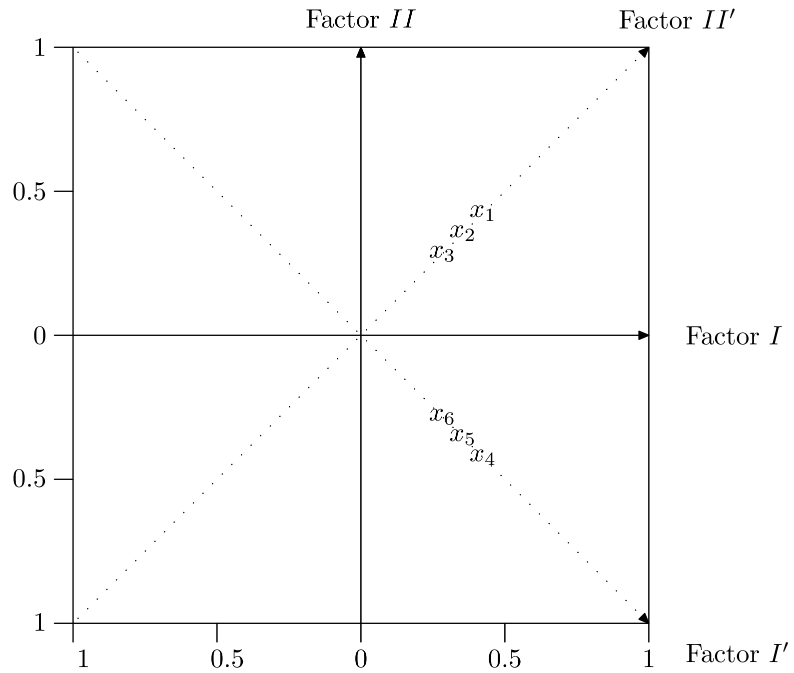

Fig. 6.1 Rotation of Two Orthogonal Factors

You can verify that, when multiplied by the above $\mathbf{T}$, the original $\mathbf{\Lambda}$ is transformed into

$$\mathbf{\Lambda}_1 = \mathbf{\Lambda T} = \begin{bmatrix} 0.0 & 0.6 \\ 0.0 & 0.5 \\ 0.0 & 0.4 \\ 0.6 & 0.0 \\ 0.5 & 0.0 \\ 0.4 & 0.0 \end{bmatrix}$$

This new factor pattern exhibits what Thurstone called "simple structure." Moreover, it agrees with our theoretical expectations, in that the 3 verbal tests load on one factor (evidently a "verbal factor"), and the 3 mathematical tests load on a second factor, which evidently represents mathematical ability.

When there are only two common factors, it is possible, by plotting the loadings, to "graphically rotate" the factors into simple structure, if such a rotation is possible. The early factor analysts, without access to the powerful computers we take for granted, made good use of this fact.

Example 6.3 (Factor Indeterminacy) Even if $\mathbf{U}^2$ is identified and $\mathbf{\Lambda}$ exhibiting satisfactory simple structure is found, the factors themselves are not uniquely determined, as the following example, taken from Steiger (1996b) shows. Suppose that the entire population of observations consists of

$$\mathbf{X} = \begin{bmatrix} 0.905 & 1.641 & 0.203 & -1.401 \\ -0.591 & -0.598 & -0.929 & -0.192 \\ -0.501 & 0.370 & 1.848 & 1.752 \\ -0.488 & -0.495 & 0.740 & -0.402 \\ -0.785 & -1.101 & -0.074 & -0.794 \\ -1.598 & 1.216 & -0.404 & -0.900 \\ 0.749 & 0.514 & -1.703 & 1.084 \\ -0.079 & -0.343 & -0.727 & 1.454 \\ 2.132 & 0.576 & 1.226 & -0.001 \\ 0.255 & -1.779 & -0.182 & -0.960 \end{bmatrix}$$

The above matrix may be conceptualized as the entire population of observations, in the sense that each of the 10 row vectors has an equal probability of occurrence. So the matrix represents the full set of outcomes in a discrete multivariate distribution where each of the 10 outcomes has probability of occurrence of 1/10.

In that case, we have

$$\mathbf{\Sigma} = \begin{bmatrix} 1.00 & 0.20 & 0.15 & 0.10 \\ 0.20 & 1.00 & 0.12 & 0.08 \\ 0.15 & 0.12 & 1.00 & 0.06 \\ 0.10 & 0.08 & 0.06 & 1.00 \end{bmatrix}, \quad \mathbf{\Sigma}^{-1} = \begin{bmatrix} 1.066 & -0.191 & -0.132 & -0.083 \\ -0.191 & 1.054 & -0.094 & -0.060 \\ -0.132 & -0.094 & 1.034 & -0.041 \\ -0.083 & -0.060 & -0.041 & 1.016 \end{bmatrix}$$

Submitting the above $\mathbf{\Sigma}$ to any standard factor analysis program yields the following solutions for $\mathbf{\Lambda}$ and $\mathbf{U}^2$:

$$\boldsymbol{\lambda} = \begin{bmatrix} 0.5 \\ 0.4 \\ 0.3 \\ 0.2 \end{bmatrix}, \quad \mathbf{U}^2 = \begin{bmatrix} 0.75 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.84 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.91 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.96 \end{bmatrix}$$

In order to "construct" a set of common factor scores that agree with the factor model and these data, we need, first of all, to find a component $p\mathbf{s}$ as described in Equations 6.8–6.10. Since there is only one factor, p is a scalar and is equal to the square root of $1 - \boldsymbol{\lambda}'\mathbf{\Sigma}^{-1}\boldsymbol{\lambda}$. After some tedious calculations, we can determine that $p = 0.775$. Hence, the indeterminate part of any common factor is a deviation score vector $p\mathbf{s}$ such that $\mathbf{X}'\mathbf{s} = \mathbf{0}$, $\mathbf{s}'\mathbf{s}/10 = 1$, and $p = 0.775$. Infinitely many such vectors exist. To produce one, simply take a vector of random numbers, convert it to deviation score form, multiply it by the complementary projector $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ to create a vector orthogonal to $\mathbf{X}$, rescale it to the appropriate length, and multiply it by p.

Two such candidates for the "indeterminate part" of the common factor are

$$p\mathbf{s}_1 = \begin{bmatrix} 0.398 \\ -0.284 \\ 0.314 \\ 1.949 \\ -0.055 \\ -0.794 \\ 0.608 \\ -0.232 \\ -0.636 \\ -0.640 \end{bmatrix}, \quad p\mathbf{s}_2 = \begin{bmatrix} 0.258 \\ -0.384 \\ 0.509 \\ -0.759 \\ 1.743 \\ -0.515 \\ 0.755 \\ -0.908 \\ -0.261 \\ -0.437 \end{bmatrix}$$

The determinate part, also known as the "regression estimates" for the factor scores, is computed directly as

$$\mathbf{X\Sigma}^{-1}\boldsymbol{\lambda} = \begin{bmatrix} 0.742 \\ -0.616 \\ 0.491 \\ -0.241 \\ -0.743 \\ -0.485 \\ 0.245 \\ -0.092 \\ 1.261 \\ -0.563 \end{bmatrix}$$

Adding the determinate and indeterminate parts together, we construct two rather different candidates for $\boldsymbol{\xi}$. They are

$$\boldsymbol{\xi}_1 = \begin{bmatrix} 1.140 \\ -0.900 \\ 0.176 \\ 1.708 \\ -0.798 \\ -1.278 \\ 0.853 \\ -0.323 \\ 0.625 \\ -1.203 \end{bmatrix}, \quad \boldsymbol{\xi}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

These candidates for $\boldsymbol{\xi}$ correlate only .399 with each other. It is possible to construct valid candidates for $\boldsymbol{\xi}$ that correlate much less.

Imagine that each of the above $\boldsymbol{\xi}_i$ represent the intelligence scores of the individuals manifesting the associated test scores in $\mathbf{Y}$. We discover that an individual manifesting score pattern $\mathbf{X}_4' = \begin{bmatrix} -0.488 & -0.495 & 0.740 & -0.402 \end{bmatrix}$ has an intelligence score of 1.708 in one version of the factor, and an intelligence score of -1 in another version. It is this singular fact, first discovered by E. B. Wilson, that seemed to compromise, irretrievably, Spearman's high hopes for measuring g.

Schönemann and Wang (1972) showed that, for orthogonal factors, assuming that $\mathbf{\Sigma}$ is a correlation matrix (i.e., that the manifest variables are standardized), the minimum correlation between equivalent factors are given by the diagonal elements of the matrix $2\mathbf{\Lambda}'\mathbf{\Sigma}^{-1}\mathbf{\Lambda} - \mathbf{I}$.

6.5 ALTERNATIVE VIEWS OF FACTOR INDETERMINACY

Maraun (1996c) reviews two fundamentally different ways of thinking about the phenomena demonstrated in Example 6.3. One approach, which he calls the *alternative solution position*, emphasizes, as in the example, that different solutions for a common factor exist, and have different implications.

The second major position discussed by Maraun (1996c), the *posterior moment position*, takes a different viewpoint. It assumes that a unique set of common factors $\boldsymbol{\xi}$ exists, and generated the data we observe via the common factor model. It then investigates the posterior distribution of the common factors, given the observed data. The conclusion is that, with equal prior distributions (in the Bayesian sense), each of the candidates for $\boldsymbol{\xi}$ has equal posterior probability.

Maraun compares the two positions carefully, and, drawing on his background in the philosophy of science, decisively favors the alternative solution viewpoint. Maraun's analysis is deep and detailed, and should be read in its original form. Maraun (1996c) deals in considerable detail with the various metaphors surrounding the use of the common factor model, and, in particular, with the notion of a *latent variable.* Maraun's clarification of the confusion between the *metaphors* of factor analysis, and the mathematics of the method, is a model of clear writing, and should be required reading for any serious student of factor analysis.

Besides the major alternative positions on factor indeterminacy, there have been a number of *responses* to the problem within the factor analytic community. These are discussed in detail by Steiger and Schönemann (1978), Steiger (1990), Maraun (1996c). Two are particularly noteworthy:

1. *Factors cannot be computed, they can only be estimated.* Although factors could indeed be computed, they could not be computed uniquely (because they are not unique). Rather than face this embarassing reality, a number of proponents of factor analysis sought other ways of thinking about the situation. One very confused line of thought led to the derivation of numerous "factor score estimates." For example, consider any common factor, such as $\boldsymbol{\xi}_1$ in Example 6.3. From standard regression algebra, the best least squares predictor of $\boldsymbol{\xi}_1$ from the variables in $\mathbf{X}$ is given by $\mathbf{X}\mathbf{\Sigma}^{-1}\boldsymbol{\lambda}$. This linear combination of the manifest variables in $\mathbf{X}$ is called the "regression estimate" of the common factor scores. It is easy to verify that, when the common factors are orthogonal, their regression estimates are not. This led to a variety of other types of estimators, each with "optimal" properties.

It can be shown that the regression estimates of the common factors are the average of all possible true factors that can be computed. To understand the fallacy of computing factor score estimates, Steiger (1990) proposed the following analogy. Suppose you have a model, called Model A. It expresses obsrved variable X in terms of a single underlying latent variable Y. The model is that

$$Y^2 - 100Y + X = 0$$

Now suppose X has a discrete distribution. Only two values ever occur. They are 99 and 2100. In this case, what can be said about latent variable Y. Suppose, for example, X is 99. Then there are two solutions for Y. We can have either $Y = 1$ or $Y = 99$.

There are many precise analogies between the indeterminacy that affects this model, and that which affects factor analysis. If all we know about Y is that it is a random variable that satisfies the model, then there are two satisfactory Y values. One could say that the "posterior distribution of Y given X assigns probability .50 to the values 99 and 1. However, to say that the "best regression estimate of Y" is 50, the average of 99 and 1, is misleading. Statistical estimation occurs when a unique parameter is approximated on the basis of a statistic, a function of a sample. In this case, there is no need for statistical estimation, as all possible values of the parameter are known. Suppose Y actually represents some kind of percentage performance. We can say that, according to the model, performance is either very bad or very good. To estimate it at 50% is both unnecessary and misleading.

2. *Factor indeterminacy vanishes in an infinite domain.* In this modification of the common factor model, the p variables currently under analysis are part of an infinite behavior domain, in which the common factor model holds. If this is true, one can make indeterminacy as small as one wishes by simply sampling enough additional tests. This view, proposed by, among others, Williams (1978), was received enthusiastically by many proponents of factor analysis. Some obvious rejoinders to this position are that (a) it is a new model, not a solution to the problems with the factor model, (b) it is a model about data that have never been observed, and so is never, at any point in time, testable. Moreover, it is extremely unlikely, *a priori* to actually hold in any set of variables.

Problems

6.1 Prove, using standard expected value algebra, that if $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ satisfy Equation 6.6–Equation 6.10, that they satisfy all the restrictions of the common factor model, i.e., they satisfy Equation 6.1 and all the orthogonality conditions on $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ are satisfied as well.

6.2 Re-express Equation 6.1 in the form $\mathbf{X} = \mathbf{C}'\mathbf{w}$ by defining $\mathbf{C}'$ and $\mathbf{w}$ as the appropriate partitioned matrix forms.

6.3 Using the basic results on the weight matrix for least squares regression estimates, derive the formula for "regression estimates" of the common factors in $\boldsymbol{\xi}$, i.e., $\hat{\boldsymbol{\xi}} = \boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\mathbf{x}$. Assume that the common factors are orthogonal. (*Hint.* Simply translate the formula in Equation 5.19, considering the observed variables in $\mathbf{x}$ to be the predictors, and the common factors in $\boldsymbol{\xi}$ to be the criterion.)

References

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability* (Vol. 5, pp. 111–150). Berkeley, CA.

Garnett, J. C. M. (1919). On certain independent factors in mental measurement. *Proceedings of the Royal Society of London, Series A*, *96*, 91–111.

Harman, H. H. (1960). *Modern factor analysis.* Chicago, IL: The University of Chicago Press.

Harman, H. H. (1967). *Modern factor analysis* (2nd ed.). Chicago, IL: The University of Chicago Press.

Hart, B., & Spearman, C. (1912). General ability: Its existence and nature. *British Journal of Psychology*, *5*, 51–84.

Lovie, P., & Lovie, A. D. (1995). Spearman and Wilson on factor indeterminacy. *British Journal of Mathematical and Statistical Psychology*, *48*, 237–253.

Maraun, M. D. (1996a). The claims of factor analysis. *Multivariate Behavioral Research*, *31*, 673–689.

Maraun, M. D. (1996b). Meaning and mythology in the factor analysis model. *Multivariate Behavioral Research*, *31*, 603–616.

Maraun, M. D. (1996c). Metaphor taken as math: Indeterminacy in the factor analysis model. *Multivariate Behavioral Research*, *31*, 517–538.

Mulaik, S. A. (1972). *The foundations of factor analysis.* New York, NY: McGraw-Hill.

Schönemann, P. H., & Wang, M.-M. (1972). Some new results on factor indeterminacy. *Psychometrika*, *37*, 47–51.

Spearman, C. (1927). *The abilities of man.* London: MacMillan.

Steiger, J. H. (1979). Factor indeterminacy in the 1930's and in the 1970's: Some interesting parallels. *Psychometrika*, *44*, 157–167.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Steiger, J. H. (1994). Factor analysis in the 1980's and the 1990's: Some old debates and some new developments. In *Trends and perspectives in empirical social research* (pp. 201–223). DeGruyter.

Steiger, J. H. (1996a). Coming full circle in the history of factor indeterminacy. *Multivariate Behavioral Research*, *31*, 617–630.

Steiger, J. H. (1996b). Dispelling some myths about factor indeterminacy. *Multivariate Behavioral Research*, *31*, 539–550.

Steiger, J. H., & Schönemann, P. H. (1978). A history of factor indeterminacy. In *Theory construction and data analysis in the behavioral sciences* (pp. 136–178). Jossey-Bass.

Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, *8*, 271–281.

Thurstone, L. L. (1935). *The vectors of mind.* Chicago, IL: The University of Chicago Press.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago, IL: The University of Chicago Press.

Williams, J. S. (1978). A definition for the common factor analysis model and the elimination of problems of factor score indeterminacy. *Psychometrika*, *43*, 293–306.