Estimation of Principal Components and Related Models by Iterative Least Squares

HERMAN WOLD
UNIVERSITY INSTITUTE OF STATISTICS
UPPSALA, SWEDEN

SUMMARY

For the estimation of principal components an iterative procedure based on the principle of least squares has recently been reported [19]. Lyttkens in his report to this Symposium [7] developed the rationale of the new technique, and showed the close connection between two approaches in multivariate analysis, (a) principal components [3, 8]; and (b) Young-Whittle's approach in factor analysis [12, 21]. The present paper focuses on the flexibility of the iterative method in adaptations and extensions in various directions. Section 1 deals with factor models with one factor and equal residual variances, and analyzes the component structure in terms of the factor structure. The ensuing relations are exploited in Section 2 to estimate the one-factor model, making use of the least-squares estimate of the first principal component. Section 3 extends the iterative method to a multidimensional approach of principal components, and to hybrid models that involve multiple regression and canonical correlations.

1. COMPONENT STRUCTURE vs. FACTOR STRUCTURE

1.1. Introduction

Sections 1 and 2 of the present paper were planned as a counterpart to [13], a Monte Carlo illustration of Whittle's application of principal-component analysis for the estimation of Young-Whittle's factor model [12], the incentive for a reconsideration being to illustrate the iterative estimation technique recently reported [7, 19]. Specific attention is now paid to similarities and differences between component analysis and factor analysis. Whereas component models and factor models under suitable conditions (equal variances of the factor residuals) have the same numerical values for the parameters β that constitute the loadings in the factor model and the direction

cosines in the component model, the two approaches differ when it comes to residuals, factors, and components. The key point is that the *models* differ; hence they call for different estimation techniques. The parting of the ways was not clear to me at the time of the Uppsala Symposium on Psychological Factor Analysis (1953), where [13] was presented and discussed. Therefore, I made it a point to emphasize the differences at issue in the first draft of the present paper as reported and discussed at the Dayton Symposium on Multivariate Analysis. By a suitable transformation, as shown in Section 2, estimates of component structure can be used to obtain estimates of factor structure. In the first draft of the paper, the transformation from component structure to factor structure was carried out in one stroke, by a joint treatment of population properties and sample properties. In the final version, to make for clarity, the transformation is carried out separately for the population and the sample.

The differences between component analysis and factor analysis are by no means a new feature. Specific reference is made to the clarifying paper by Rao [9]. To quote a statement of Rao's that bears upon collective analysis of the variates, principal components aim at reproducing their entire moment matrix, theoretical μ_{ij} or observed m_{ij}, whereas factor analysis only aims at reproducing the nondiagonal elements, $i \neq j$. Thus it is clear from Rao's review that the parting of the ways lies in the model itself, and is not primarily a matter of different estimation techniques, such as maximum likelihood vs. least squares. To some extent the differences between the two types of models have been obscured by other issues, notably collective analysis vs. individual analysis, and the representation of the individual factors or components as parameters vs. random variates.

The approach of the present paper is limited in two respects. One is that the method works only for factor models with equal (or known) residual variances. The other is that the estimation of factor structure provides estimates not of the individual factor values ζ_t but of the conditional expectation of ζ_t for given individual component ξ_t. In practice it is an important problem to estimate the factor structure when the residual variances may be different. For this extension reference is made to the techniques designed by Jöreskog [4; 5; 19, Chap. 8]. It is an interesting question how the estimation of factor structure in Section 2 compares with Jöreskog's and other techniques. This is, however, an order of problems that falls outside the scope of the present paper.

1.2. Principal Components vs. Factor Representation

Throughout the paper we shall be concerned with one and the same set of variates—theoretical,

$$\eta_1, \cdots, \eta_p \tag{1a}$$

or observed in a sample of n observations,

$$y_{i1}, \ldots, y_{in}; \qquad i = 1, \ldots, p \tag{1b}$$

The number of variates, p, will be kept fixed. We shall consider two theoretical representations,

$$\eta_i = \sum_{a=1}^{k'} \beta_{ia}\xi_a + \varepsilon_i \tag{2}$$

$$= \sum_{a=1}^{k} \beta_{ia}\zeta_a + \delta_i \tag{3}$$

where the parameters β_{ia} are assumed to satisfy conditions

$$\sum_{i=1}^{p} \beta_{ia}^2 = 1 \tag{4a}$$

$$\sum_{i=1}^{p} \beta_{ia}\beta_{ib} = 0; \qquad a \neq b. \tag{4b}$$

Here and in the following, the unspecified integers run as follows:

$$a, b = 1, \ldots, k \text{ or } k'; \qquad i, j = 1, \ldots, p; \qquad t, u = 1, \ldots, n.$$

Model (2) will be specified as a component structure, model (3) as a factor structure. Hence the parameters β_{ia} figure as direction cosines of the principal components ξ_a in model (2), and as loadings of the factors ζ_a in model (3). To bring the similarity between approaches (2) and (3) in relief, we deviate somewhat from current usage, inasmuch as the normalization (4a) is current in component analysis, whereas factor analysis usually adopts the normalization

$$E(\zeta_a^2) = 1.$$

Any two components ξ_a and ξ_b are assumed to have zero product moments, just as is customary with factor models,

$$E(\xi_a\xi_b) = 0, \quad E(\zeta_a\zeta_b) = 0; \qquad a \neq b \tag{5}$$

and it is known that this involves no loss of generality (see [12]).

Both representations (2) and (3) will be specified in terms of conditional expectations; thus

$$E(\eta_i \mid \xi_1, \ldots, \xi_{k'}) = \sum_{a=1}^{k'} \beta_{ia}\xi_a \tag{6}$$

$$E(\eta_i \mid \zeta_1, \ldots, \zeta_k) = \sum_{a=1}^{k} \beta_{ia}\zeta_a. \tag{7}$$

Hence we shall say that the systematic part (6) in relation (2) is a *predictor*

of the variate η_i, and similarly for the systematic part (7) of (3); see [18]. Assumption (6) implies

$$E(\varepsilon_i) = 0\,; \qquad E(\xi_a \varepsilon_i) = 0 \tag{8}$$

and similarly for (7),

$$E(\delta_i) = 0\,; \qquad E(\zeta_a \delta_i) = 0\,. \tag{9}$$

In model (2) the components ξ_a are assumed to allow the reverse representation

$$\xi_a = \sum_{i=1}^{p} \beta_{ia} \eta_i\,. \tag{10a}$$

Substituting (10a) to the right in (2), reduction by the use of (4) gives

$$\sum_{i=1}^{p} \beta_{ia} \varepsilon_i = 0\,. \tag{10b}$$

The equivalent relations (10a and b) will be referred to as the *component property* of the model (cf. [7]).

In model (3) the residuals δ_i are assumed to have zero product moments:

$$E(\delta_i \delta_j) = 0 \tag{11}$$

which in conjunction with (9a) implies that the factor residuals are mutually uncorrelated; this will be referred to as the *factor property* of the model.

Comments.[1] (a) Reference is made to the geometric interpretation of the component model (2) in terms of lines and planes of closest fit in the joint distribution of the variates under analysis, theoretical (1a) or observed (1b) [3, 8, 12]. The first component ξ_1 $(a = 1)$ has the largest variance and gives the line of closest fit in the sense of least squares. If the joint distribution is nonsingular, the plane of closest fit is orthogonal to the pth component.

The classic estimation method for component structure [3, 8] gives the variances and loadings of the components as eigenvalues and eigenvectors of the product-moment matrix μ, m of the variates under analysis, theoretical

$$\mu = [\mu_{ij}]\,; \qquad \mu_{ij} = E(\eta_i \eta_j) \tag{12a}$$

or observed,

$$m = [m_{ij}]\,; \qquad m_{ij} = \frac{1}{n} \sum_{t=1}^{n} y_{it} y_{jt}\,. \tag{12b}$$

[1] The reader is assumed to have a general orientation about factor and component analysis. The introductory sections of [5] are very instructive for the purpose. For the general background in multivariate analysis, see [1, 6].

(b) Factor models originate from experimental psychology [10, 11]. Distinction is made between *individual* and *collective* analysis; that is, individual representation of the variates η_i as in (2) or (3) vs. models formulated in terms of product moments (12a). Speaking generally, there is no difference in principle between an individual model and the corresponding collective model. Factor models are usually specified in individual terms; factor estimation techniques make use of observed product moments (12b), and the statement of numerical results and verbal conclusions is often oriented toward collective rather than individual aspects. Component models are usually formulated in collective terms. The work of Whittle [12] was pioneering in that it emphasized the equivalence in the component approach between the basic assumptions of collective analysis and individual analysis.

We see that the factor and component properties (10) and (11) are mutually exclusive. This is clearly so, since (10b) implies

$$E\left[\left(\sum_{i=1}^{p} \beta_{ia}\varepsilon_i\right)^2\right] = 0 \tag{13}$$

and this could not be true if the component residuals ε_i and ε_j were to satisfy an assumption of type (11).

Note the difference in current terminology between the numbers k and k' in models (2) and (3). For k-factor models (3) such that all residuals δ_i have the same variance $\sigma^2(\delta_i)$ it has been shown by Whittle [12] that in the corresponding component structure (2), with $k' = p$, there are k components that have larger variance than the remaining $p - k$ ones, all of which have the same variance $\sigma^2(\varepsilon_i)$. In such a case, as shown by Lyttkens [7], the variances of these components are given by

$$E(\xi_a{}^2) = \sum_i \sum_j \beta_{ia}\beta_{ja}E(\eta_i\eta_j); \qquad a = 1, \ldots, k. \tag{14}$$

(c) In specifying models (2) and (3) in terms of predictors as in (6) and (7), rather than in terms of residual properties (8) or (9) as is customary, the operative use of the models for predictive purposes is emphasized (cf. [15, 17, 18]). For given components ξ_a, the parameters β_{ia} have least-squares properties in model (2), that is, the residuals ε_i have the smallest possible standard deviation, and for given factors ζ_a the same β_{ia}'s have least-squares properties in model (3).

If the probability distribution of the variates η_i is jointly normal, specification (6) is equivalent to (8), and (7) equivalent to (9).

1.3. Young-Whittle's Approach with One Factor and Equal Residual Variances

Following Young [21] and Whittle [12], we shall consider a factor model (3) with equal residual variances. To estimate such a model Whittle applies the

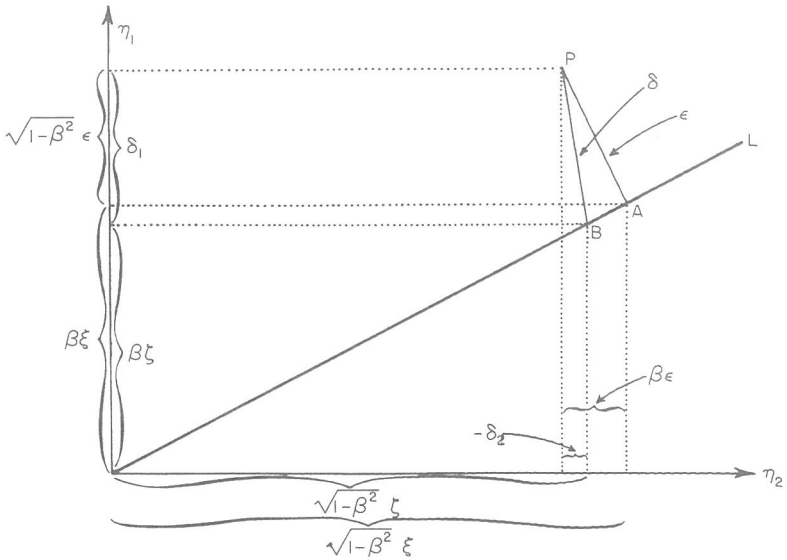

Fig. 1. The dual representation (15) and (16) of two variates η_1 and η_2.

principle of least squares, minimizing the total residual variance. In the case of one factor, as will be briefly reviewed in this section, his estimation procedure gives consistent estimates for the loadings β_i, and a consistent estimate not of the factor ζ but of the first principal component ξ.

The Dual Representation. We form the factor model (3) in the special case of one factor, say ζ,

$$\eta_i = \beta_i\zeta + \delta_i \tag{15a}$$

and assume that all residuals δ_i have the same standard deviation,

$$\sigma(\delta_i) = \text{const.} = \frac{1}{p^{1/2}}\,\sigma(\delta), \tag{15b}$$

where

$$\delta = (\delta_1{}^2 + \cdots + \delta_p{}^2)^{1/2} \tag{15c}$$

is the total residual of the one-factor model.

Further we shall represent the variates η_i in terms of component model (2) with one component ξ and the same parameters β_i,

$$\eta_i = \beta_i\xi + \varepsilon_i. \tag{16a}$$

Here we rewrite the residuals in the normalized form

$$\varepsilon_i = \varphi_i\varepsilon \tag{16b}$$

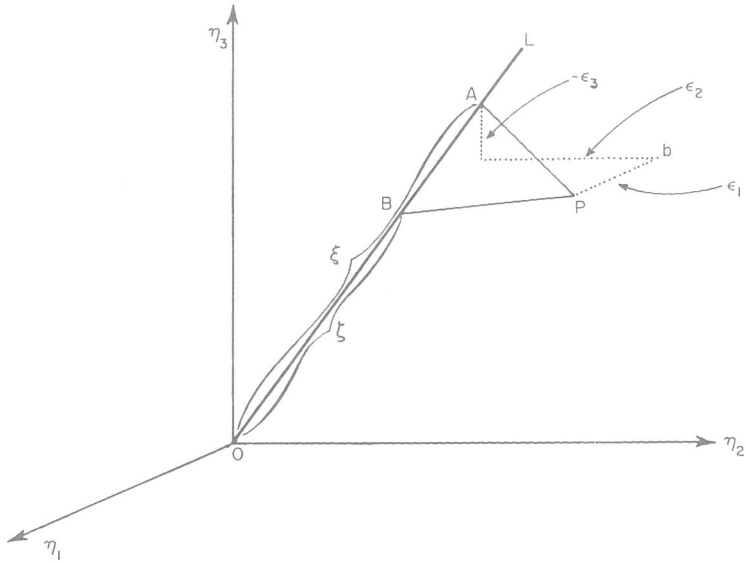

Fig. 2. The dual representation (15) and (16) of three variates η_1, η_2, and η_3.

with

$$\varepsilon = (\varepsilon_1{}^2 + \cdots + \varepsilon_p{}^2)^{1/2}\,; \qquad \varphi_1{}^2 + \cdots + \varphi_p{}^2 = 1 \tag{16c}$$

where ε is the total residual of the component representation (16a).

The dual representation (15) and (16) of the variates η_i is illustrated in Fig. 1 for $p = 2$ and in Fig. 2 for $p = 3$. The graphs show an arbitrary point $P = (\eta_1, \ldots, \eta_p)$ as generated from model (15) with factor $\zeta(=OB)$ and residuals δ_i. The total factor residual δ is given by the line PB. The projection of P on the line L that extends OB gives the point A and thereby the component $\xi(=OA)$. The line PA gives the total component residual ε, and the projection of PA on the coordinate axes give the component residuals ε_i. We see that the coefficients φ_i may be interpreted as direction cosines of ε. Further we note that the φ_i's for $p > 2$ are random relative to the direction cosines β_i, and have $p - 2$ degrees of freedom.

Orthogonality Properties of Component and Factor Residuals. With reference to Figs. 1 and 2, three different notions of orthogonality are in play. We shall see that component and factor residuals have one mode of orthogonality in common, whereas the two others are characteristic of component and factor residuals, respectively.

(a) Geometric orthogonality, first mode. According to (15a), each factor residual δ_i is measured along the coordinate axis of the corresponding variate η_i. Since the coordinate axes are mutually orthogonal, the residuals δ_i

are orthogonal in the same geometric sense. The situation is the same for the component residuals ε_i. Hence component and factor residuals have this first mode of orthogonality in common.

(b) Geometric orthogonality, second mode. The total component residual ε is the orthogonal distance from the observational point P to the k'-dimensional hyperplane spanned by the k' components ξ_a. This mode of orthogonality is characteristic of the total component residual ε, and accordingly is not shared by the total factor residual δ.

(c) Stochastic orthogonality. The factor residuals δ_i have zero expectation and zero product moments, giving

$$E(\delta_i\delta_j) = E(\delta_i)E(\delta_j) = 0. \tag{17}$$

Hence they are mutually orthogonal in the stochastic sense. Similarly, if we interpret the residuals δ_i as elements in a Hilbert space with inner products (δ_i, δ_j) defined by the product moments $E(\delta_i\delta_j)$, the elements δ_i and δ_j are mutually orthogonal in the terminology of Hilbert-space geometry; that is, if we project one element upon another, the projection will be zero. This mode of orthogonality is characteristic of the factor residuals δ_i; as pointed out in connection with (13), it is not shared by the component residuals ε_i.

Component Structure (16) *Expressed in Terms of One-Factor Structure* (15).

Theorem 1. *For any one-factor Young-Whittle model* (15) *the relations*

$$\xi = \sum_i \beta_i\eta_i \tag{18}$$

$$\varepsilon_i = \eta_i - \beta_i\xi \tag{19}$$

whence

$$\varepsilon = (\varepsilon_1{}^2 + \cdots + \varepsilon_p{}^2)^{1/2}; \qquad \varphi_i = \varepsilon_i/\varepsilon \tag{20}$$

define a component representation (16) *with the following properties*:

$$E(\eta_i{}^2) = \beta_i{}^2E(\zeta^2) + \frac{1}{p}\sigma^2(\delta) \tag{21}$$

$$= \beta_i{}^2E(\xi^2) + \sigma^2(\varepsilon_i{}^2) \tag{22}$$

$$E(\xi) = E(\zeta) \tag{23a}$$

$$E(\xi^2) = E(\zeta^2) + \frac{1}{p}\sigma^2(\delta) \tag{23b}$$

$$\sigma(\varepsilon) = \left(\frac{p-1}{p}\right)^{1/2}\sigma(\delta) \tag{24}$$

$$E(\xi\zeta) = E(\zeta^2). \tag{25}$$

Proof. Relations (18) to (20) are immediate implications of (10a) and the normalized expression (16c) for the component residuals ε_i. Relations (21) and (22) follow from definitions (15) and (16) and the general properties (8) and (9).

Substituting (15a) in (18) we obtain

$$\xi = \zeta + \sum_i \beta_i \delta_i \tag{26}$$

which gives (23a). Further, making use of (9) and (15b),

$$E(\xi^2) = E(\zeta^2) + \sigma^2(\delta_i) \qquad E(\xi\zeta) = E(\zeta^2)$$

which verifies (23b) and (25). Finally, summing over i in (21) and (22) and paying regard to (23),

$$\sum_i E(\eta_i{}^2) = E(\zeta^2) + \sigma^2(\delta) = E(\xi^2) + \sigma^2(\varepsilon) \tag{27a}$$

$$= E(\zeta^2) + \frac{1}{p}\sigma^2(\delta) + \sigma^2(\varepsilon) \tag{27b}$$

where (27a and b) imply (24). Theorem 1 is proved.

Comment. Writing

$$\xi^* = E(\xi \mid \zeta) \tag{28}$$

we note that (26) gives the conditional expectations

$$\xi^* = \zeta \qquad E[(\xi^*)^2 \mid \zeta] = E(\xi^*\zeta \mid \zeta) = \zeta^2.$$

This gives the unconditional expectations

$$E(\xi^*) = E(\zeta) \qquad E[(\xi^*)^2] = E(\xi^*\zeta) = E(\zeta^2).$$

1.4. NILES Estimation of Principal Components[2]

We shall consider a $p \times m$ array of observations (1b) which we assume to be a sample of variates (1a) ruled by the dual model (2) and (3). For two models as estimated from the sample we write

$$y_{it} = \sum_{a=1}^{k'} b_{ia} x_{at} + e_{it} \tag{29}$$

$$= \sum_{a=1}^{k} b_{ia} z_{at} + d_{it}. \tag{30}$$

The NILES procedure gives estimates b_{ia}, x_{ia}, e_{it} of parameters, components, and residuals in the component model (2). In Section 2 we shall

[2] NILES = Nonlinear Iterative Least Squares. See [19]; also cf. [16]

adapt the procedure to obtain estimates for the elements z and d of the factor model (3).

NILES estimation of the component model (2) *is an iterative regression procedure that may be summarized as follows.*

(a) The procedure works stepwise, estimating component by component, $a = 1, \ldots, k'$. In each step the component $x_{a1}, \ldots, x_{an}$ and its direction cosines $b_{1a}, \ldots, b_{pa}$ are estimated by the iterative procedure. Thus when the $(a-1)$st component and its cosines have been estimated, the procedure obtains a sequence of iterative estimates of component and cosines in the ath step,

$$x_{at}^{(s)}, b_{ia}^{(s)}; \qquad s = 1, 2, \ldots \tag{31a}$$

giving in the limit

$$x_{at} = \lim x_{at}^{(s)}; \qquad b_{ia} = \lim b_{ia}^{(s)} \qquad (s \to \infty) \tag{31b}$$

as the NILES estimates for the ath component ξ_a and its direction cosines β_{ia} in model (2).

(b) To start the iteration (31) we make $s = 1$, and take for $b_{ia}^{(1)}$ a set of p arbitrary numbers

$$b_{1a}^{(1)}, \ldots, b_{pa}^{(1)} \qquad \text{such that} \quad (b_{1a}^{(1)})^2 + \cdots + (b_{pa}^{(1)})^2 = 1. \tag{32}$$

For general s, when the cosines proxy

$$b_{1a}^{(s)}, \ldots, b_{pa}^{(s)} \qquad \text{with} \quad (b_{1a}^{(s)})^2 + \cdots + (b_{pa}^{(s)})^2 = 1 \tag{33}$$

has been calculated, the iteration (31) proceeds with the following "criss-cross" regressions:

First, we fix t consecutively $(t = 1, \ldots, n)$ and calculate for each t the regression of the tth column y_{it} $(i = 1, \ldots, p)$ on the cosines (33). This gives the regression coefficients

$$x_{at}^{(s+1)} = \sum_{i=1}^{p} b_{ia}^{(s)} y_{it}; \qquad t = 1, \ldots, n \tag{34}$$

which we take for the $(s+1)$st component proxy.

Second, we fix i consecutively $(i = 1, \ldots, p)$ and calculate for each i the regression of the ith row y_{it} $(t = 1, \ldots, n)$ on the component proxy (34). This gives regression coefficients

$$B_{ia}^{(s+1)} = \sum_{t=1}^{n} x_{at}^{(s+1)} y_{it} \Big/ \sum_{t=1}^{p} (x_{at}^{(s+1)})^2; \qquad i = 1, \ldots, p \tag{35a}$$

which we use as auxiliary entities to calculate the $(s+1)$st cosines proxy,

$$b_{ia}^{(s+1)} = \frac{B_{ia}^{(s+1)}}{\left[\sum_{j=1}^{p} (B_{ja}^{(s+1)})^2\right]^{1/2}}; \qquad i = 1, \ldots, p. \tag{35b}$$

In words, we normalize the coefficients B_i to a unit square sum to obtain the cosines b_i.

(c) When the calculations (32) to (35) have given estimates for the first components and their direction cosines, say $a = 1, \ldots, h$, the $(h + 1)$st round begins by calculating the residuals in the hth step,

$$e_{ith} = y_{it} - \sum_{a=1}^{h} b_{ia}x_{at}\,; \qquad \begin{matrix} i = 1, \ldots, p \\ t = 1, \ldots, n \end{matrix} \tag{36}$$

Taking the residuals thus defined to be our observations y_{it} in the $(h + 1)$st step, we apply procedure (2) anew, obtaining

$$e_{ith} = b_{i,h+1}x_{h+1,t} + e_{it,h+1} \qquad y_{it} = \sum_{a=1}^{h+1} b_{ia}x_{at} + e_{it,h+1} \tag{37}$$

Comments. (a) *Convergence.* As shown by Lyttkens [7], the NILES procedure of the previous section converges, giving well-defined estimates b_{ia}, x_{at}, e_{it} for the direction cosines β_{ia}, component sample values ξ_{at}, and residual sample values ε_{it}, and the resulting cosines estimates b_{ia} are equivalent to those given by Hotelling's classic method [3]. In his treatment of the least-squares aspects of principal components, Lyttkens leans heavily on the fundamental work of Whittle [12].

(b) *Consistency.* Least-squares estimates of linear predictors are known to be consistent (that is, a parameter estimate b will tend in probability to the corresponding theoretical parameter β as the sample size is allowed to increase indefinitely) under very general conditions of statistical regularity [14]. The argument extends to NILES estimation of component models (2) as specified in terms of predictors (6). The essential requirement is that as the sample size n increases, each observed moment m_{ij} as defined by (12) tends to the corresponding theoretical moment μ_{ij},

$$\lim_{n\to\infty} \text{prob } m_{ij} = \mu_{ij}\,; \qquad i, j = 1, \ldots, p. \tag{38}$$

In different phrasing, condition (38) requires that the sample (1b) under analysis be *ergodic* with respect to the observed product moments m_{ij}.

The key argument in proving the consistency of the NILES estimates (31b) is that in each step of the iterative procedure (34) and (35) every proxy $b_{at}^{(s)}$ is a continuous function of the moments m_{ij}. In conjunction with the ergodicity assumption (38) this implies

$$\lim_{n\to\infty} \text{prob } b_{ia} = \beta_{ia} \tag{39}$$

as stated. In view of (34), the component property (10a) carries over to the least-squares estimates b_{ia} and x_{at}, and gives

$$x_{at} = \sum_{i=1}^{p} b_{ia}y_{it} \tag{40}$$

Hence the consistency (39) of the cosine estimates b_{ia} extends to components and residuals, giving

$$\lim_{n\to\infty} \text{prob } x_{ia} = \xi_{ia}\,; \qquad \lim_{n\to\infty} \text{prob } e_{it} = \varepsilon_{it} \tag{41}$$

Some qualification is needed as to the order of the components and the case of components ξ_a and ξ_b which are equal in mean square, but these exceptional cases can readily be checked and taken into account (see [7]).

2. COMPONENT STRUCTURE AS A BASIS FOR THE ESTIMATION OF ONE-FACTOR STRUCTURE

Considering the one-factor model (15) with equal residual variances, the problem dealt with in this section is to use the structure of the first principal component to assess the factor structure. For the limited purpose of this paper, it suffices to consider the case of one-factor structure, inasmuch as (a) the one-to-one correspondence between factor and largest component in the one-factor structure, (15) and (16), extends to the general case of k-factor models, as we know from the work of Whittle [12] and Lyttkens [7], and (b) the NILES estimates procedure (Section 1.4) gives the components one by one, in the order of the component variance.

The transfer from component to factor structure makes no change in the direction-cosines loading parameters β_i, since these are the same in the component and factor models, (15) and (16). The components and factors differ in the two models, as do the residuals. There is a radical difference here, for the component ξ and the component residuals ε_i can be assessed to be exact linear expressions in the variates η_i, as we know from Theorem 1, whereas the corresponding transfer to factor structure necessarily brings in the unknown factor residuals

$$\delta_i = \eta_i - \beta_i\zeta$$

with the result that the transfer only provides conditional expectations for the factor values ζ.

The population and sampling aspects of the problem will be dealt with separately. As to population properties, we regard the observations (16) under analysis to be generated from the factor model (15) with known theoretical matrix μ_{ij}, unknown loadings β_i, and unknown individual values ζ_t and δ_{it} for the factor and the residuals of the factor structure. We take the loadings β_i to be assessed from the theoretical moments μ_{ij} by Hotelling's classic method. Then relation (10a) provides the corresponding individual values for the first principal component ξ, giving

$$\xi_t = \sum_{i=1}^{p} \beta_i y_{it}\,; \qquad \varepsilon_{it} = y_{it} - \beta_i\xi_t. \tag{42}$$

Writing

$$\zeta_t^* = E(\zeta_t \mid \xi_t) \tag{43a}$$

for the conditional expectations of a factor value ζ_t for known component value ξ_t, and

$$\delta_{it}^* = y_{it} - \beta_i \zeta_t^* \tag{43b}$$

for the corresponding factor residuals, Section 2.1 deals with the assessment of ζ_t^* and δ_t^* in terms of component structure. Furthermore, Section 2.1 is concerned with the collective properties of ζ_t and ζ_t^*, notably the mean square $E(\zeta^2)$.

As to the sampling aspects, we consider representation (29) of the observed sample in terms of the least-squares estimates b_i of the direction cosines of the first principal component and the estimates x_t and e_{it} of the individual values of the component and the component residuals. We take the cosines b_i to be estimated either by Hotelling's classic method or by the NILES procedure (Section 1.4), which in this respect are equivalent. The estimates x_t and e_{it} will be given by

$$x_t = \sum_{i=1}^{p} b_i y_{it}\,; \qquad e_{it} = y_{it} - b_i x_t \tag{44}$$

in accordance with (36) and (40). Writing

$$z_t^*, \quad d_t^*$$

for the sample estimates of the quantities (43a and b), Section 2.2 deals with the correction of the factor-structure estimates for finite-sample bias.

2.1. Population Properties

We shall be concerned with two types of problems:

(a) To estimate collective features of the factor structure in terms of the collective component structure.

(b) To estimate the factor structure when the variates $\eta_1, \ldots, \eta_p$ are known, or, equivalently, when the component ξ is known.

Factor Estimation. The following two simple theorems enter in order under the headings (a) and (b). The first is an immediate implication of relations (23b) and (24).

Theorem 2. *For any one-factor model* (15) *with equal residual variances, the relations*

$$E(\zeta^2) = E(\xi^2) - \frac{1}{p-1} \sum_{i=1}^{p} \sigma^2(\varepsilon_i) \tag{45}$$

gives the factor mean square in terms of the first principal component.

Theorem 3. *A least-squares estimate of the factor ζ for known component ξ is given by*

$$\zeta^* = \lambda\xi \tag{46a}$$

where

$$\lambda = \frac{E(\zeta^2)}{E(\zeta^2) + (1/p)\sigma^2(\delta)} = 1 - \frac{\sigma^2(\varepsilon)}{(p-1)E(\xi^2)}. \tag{46b}$$

Proof. Relation (46a) being in the nature of linear regression, we obtain

$$\lambda = \frac{E(\xi\zeta)}{E(\xi^2)}$$

which by the use of (23) and (25) transforms to (46b).

Comment. The following theorem brings in relief that the unknown factor ζ does not coincide with its estimate ζ^* for known component ξ.

Theorem 4. *The conditional estimate ζ^* is a random variate with the following collective properties:*

$$E(\zeta^*) = \frac{E(\zeta^2)}{E(\zeta^2) + (1/p)\sigma^2(\delta)}\, E(\zeta) \tag{47}$$

$$E[(\zeta^*)^2] = \frac{E^2(\zeta^2)}{E(\zeta^2) + (1/p)\sigma^2(\delta)} \tag{48}$$

$$E[(\zeta^* - \zeta)^2] = \frac{(1/p)E(\zeta^2)\sigma^2(\delta)}{E(\zeta^2) + (1/p)\sigma^2(\delta)}. \tag{49}$$

Proof. As to (47), relation (46a) implies

$$E(\zeta^* \mid \xi) = \lambda\xi \tag{50}$$

which gives the unconditional expectation

$$E(\zeta^*) = \lambda E(\xi) \tag{51}$$

in accordance with (47). As to (48), we infer from (46a) that

$$E[(\zeta^*)^2] = \lambda^2\xi^2.$$

Hence

$$E[(\zeta^*)^2] = \lambda^2 E[\xi^2]$$

which by use of (23) and (46b) reduces to (48). As to (49), we obtain by the same argument,

$$E(\zeta^*\zeta \mid \xi) = \lambda\xi^2$$

which gives

$$E(\zeta^*\zeta) = \lambda^2 E(\xi^2).$$

Hence

$$E[(\zeta^* - \zeta)^2] = E[(\zeta^*)^2] - 2E(\zeta^*\zeta) + E(\zeta^2)$$
$$= E(\zeta^2) - \frac{E^2(\zeta^2)}{E(\zeta^2) + (1/p)\sigma^2(\delta)}$$

in accordance with (49).

Residual Estimation. The difference between factor values ζ_t and their estimates ζ_t^* carries over to the factor residuals

$$\delta_{it} = y_{it} - \beta_i \zeta_t \tag{52a}$$

and the corresponding conditional residuals

$$\delta_{it}^* = y_{it} - \beta_i \zeta_t^*. \tag{52b}$$

The unknown factor residuals δ_{it} have zero expectation, $E(\delta_{it}) = 0$, as we know from (9a). Expressing their variance in terms of component structure, we obtain by (15) and (24),

$$\sigma^2(\delta_i) = \frac{1}{p}\sigma^2(\delta) = \frac{1}{p-1}\sigma^2(\varepsilon) = \frac{1}{p-1}\sum_i \sigma^2(\varepsilon_i). \tag{53}$$

The conditional residuals δ_i^* are biased,

$$E(\delta_i^*) = \beta_i(1 - \lambda)E(\xi) \tag{54}$$

as readily verified. By deductions similar to the proof of Theorems 3 and 4, their theoretical mean square comes out as

$$E[(\delta_i^*)^2] = \beta_i^2(1 - \lambda)^2 E(\xi^2) + \sigma^2(\varepsilon_i). \tag{55}$$

2.2. Sample Estimates of One-Factor Structure. Corrections for Finite-Sample Bias

Least-squares estimates of the structure of the first principal component being given by (44), formulas (45) and (53) carry over to yield sample estimates of the one-factor structure. Writing

$$m(x^2) = \frac{1}{n}\sum_{t=1}^{n} x_t^2; \qquad m(y_i^2) = \frac{1}{n}\sum_{t=1}^{n} y_{it}^2 = m_{ii} \tag{56a}$$

$$s^2(e_i) = \frac{1}{n}\sum_{t=1}^{n} e_{it}^2; \qquad s^2(e) = \sum_{i=1}^{p} s^2(e_i) \tag{56b}$$

for the observed mean squares of the first principal component and of the

corresponding residual variances and similarly for the factor structure, we obtain

$$m(z^2) = m(x^2) - \frac{1}{p-1}\, s^2(e) \tag{57}$$

$$s^2(d_i) = \frac{1}{p}\, s^2(d) = \frac{1}{p-1}\, s^2(e) = \frac{1}{p-1} \sum_i s^2(e_i). \tag{58}$$

The corresponding estimates for the individual factor structure are

$$z_t^* = \left[1 - \frac{1}{p-1}\, \frac{s^2(e)}{m(x^2)}\right] x_t \tag{59}$$

$$d_{it}^* = y_{it} - b_i z_t^*. \tag{60}$$

Formulas (57) and (58) are in the nature of large-sample estimates, with estimation errors that tend to zero as the sample size increases, whereas (59) and (60) are subject to an estimation error that does not vanish in indefinitely large samples. In the following paragraphs we shall adduce three simple corrections for finite-sample bias in the estimation, one for each of the elements β_i, ζ_t, δ_{it} of the one-factor model under consideration.

The limited scope of these subsections must be emphasized. All through we assume that the sample is large relative to the number of variates,

$$p \prec n. \tag{61}$$

The argument is partly heuristic, and the assessment of large-sample standard errors and confidence intervals is not more than mentioned.

The Residual Variance. The one-factor model (15) involves p parameters to be estimated, namely $\beta_1, \ldots, \beta_p$, or rather $p-1$ if we take into account the normalization (4a) to a unit square sum. We see that the estimation of the individual factor values ζ_t involves no loss of degrees of freedom, since the estimates are given by (59) on the basis of the observations (1b), formula (44), and the estimated parameters b_i.

The model under estimation involves np residuals, as many as the available observations (1b). By the argument of the previous paragraph the residual variance is estimated on the basis of $np - (p-1)$ degrees of freedom. Hence, letting "est" denote estimates corrected for finite-sample bias,

$$\text{est } \sigma^2(\varepsilon) = \frac{p}{np - p + 1} \sum_i \sum_t e_{it}^2 \tag{62}$$

and in consequence,

$$\text{est } \sigma^2(\delta_i) = \frac{1}{p}\, \text{est } \sigma^2(\delta) = \frac{1}{p-1}\, \text{est } \sigma^2(\varepsilon)$$

$$= \frac{p}{(p-1)(np - p + 1)} \sum_i \sum_t e_{it}^2. \tag{63}$$

We see that the correction needed in the standard deviation $s(d)$ is of relative magnitude const./n,

$$\text{est } \sigma(\delta_i) \sim \left[1 + \frac{p-1}{2np}\right] s(d_i). \tag{64}$$

Factor and Component Corrections. To repeat from Section 1.3, the first component in (16) "explains more" than the factor in (15), or, equivalently, $\sigma(\varepsilon) < \sigma(\delta)$. This feature is somewhat exaggerated in the sample, since $s^2(e)$ comes out on the low side, owing to the loss of degrees of freedom, as we know from Section 2.2. Correcting for this bias, we obtain

$$\begin{aligned} \text{est } E(\xi^2) &= \sum_i m(y_i^2) - \text{est } \sigma^2(\varepsilon) \\ &= \sum_i m(y_i^2) - \frac{p}{np - p + 1} \sum_i \sum_t e_{it}^2. \end{aligned} \tag{65}$$

$$\begin{aligned} \text{est } E(\zeta^2) &= \sum_i m(y_i^2) - \text{est } \sigma^2(\delta) \\ &= \sum_i m(y_i^2) - \frac{p^2}{(np - p + 1)(p-1)} \sum_i \sum_t e_{it}^2. \end{aligned} \tag{66}$$

In formula (59) for the individual factor values, the finite-sample bias affects the negative term under the root sign, and we see that both the numerator and denominator make the term somewhat too small. As a first approximation we obtain

$$\text{est } \zeta_t^* = \left[1 - \frac{1}{p-1} \frac{s^2(e)}{m(x^2)} (1 + \Delta_1)\right] x_t \tag{66a}$$

with

$$\Delta_1 = \frac{p-1}{np} \frac{\sum_i m(y_i^2)}{m(x^2)}. \tag{66b}$$

The Loadings b_i. Here we encounter two sources of finite-sample bias. These will be dealt with separately, and the ensuing corrections are to be applied in conjunction.

One is the counterpart to the bias in the factor z_t. Since the factor estimates z_t come out with a mean square somewhat too large, the negative correction for bias will have to be balanced by a corresponding positive correction in the loadings b_i to maintain the least-squares approximation

$$b_i z_t \sim \beta_i \zeta_t \tag{67}$$

Denoting this partial amendment by $\text{est}^{(1)}$, we obtain

$$\text{est}^{(1)}\,\beta_i = \left[1 + \frac{1}{p-1}\frac{s^2(e)}{m(x^2)}(1+\Delta_1)\right]b_i \tag{68}$$

where Δ_1 is the same quantity as in (66).

The second source of bias in b_i is the normalization (35b) to a unit square sum. Considering h replications of the sample, let $b_i(r)$ denote the direction cosines of the first principal component as estimated from the rth replication, and write

$$\bar{b}_i = \frac{1}{h}\sum_{r=1}^{h} b_i(r)\,; \qquad D(b_i) = \left\{\frac{1}{h}\sum_{r=1}^{h}[b_i(r) - \bar{b}_i]^2\right\}^{1/2} \tag{69}$$

for the mean and standard deviation of b_i as formed on the basis of the h replications. The identity

$$\frac{1}{h}\sum_{r=1}^{h} b_i^2(r) = \bar{b}_i^2 + D^2(b_i)$$

gives

$$\begin{aligned}\sum_{i=1}^{p}\bar{b}_i^2 + \sum_{i=1}^{p} D^2(b_i) &= \frac{1}{h}\sum_{i=1}^{p}\sum_{r=1}^{h} b_i^2(r) \\ &= \frac{1}{h}\sum_{r=1}^{h}\sum_{i=1}^{p} b_i^2(r) = 1\end{aligned}$$

where the last reduction makes use of (38). Hence

$$\sum_{i=1}^{p}\bar{b}_i^2 < 1 = \sum_{i=1}^{p}\beta_i^2$$

which shows that $\bar{b}_i$ and b_i tend to be too small in absolute magnitude. To counteract this bias we assume as a rough approximation that the requisite correction is multiplicative, and independent of i. This gives, indicating the second correction by $\text{est}^{(2)}$,

$$\text{est}^{(2)}\beta_i = b_i(1+\Delta_2) \tag{70a}$$

where

$$\Delta_2 \sim \tfrac{1}{2}\sum_i D^2(b_i) + \tfrac{3}{8}\Big[\sum_i D^2(b_i)^2\Big]. \tag{70b}$$

In fact, taking the average of the corrected loadings over the replications, and forming the square sum of the averages, we obtain

$$\begin{aligned}\sum_i \bar{b}_i^2(1+\Delta_2)^2 &= (1 + 2\Delta_2 + \Delta_2^2)\sum_i \bar{b}_i^2 \\ &= (1 + 2\Delta_2 + \Delta_2^2)\Big[1 - \sum_i D^2(b_i)\Big] \\ &= 1 + c\,\Delta_2^3 + \cdots \sim \sum_i \beta_i^2\end{aligned} \tag{71}$$

which shows that the correction (70) takes into account first and second powers of $\sum D^2(b_i)$.

Comment. According to (70b) the second correction for the loadings estimates is of the same order of magnitude as the sum of squares of the standard deviations $d(b_i)$ of the loadings estimates. Hence for the correction to work, the sample size n must be so large that this square sum is fairly small.

The assessment of standard deviations $D(b_i)$ falls outside the scope of this paper. In connection with the correction (70) we quote the following large-sample formula,

$$D(b_i) \sim \frac{\sigma(\varepsilon_i)}{[E(\xi^2)]^{1/2}\left[1 - \dfrac{\sigma^2(\varepsilon_i)}{E(\xi^2)(1-\beta_i^2)}\right]} \tag{72}$$

with reference to E. Lyttkens,[3] who has recently obtained the formula by a new argument, under assumptions that involve an extension of related earlier results.

2.3. Illustrations by Monte Carlo Experiments

We shall now in all brevity report some experiments that have been carried out to check and illustrate the NILES procedure (Section 1.4) for the estimation of component structure and its adaptation (Section 2.1) for the estimation of one-factor structure.

In the terminology of experimental psychology, our Monte Carlo experiments simulate the observations of 4 ability tests as applied to 50 persons. This makes 200 observations in each sample. Our experiments involve 100 replications of the sample, once with fixed and once with changing factor values. This makes in all $200 \times 200 = 40{,}000$ observations. Further we shall report other similar experiments, once with 25% of the data missing, and once with 50%. All in all, the Monte Carlo study involves 120,000 simulated observations. The numerical work was performed in a pilot round at Battelle Memorial Institute, Columbus, Ohio, and in the production phase at the University of Uppsala, Sweden, using IBM-1620 and CDC 3600 computers.[4]

The One-Factor Structure Taken to Generate the Data of the Experiments

I. *Number of variates in the data set* (1)

$$p = 4 \tag{73}$$

[3] See [19, Chap. 1].

[4] The programs have been written and run by Fil. Kand. S. Wold, Dept. of Chemistry, Univ. of Uppsala, and Fil. Kand. D. Jonsson and Kand. L. Bodin, Dept. of Statistics, Univ. of Uppsala, Sweden.

II. *Model.* The factor model (3) with one factor and equal residual variances,

$$k = 1\,; \qquad \sigma(\delta_i) = \frac{1}{p^{1/2}}\,\sigma(\delta) = \frac{1}{2}\,\sigma(\delta) = \text{const.} \tag{74}$$

III. *Sample size n and number of replications h*

$$n = 50\,; \qquad h = 100. \tag{75}$$

IV. *Loadings*

$$\beta_1 = -\frac{2}{10^{1/2}}\,; \quad \beta_2 = -\frac{1}{10^{1/2}}\,; \qquad \beta_3 = \frac{1}{10^{1/2}}\,; \qquad \beta_4 = \frac{2}{10^{1/2}}\,. \tag{76}$$

V. *Factor values.* In each sample

$$\zeta_1, \ldots, \zeta_{50} \quad \text{are} \quad \text{n.i.d.} \quad (0,1). \tag{77}$$

That is, the individual factor values ζ_t are generated so as to be normally distributed and mutually independent with zero expectation and unit standard deviation.

VI. *Residuals.* In each sample

$$\delta_{i1}, \ldots, \delta_{i,50} \quad \text{are} \quad \text{n.i.d.} \quad (0,1). \tag{78}$$

Since $p = 4$, each sample involves 200 independent residuals δ_{it}.

VII. *Replications with fixed vs. changing factor values.* The experiments have been carried through in two versions:

A. Keeping the factor values (77) fixed in the 100 replications.

B. Generating the factor values (77) anew for each replication.

Having now specified the one-factor model (15) that generates the data of our experiments, we note the following properties of the model.

Product moment matrix (12a) of the variates η_i:

$$[\mu_{ij}] = [E(\eta_{it}\eta_{jt})] = \begin{bmatrix} 1.4 & 0.2 & -0.2 & -0.4 \\ 0.2 & 1.1 & -0.1 & -0.2 \\ -0.2 & -0.1 & 1.1 & 0.2 \\ -0.4 & -0.2 & 0.2 & 1.4 \end{bmatrix}$$

in accordance with (76) to (78).

Standard deviation of the total factor residual is

$$\sigma(\delta) = 2\sigma(\delta_i) = 2 \tag{79}$$

in accordance with (73) and (74).

Comments. (a) Small residuals make small differences between the factor

and component models, (15) and (16). To bring the differences in relief we have used substantial residuals; specifications (77) and (78) give

$$\sigma(\delta_i) = \sigma(\zeta_t) = [E(\zeta_t^2)]^{1/2}$$

which makes larger residuals relative to the factors than is usually met in practice.

(b) Versions A and B of specification VII are analogous to the two standard situations in regression analysis known as fixed vs. random explanatory variates [2]. Just as for many problems in regression analysis, we shall see that in the estimation of component and factor structure the difference between the two versions makes little or no difference in the first approximation.

The Experiments Reinterpreted in Terms of Component Structure. Reinterpreting the data of our experiments as generated by a component model (2) with $k > 1$, Theorem 1 allows us to express the first principal component and its constituent relation (18) in terms of the one-factor structure specified by (73) to (78). The following results are obtained.

(a) *Direction cosines β_i of the first principal component.* Same as the loadings (76). This is in accordance with (18a).

(b) *Individual values for the first component*

$$\xi_t = (1/10^{1/2})(-2y_{1t} - y_{2t} + y_{3t} + 2y_{4t}) \tag{80}$$

in accordance with (18).

(c) *Component residuals*

$$\begin{aligned}
\varepsilon_{1t} &= \tfrac{1}{10}(6y_{1t} - 2y_{2t} + 2y_{3t} + 4y_{4t}) \\
\varepsilon_{2t} &= \tfrac{1}{10}(-2y_{1t} + 9y_{2t} + y_{3t} + 2y_{4t}) \\
\varepsilon_{3t} &= \tfrac{1}{10}(2y_{1t} + y_{2t} + 9y_{3t} - 2y_{4t}) \\
\varepsilon_{4t} &= \tfrac{1}{10}(4y_{1t} + 2y_{2t} - 2y_{3t} + 6y_{4t})
\end{aligned}$$

in accordance with (19).

(d) *Mean square of the first component*

$$E(\xi^2) = 2E(\zeta^2) = 2 \tag{81}$$

in accordance with (23) and (74).

(e) *Residual variances*

$$\sigma^2(\varepsilon_{1t}) = 0.6\,; \qquad \sigma^2(\varepsilon_{2t}) = \sigma^2(\varepsilon_{3t}) = 0.9\,; \qquad \sigma^2(\varepsilon_{3t}) = 0.6 \tag{82}$$

in accordance with (c). Hence the total residual variance

$$\sigma^2(\varepsilon) = 3 \tag{83}$$

in accordance with (20) and (24).

Comments. The features under (a) to (e) suffice to determine the transformation from component to factor structure. The following points provide further details about component vs. factor structure.

(f) *The component property*

$$\beta_1\varepsilon_{1t} + \beta_2\varepsilon_{2t} + \beta_3\varepsilon_{3t} + \beta_4\varepsilon_{4t} = 0 \tag{84}$$

in accordance with (18), and as readily verified from (76) and (c) above.

(g) *The least-squares property* (8)

$$E(\varepsilon_{it}) = E(\xi_t\varepsilon_{it}) = 0 \tag{85}$$

in accordance with (8), and readily verified by the use of (b) and (c). For example, transforming $E(\xi_t\varepsilon_{1t})$ by expressing the expectation in terms of $E(\zeta^2)$ and $E(\delta_i{}^2)$, the terms that involve $E(\zeta^2)$ cancel out if we make repeated use of the relation $6\beta_1 - 2\beta_2 + 2\beta_3 + 4\beta_4 = 0$, while the terms that involve $E(\delta_i{}^2)$ cancel out, since $2 \times 6 - 1 \times 2 - 1 \times 2 - 2 \times 4 = 0$.

(h) *The factor property.* Thanks to the specification (78), the factor residuals δ_{it} satisfy the factor property (11). The component residuals ε_{it}, on the other hand, do not possess this property; a case in point is

$$E(\varepsilon_{1t}\varepsilon_{2t}) = -0.2 \neq 0 \tag{86}$$

as is readily verified from (c) above.

Monte Carlo Results. With reference to Table I, we shall now summarize the numerical results of our experiments. The loadings (= direction cosines) β_i being the same for factor and component models, (15) and (16), under analysis, part (1) of the table reports about the estimated loadings. Then in parts (2) and (3) we proceed to the estimates of residuals, components, and factors. Dealing in (2) and (3) with population and sampling aspects of component vs. factor structures, we return finally to the sampling aspects of the estimates b_i reported in part (1). All through we limit the numerical illustrations to collective features of the factor and component structures.

I(1) *The loadings estimates.* For the first of the 100 replications of experiment A the NILES estimates for the four loadings are

$$b_1 = -0.793\,; \quad b_2 = -0.209\,; \quad b_3 = 0.262\,; \quad b_4 = 0.508. \tag{87}$$

The deviations from the theoretical loadings (33) are

$$\Delta_1 = -0.161\,; \quad \Delta_2 = 0.107\,; \quad \Delta_3 = -0.054\,; \quad \Delta_4 = -0.124. \tag{88}$$

Next we turn to averages and standard deviations calculated in accordance with (69). Table I(1) shows the four average estimates

$$\bar{b}_i = \tfrac{1}{100}\sum b_i \tag{89}$$

TABLE I

The One-Factor Model (15) *with* $n = 50$, $p = 4$

I(1)[a,b]			A		B		
	i	β_i	b_i	$D(b_i)$	$\bar{b}_i$	$D(b_i)$	(72)
	1	−0.632	−0.573	0.182	−0.572	0.174	0.155
	2	−0.316	−0.257	0.258	−0.312	0.185	0.187
	3	0.316	0.331	0.199	0.304	0.189	0.187
	4	0.632	0.555	0.221	0.570	0.239	0.155

I(2)[a,c]	A				B			
Repl.	$m(e^2)$	$m(d^2)$	est $m(\delta^2)$	$m(\delta^2)$	$m(e^2)$	$m(d^2)$	est $m(\delta^2)$	$m(\delta^2)$
1	0.612	0.815	0.828	0.836	0.612	0.815	0.828	0.836
2	0.752	1.003	1.018	1.009	0.654	0.874	0.886	0.923
3	0.617	0.823	0.836	0.814	0.661	0.882	0.895	0.896
4	0.699	0.933	0.947	0.991	0.830	1.107	1.124	1.213
5	0.857	1.143	1.160	1.204	0.721	0.961	0.975	0.980
Aver.	0.714	0.952	0.967	1.003	0.716	0.955	0.969	0.999

I(3)[a,d]	A				B			
Repl.	$m(x^2)$	$m(z^2)$	est $m(\zeta^2)$	$m(\zeta^2)$	$m(x^2)$	$m(z^2)$	est $m(\zeta^2)$	$m(\zeta^2)$
1	1.534	0.719	0.706	0.903	1.534	0.719	0.706	0.903
2	2.020	1.017	1.002	0.903	2.147	1.274	1.261	1.047
3	1.788	0.965	0.952	0.903	1.529	0.647	0.634	0.794
4	2.284	1.352	1.338	0.903	2.480	1.374	1.357	1.130
5	2.181	1.038	1.021	0.903	2.178	1.217	1.202	0.972
Aver.	2.114	1.161	1.147	0.903	2.129	1.147	1.160	1.033

[a] Two Monte Carlo experiments with 100 replications: A, fixed factor values; B, shifting factor values.

[b] Loadings: theoretical, β_i; average of estimates b_i in 100 replications, $\bar{b}_i$. Standard errors of loadings estimates b_i: assessed from 100 replications, $D(b_i)$; theoretical, by (72).

[c] Observed mean squares of total component residuals, $m(e^2)$. Mean squares of total factor residuals: observed, $m(d^2)$; corrected for finite sample bias, est $m(\delta^2)$; actual, $m(\delta^2)$. Items for five replications, and averages for 100 replications.

[d] Observed mean squares of first principal component, $m(x^2)$. Mean squares of factor ζ: observed, $m(z^2)$; corrected for finite-sample bias, est $m(\zeta^2)$; actual, $m(\zeta^2)$. Items for five replications, and averages for 100 replications.

as based on the 100 replications. We see that the averages compare fairly well with the theoretical loadings β_i shown in the second column. In accordance with a previous remark, the results are quite similar for the series A and B of the experiment. On closer comparison with the theoretical loadings β_i, we see that the large loadings b_i and b_4 come out somewhat too small in absolute value. We shall return to this feature later.

Table I(1) further gives the standard errors $D(b_i)$ of the loadings estimates b_i, observed values calculated from

$$D(b_i) = [\tfrac{1}{100} \sum (b_i - \bar{b}_i)^2]^{1/2} \tag{90}$$

and theoretical values given by (72), which covers both cases A and B, and in the present case gives

$$D(b_1) = D(b_4) = (0.024)^{1/2} = 0.155; \qquad D(b_2) = D(b_3) = (0.036)^{1/2} = 0.187$$

The standard errors are about 0.15 to 0.23, well covering the deviations (88) that we have quoted for the first replication under A.

I(2) *Residual variance.* Writing

$$m(\delta^2) = \tfrac{1}{200} \sum_i \sum_t \delta_{it}^2 \tag{91}$$

for the mean square of the 200 residuals generated for each replication of the experiment, Table I(2) gives $m(\delta^2)$ for the first five replications in each version A and B, and its average over all 100 replications. Indicating its variability by the formula

$$\text{Expectation} \pm \text{standard deviation; that is, } E(\delta^2) \pm \frac{\sigma(\delta^2)}{(np)^{1/2}}$$

the specification that δ_{it} is n.i.d. (0.1) gives

$$m(\delta^2) = 1 \pm 0.1; \qquad m(\delta^i) = 1 \pm 0.01 \tag{92}$$

for each replication and, respectively, for the over-all average. As seen from column $m(\delta^2)$ in Table I(2), the factor residuals generated in the experiments are in accordance with this theoretical variability.

Column $m(e^2)$ refers to the total component residuals e_t as obtained by the NILES estimation procedure (Section 1.4). Since $\bar{e}_t = 0$, the mean square $m(e^2)$ can be interpreted as the observed variance $s^2(e)$. Hence $m(e^2) = s^2(e)$ may be interpreted as an estimate of either $\sigma^2(\varepsilon)$ or $m(\varepsilon^2)$. By (24), $\sigma^2(\varepsilon) = (p-1)/p = 0.750$. The observed estimates are in accordance with this theoretical value, as seen from the columns $m(e^2)$ under A and B.

Column $m(d^2)$ refers to the total factor residuals d_t, and is calculated from $ps^2(e)/(p-1)$, in accordance with (58). Just as for the component residuals e_t, we have $m(d^2) = s^2(d)$, and $m(d^2)$ may be interpreted as an estimate either of

$E(\delta^2)(=1)$ or $m(\delta^2)$. The accordance between theoretical and observed values is necessarily the same as for column $m(e^2)$.

I(3) *Component and factor estimates.* Column $m(\zeta^2)$ in Table I(3) gives the mean square

$$m(\zeta^2) = \tfrac{1}{50} \sum_t \zeta_t^2 \tag{93}$$

of the 50 factor values ζ_t generated for the first five replications of experiments A and B, and its average over all 100 replications. Here the variability is twice as large as in (92),

$$m(\zeta^2) = 1 \pm 0.2$$

and the numerical results are seen to be in agreement with this theoretical formula.

Column $m(x^2)$ gives the mean square of the first principal component as assessed by the NILES procedure (Section 1.4). Interpreting $m(x^2)$ as an estimate of $E(\xi^2) = 2$ as given by (81), we see that the observed $m(x^2)$ are in fair agreement with expectation.

Column $m(z^2)$ gives the mean square of the factor ζ as assessed by (57). Interpreting $m(z^2)$ as an estimate either of $m(\zeta^2)$ as defined by (93) or, more crudely, as an estimate of $E(\zeta^2) = 1$, we see that the observed $m(z^2)$ are in fair agreement with the theoretical values.

Finite-sample corrections. The numerical results we have taken up thus far are based on Theorems 1 and 2, and thus refer to the population aspects of the transfer from component structure to factor structure. We shall now turn to the corrections for finite-sample bias as calculated from the formulas given in Section 2.2.

As to the residual variances, we see from (62) and (63) that the corrections in $m(e^2)$ and $m(d^2)$ are given by the multiplicative factor

$$\frac{np}{np - p + 1} = \frac{200}{197} \sim 1.015.$$

Table I(2) gives the corrected estimate for $m(\delta^2)$, denoted by est $m(\delta^2)$. By 1.5% higher than the uncorrected estimate $m(d^2)$, we see that the correction, although quite small, goes in the right direction.

As to the mean squares of factors and components, the finite-sample correction as given by (65) and (66) is 1.5% of the corresponding residual variance. The corrected estimate for the factor mean square $m(\zeta^2)$ is given in Table I(3) under est $m(\delta^2)$. Again we see that the correction is quite small, and that it works in the appropriate direction.

Finally we turn to the two corrections (68) and (70) for the loadings estimates b_i. Both corrections Δ_1 and Δ_2 are positive, and thus have the

appropriate sign, as seen from Table I(1). In the present case Δ_1 is quite small, the ensuing correction amounting to about 1% of b_i. The correction by Δ_2 is more important, and if we evaluate $D^2(b_i)$ by the empirical variances reported in Table I(1), we find that it amounts to about 9% of b_i. In all, the two corrections thus increase b_i by 10%. We see from Table I(1) that this correction brings up b_i and $\bar{b}_i$ to the theoretical value β_i.

Standard errors of the loadings estimates. As seen from Table I(1), the standard errors $d(b_i)$ of the loadings estimates are on the whole somewhat larger for case A than for case B. Lyttkens' formula (72) gives more close agreement in case B than in case A, especially so for the standard errors of b_2 and b_3.

Principal-Components Estimation When Some of the Data are Missing. We shall here report an adaptation of the NILES procedure (Section 1.4) to situations when the observational material is incomplete. The device, due to Christoffersson [19, Chap. 4], is to supplement the missing data by dummy observations y_{it}^* such that when the estimation procedure has been applied to the data thus supplemented, the dummy observations y_{it}^* coincide with their principal-components representation

$$y_{it}^* = \sum_a b_{ia} x_{at} \tag{94}$$

or, otherwise expressed, giving zero residuals for the dummy observations. In the case of one component, $k = 1$, the device requires no change in the NILES procedure other than to replace all regressions in (34) and (35a) by weighted regressions, making the weight $w = 1$ for the existing observations, and $w = 0$ for the missing ones. In the case of two or more components, the computational device is somewhat more complicated.

TABLE II[a]

The One-Factor Model with 25% or 50% of the Data Missing

		25% missing				50% missing			
		A		B		A		B	
i	β_i	$\bar{b}_i$	$D(b_i)$	$\bar{b}_i$	$D(b_i)$	$\bar{b}_i$	$D(b_i)$	$\bar{b}_i$	$D(b_i)$
1	−0.632	−0.590	0.151	−0.571	0.182	−0.593	0.189	−0.565	0.182
2	−0.316	−0.294	0.223	−0.297	0.271	−0.314	0.214	−0.312	0.184
3	0.316	0.296	0.209	0.255	0.225	0.247	0.210	0.304	0.190
4	0.632	0.581	0.154	0.569	0.198	0.543	0.261	0.574	0.243

[a] Same model and table code as in Table I.

A Monte Carlo experiment with the dummy device is reported in Table II. The data are generated in the same way as in Table I, except that either one of four or one of two observations is missing. The missing observations are arranged in a systematic pattern; for example, with 25% missing data one observation is excluded for each t, the observation y_{it} with $i \equiv t \pmod 4$.

Comments. (a) Comparing with Table I(1) we see that the omission of as much as 25% or even 50% of the data does not make much difference in the estimation of the loadings. For both versions A and B, the average estimates $\bar{b}_i$ show fair agreement with the theoretical loadings β_i. There is the same tendency as in Table I(1) to a downward bias in the numerical value of the numerically largest loadings b_1 and b_4. What is perhaps more surprising, the observed standard deviations $D(b_i)$ come out with about the same size as in the case of complete observations, Table I(1), maybe with some very slight increase with the percentage of missing data. The requisite modification of the theoretical formula (72) will not be taken up here.

To conclude, our Monte Carlo experiments suggest that the sampling properties of the loadings estimates b_i are influenced primarily by the size of n, not so much by p or by the percentage of missing data for each t.

(b) In applied work there are many situations where principal-components analysis has to cope with incomplete data. For one thing, this is so because there are gaps even in very complete data. Sometimes, and this is of great importance for the scope of the approach, it lies in the very nature of the situation under analysis that the available information is incomplete. A case in point is horse-race data [19]. Letting y_{it} stand for the recorded time from start to goal for horse t in race i, complete data would here mean that all horses under comparison participate in each race; in reality, only some 10 horses are set up at the same start. Another case in point is the incompleteness that occurs in controlled experiments in which for some reason or other the number of replications is limited.

The dummy assumption is just one, and perhaps the simplest, way to handle incomplete data. Speaking generally, approach (94) is *neutral* in the sense that missing data are assumed to occur just by chance, without affinity to any of the components or loadings.

3. EXTENSIONS OF THE PRINCIPAL-COMPONENTS APPROACH

To illustrate the flexibility of the principal-components model (2) and its estimation by the NILES procedure, we shall in all brevity refer to three generalizations of the approach; cf. [19, Chap. 3 and 4].

3.1. Hybrid Model of Principal Components and Canonical Correlation

In the case of one component the model has the form

$$\alpha_1 y_{it} + \alpha_2 u_{it} = \beta_i \xi_t + \varepsilon_{it} \tag{95}$$

with

$$\alpha_1^2 + \alpha_2^2 = 1\,; \qquad \sum_i \beta_i^2 = 1. \tag{96}$$

As in Section 2.3 we consider the application to horse-race data, assume that the primary purpose of the analysis is to assess the racing quality of the horses, and adopt the device to rank them according to their component values ξ_t. In Section 2.3 the ranking exploits the information embodied in the observations y_{it} on racing times; in model (95) the ranking makes joint use of observations on y_{it} and some other variate u_{it}, for example, the saddle weight that horse t carries as a handicap in race i.

A NILES estimation procedure is available for the joint estimation of the parameters α_1, α_2, γ_i, and ξ_t (see [19, Chap. 3]). The model and the estimation procedure extend (a) to any number of left-hand variates y, u, ..., and (b) to two or more components ξ_{at} in the right-hand member.

3.2. Hybrid Model of Principal Components and Multiple Regression

The model is formally related to (95):

$$y_{it} = \beta_i \xi_t + \gamma_i u_{it} + \varepsilon_{it} \tag{97}$$

with

$$\sum_i \beta_i^2 = 1. \tag{98}$$

Again considering the model as applied to horse-race data, we assume that the purpose of the analysis is now to forecast the racing time y_{it} of horse t in race i, using the horse component ξ_t, the race loading γ_i, and the saddle weight u_{it} as predictive elements.

The NILES procedure for model (95) [Ref. 19, Chap. 3] can readily be adapted for the estimation of model (96). The approach extends (a) to two or more observational variates u, v, ... in the right-hand member, and (b) to two or more principal components ξ_{at}.

3.3. Principal Components for a Multidimensional Array of Observations

This generalization is in a direction other than (95) or (96), the model being of the type

$$y_{ijt} = \beta_i \gamma_j \xi_t + \varepsilon_{ijt}$$

$$\sum \beta_i^2 = 1\,; \qquad \sum \gamma_j^2 = 1 \qquad \begin{cases} i = 1, \ldots, p \\ j = 1, \ldots, q \\ t = 1, \ldots, n. \end{cases} \tag{99}$$

The model can be consistently estimated by the following NILES estimation procedure [20].

Initial values $b^{(1)}$ and $g_j^{(1)}$ are chosen arbitrarily, say

$$b_i^{(1)} = u_1 y_{iAB}, \qquad g_j^{(1)} = v_1 y_{CjB}$$

where A, B, and C are fixed subscripts, and the multiplicative factors u_1 and v_1 are determined so as to make

$$\sum (b_i^{(1)})^2 = \sum (g_j^{(1)})^2 = 1.$$

When the proxies $b_i^{(s)}$ and $g_j^{(s)}$ have been calculated, the procedure continues as follows:

$$x_t^{(s+1)} = \sum_i \sum_j b_i^{(s)} g_j^{(s)} y_{ijt}\,; \qquad t = 1, \ldots, n$$

$$b_i^{(s+1)} = u_{s+1} \sum_j \sum_t g_j^{(s+1)} x_t^{(s+1)} y_{ijt} \Big/ \sum_j (g_j^{(s+1)})^2 \sum_t (x_t^{(s+1)})^2\,; \qquad i = 1, \ldots, p$$

$$g_j^{(s+1)} = v_{s+1} \sum_i \sum_t b^{(s+1)} x_t^{(s+1)} y_{ijt} \Big/ \sum_i (b_i^{(s+1)})^2 \sum_t (x_t^{(s+1)})^2\,; \qquad j = 1, \ldots, q$$

where the multiplicative factors u_{s+1} and v_{s+1} are to be determined so as to make

$$\sum_i (b_i^{(s+1)})^2 = \sum_j (g_j^{(s+1)})^2 = 1.$$

The procedure gives

$$b_i = \lim b_i^{(s)}\,; \quad g_j = \lim g_j^{(s)}\,; \quad x_t = \lim x_t^{(s)} \qquad (s \to \infty)$$

as NILES estimates for the parameters β_i, γ_j, and ξ_t.

ACKNOWLEDGMENTS

The author is greatly indebted to Professor P. Whittle, University of Manchester, and Professors K. G. Jöreskog and E. Lyttkens, University of Uppsala, for having read and commented upon the paper in draft form.

REFERENCES

1. ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
2. CRAMÉR (1945). *Mathematical Methods of Statistics*. Almqvist and Wiksell, Uppsala. (1946), Princeton Univ. Press, Princeton, New Jersey.

3. HOTELLING, H. (1933). Analysis of a complex of variables into principal components. *J. Educational Psychology* **24** 417–441, 498–520.
4. JÖRESKOG, K. G. (1962). On the statistical treatment of residuals in factor analysis. *Psychometrika* **27** 335–354.
5. JÖRESKOG, K. G. (1963). *Statistical Estimation in Factor Analysis*. Almqvist and Wiksell, Uppsala.
6. LAWLEY, D. N. (1963). *Factor Analysis as a Statistical Method*. Butterworth, London.
7. LYTTKENS, E. (1966). On the fix-point property of Wold's iterative estimation method for principal components. Paper in this volume, pp. 335–350.
8. PEARSON, K. (1901). On lines and planes of closest fit to systems of points in space. *Phil. Mag.* **6** 559–572.
9. RAO, C. R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika* **20** 93–111.
10. SPEARMAN, C. (1904). General intelligence, objectively determined and measured. *Amer. J. Psychology* **15** 201–293.
11. THURSTONE, L. L. (1931). Multiple factor analysis. *Psychological Review* **38** 406–427.
12. WHITTLE, P. (1952). On principal components and least square methods of factor analysis. *Skand. Aktuarietidskr.* **36** 223–239.
13. WOLD, H. (1953). A Monte Carlo illustration of principal components analysis. *Uppsala Symposium on Psychological Factor Analysis*, 17–19 *March* 1953. Nordisk Psykologi's Monograph Series, No. 3, pp. 43–64. Munksgaard, Copenhagen.
14. WOLD, H. (1963). On the consistency of least squares regression. *Sankhya* **25A** 211–215.
15. WOLD, H. (1964). Forecasting by the chain principle, *Econometric Model Building. Essays on the Causal Chain Approach*, pp. 5–36. North-Holland, Amsterdam.
16. WOLD, H. (1965). A fix-point theorem with econometric background, I–II. *Ark. Mat.* **6** (11–12) 209–240.
17. WOLD, H. (1965). Toward a verdict on simultaneous equation systems. *Transactions of the Study Week on Econometrics*, *October* 1963. *Scripta Varia* **28**; I, 115–185. Pontifical Academy of Sciences, Vatican City.
18. WOLD, H. (1966). The approach of model building. Cross-roads of probability theory, statistics, and theory of knowledge. *Transactions des Entretiens de Monaco* 1964. Centre International d'Étude des Problèmes Humains, Monaco.
19. WOLD, H. (1966). Nonlinear estimation by iterative least squares procedures. *Festschrift Jerzy Neyman*. Wiley, New York.
20. WOLD, S. (1965). Principal components analysis of multidimensional arrays of observations. *Seminar communication*. Univ. Institute of Statistics, Uppsala, Sweden.
21. YOUNG, G. (1941). Maximum likelihood estimation and factor analysis. *Psychometrika* **6** 49–53.