COMPONENT ANALYSIS IN CROSS-SECTIONAL AND LONGITUDINAL DATA

ROGER E. MILLSAP
BARUCH COLLEGE
CITY UNIVERSITY OF NEW YORK

WILLIAM MEREDITH
UNIVERSITY OF CALIFORNIA-BERKELEY

An extension of component analysis to longitudinal or cross-sectional data is presented. In this method, components are derived under the restriction of invariant and/or stationary compositing weights. Optimal compositing weights are found numerically. The method can be generalized to allow differential weighting of the observed variables in deriving the component solution. Some choices of weightings are discussed. An illustration of the method using real data is presented.

Key words: invariance, stationarity, linear regression, three-mode factor analysis.

Introduction

Meredith and Millsap (1985) utilize an approach to component analysis, which can be traced back to Pearson (1901), that emphasizes the regression of observed variables on component scores. Let $X$ be an $n$-dimensional random variable with elements $x_j, j = 1, \ldots, n$. Without loss of generality, assume that $E(X) = 0$ and $E(XX') = \Sigma$, with rank ($\Sigma$) = $n$, $\Sigma$ being the mathematical expectation operator. Define $W$ as an $n \times m$ ($m < n$) matrix of compositing weights and further define an $m$-dimensional component score random variable $Z$ by the relation

$$Z = W^*X.$$  \hspace{1cm} (1)

The component scores $Z$ should provide an adequate representation of the observed variables $X$. A criterion for "adequacy" uses the residuals about the linear regression of the observed variables on the component scores. Let $\hat{X}$ be a prediction for $X$ in this regression, that is

$$\hat{X} = PZ.$$  \hspace{1cm} (2)

with $P$ an $n \times m$ matrix of regression weights, that is, a component pattern matrix. Define $E = X - \hat{X}$ as a vector of error random variables with elements $e_j, j = 1, \ldots, n$. We
choose \( P \), as a function of \( W \), to minimize the expected squared-error loss or risk function

\[
R(W; \Sigma) = e\left( \sum_j e_j^2 \right).
\]

By this criterion, the best linear prediction, \( \hat{X} \), provided by \( Z \) for a fixed \( W \) is

\[
\hat{X} = E(XZ)[E(ZZ')]^{-1}Z = \Sigma W(W\Sigma W)^{-1}W^*X.
\]

The risk function can be written

\[
R(W; \Sigma) = e[\text{tr} (EE')] = \text{tr} \left[ \Sigma - \Sigma W(W\Sigma W)^{-1}W^*\Sigma \right],
\]

where \( \text{tr} (\cdot) \) is the trace operator. The risk is a function of \( W \), and we can choose \( W \) to minimize \( R(W; \Sigma) \), or to maximize

\[
F(W; \Sigma) = \text{tr} \left[ W^*\Sigma^2 W(W\Sigma W)^{-1} \right].
\]

Let the spectral representation of \( \Sigma \) be \( \Sigma = QAQ' \), with \( Q \) orthonormal and \( \Lambda \) diagonal, \( \lambda_1 > \lambda_2 > \cdots > \lambda_n > 0 \). For \( m < n \), \( F(W; \Sigma) \) is maximized by \( W = Q_m T \), where \( Q_m \) is an \( n \times m \) matrix whose columns are the \( m \) eigenvectors corresponding to the \( m \) largest eigenvalues in \( \Sigma \), and \( T \) is an \( m \times m \) nonsingular matrix (Rao, 1973). For simplicity of presentation, we assume no equal eigenvalues.

The risk in (5) can be generalized to allow differential weighting of the elements of \( E \). Let \( G \) be an \( n \times n \) symmetric positive definite matrix of weights, with elements \( g_{i,j} \), \( i, j = 1, \ldots, n \). A weighted risk function is

\[
R(W; G, \Sigma) = e\left( \sum_j \sum_k e_j e_k g_{i,j} \right) = e[\text{tr} (GE)]
\]

\[
= \text{tr} \left[ G\Sigma - G\Sigma W(W\Sigma W)^{-1}W^*\Sigma \right].
\]

Let \( G = AA' \) be a factorization of \( G \). The matrix \( W \) which minimizes (7) can be shown to be \( AQ_m T \), with \( T \) an \( m \times m \) nonsingular matrix and \( Q_m \) the \( n \times m \) matrix whose columns are the \( m \) eigenvectors corresponding to the \( m \) largest eigenvalues of \( \Gamma = A\Sigma A \).

Multiple Occasion and Multiple Group Analyses

Multiple occasion or multiple group component analysis can be formulated in a way that is analogous to the regression formulation in the single group/occasion case. Let \( X_{ik} \) be an \( n \)-dimensional random variable measured in the \( k \)-th group on the \( i \)-th occasion, \( k = 1, \ldots, r, i = 1, \ldots, p_{ik} \). To simplify notation, in the following development we drop the subscript \( k \) on \( p \), but note that our procedures do not require an equal number of measurement occasions per group. Without loss of generality, let \( \mu(Y_{ik}) = 0 \) for all \( i, k \), and let \( \Sigma(Y_{ik}, Z_{jk}) = \Sigma_{i,j,k} \) be the covariance matrix between \( X_{ik} \) and \( X_{jk} \) in the \( k \)-th group. We will assume that rank \( (\Sigma_{i,j,k}) = n, k = 1, \ldots, r \).

The procedures we will develop are “model free”. No structural or distributional assumptions need be made in deriving these methods except the assumption of the existence of first and second moments of the random variables involved. However, we wish to emphasize the importance of developing models of change over time. For example, powerful models of the Markov sort could be very illuminating in a longitudinal context. In contrast, our purposes are descriptive, exploratory, and data analytic in orientation.

We define a \( m \times 1 \) random vector of component scores for the \( k \)-th group on the
$Z_{ik} = W^*X_{ik}$

$W$ is an $n \times m$ matrix of compositing weights, fixed to be constant over groups and occasions. Two important concepts underlying the above are the notions of "invariance" and "stationarity": that something should remain fixed over groups and occasions, respectively. Our choice for the stationary and invariant feature is the matrix of compositing weights used to derive the component scores. Levin (1966) and ten Berge (1986) also discuss this choice. The component scores are allowed to vary somewhat freely over groups and occasions, and are defined "internally" as linear functions of the observed variables within a given group and occasion.

As in the single occasion case, the component scores should provide an adequate representation of the observed variables within occasions and groups. Let a predicted value for $X_{ik}$ determined by its regression on $Z_{ik}$ be defined as

$\hat{X}_{ik} = P_{ik}Z_{ik}$

The $n \times m$ component pattern matrix $P_{ik}$ contains the regression weights for predicting $X_{ik}$ from $Z_{ik}$. The pattern matrix will ordinarily vary over groups and occasions, even though $W$ does not. Let $E_{ik} = X_{ik} - \hat{X}_{ik}$ be the vector of error random variables in the approximation of $X_{ik}$ by $\hat{X}_{ik}$. For a fixed $W$, the pattern matrix $P_{ik}$ can be chosen to minimize the expected squared-error loss or risk function

$R_{ik}(W) = \epsilon(E_{ik}', E_{ik})$

$= \epsilon[(X_{ik} - \hat{X}_{ik})(X_{ik} - \hat{X}_{ik})]$.  

(10)

The above risk is minimized when

$P_{ik} = \Sigma_{ik}W(W^*\Sigma_{ik}W)^{-1}$

(11)

Taken together, these pattern matrices minimize the summed risk as a function of $W$, namely

$R(W) = \sum_i \sum_k R_{ik}(W)$.  

(12)

With substitutions from (9), (10), and (11), the total risk is

$R(W) = \sum_i \sum_k \epsilon[\text{tr} [(X_{ik} - \hat{X}_{ik})(X_{ik} - \hat{X}_{ik})']]$

$= \sum_i \sum_k \text{tr} [\Sigma_{ik} - \Sigma_{ik}W(W^*\Sigma_{ik}W)^{-1}W^*\Sigma_{ik}W]$.  

(13)

We can choose $W$ to minimize this risk, or to maximize

$F(W) = \sum_i \sum_k \text{tr} [W^*\Sigma_{ik}^2W(W^*\Sigma_{ik}W)^{-1}W^*\Sigma_{ik}W]$.  

(14)

An optimal solution for $W$ can be found numerically, as discussed below.

The risk function in (10) can be generalized to allow differential weighting of the elements of $E_{ik}$. Let $G_{ik}$ be an $n \times n$ symmetric positive definite matrix of weights to be applied in the $k$-th group at the $i$-th occasion. The weighted risk function is

$R_{ik}(W; G_{ik}) = \epsilon(E_{ik}', G_{ik}E_{ik}) = \epsilon[\text{tr} (G_{ik}E_{ik}E_{ik}')]$

$= \text{tr} [G_{ik}\Sigma_{ik} - G_{ik}\Sigma_{ik}W(W^*\Sigma_{ik}W)^{-1}W^*\Sigma_{ik}W]$.  

(15)
A total risk function $R(W; G)$ is obtained through summation over $i$ and $k$ in (15), and gives the following criterion to be maximized in choosing $W$

$$F(W; G) = \sum_i \sum_k \text{tr} \left[ W \Sigma_{ik} G_{ik} \Sigma_{ik} W(W \Sigma_{ik} W)^{-1} \right].$$  \hspace{1cm} (16)

The criterion in (16) considers only within-occasion covariance information. The component method presented here can be extended to consider between-occasion covariances as well (Millsap, 1986), but we will not discuss this extension.

**Computing Algorithm**

**Existence of maxima**

The existence of maxima for (16) can be demonstrated by rewriting (16) as a special case of a more general function. The requirement of stationarity/invariance for $W$ in (16) can be removed, giving a general function $F(W_{ik}; G)$ which allows $W$ to vary over groups and occasions. This general function has a global maximum and a global minimum, using arguments from the single group/occasion case as in the discussion following (7). The range of $F(W_{ik}; G)$ is therefore bounded.

The requirement of stationarity/invariance for $W$ can be implemented by maximizing $F(W_{ik}; G)$ under linear constraints on the $W_{ik}$ matrices. The function $F(W_{ik}; G)$ can be compactly expressed

$$F(W_{ik}; G) = \text{tr} \left[ D_W D_S D_G D_W (D_W D_S D_W)^{-1} \right],$$  \hspace{1cm} (17)

where $D_W$, $D_S$, and $D_G$ are block diagonal supermatrices. The block diagonal matrices in $D_W$ are the $W_{ik}$. The block diagonal matrices in $D_S$ are the $\Sigma_{ik}$, and the block diagonal matrices in $D_G$ are the $G_{ik}$.

As discussed above, the maximum of (17) requires specification of an $n \times m$ matrix $W_{ik}$ in each of $r$ groups on $p$ occasions, giving a total of $n \cdot m \cdot r \cdot p = t$ parameters to be estimated. Let $V$ be a $t \times 1$ supervector containing these unknown parameters. The requirement of stationarity/invariance on the $W_{ik}$ consists of $t - (n \cdot m)$ linear equality constraints on $V$. These equality constraints can be represented in the matrix equation $CV = 0$ for a $(t - (n \cdot m)) \times t$ matrix $C$. The maximum of (16) is equal to the maximum of (17) under these linear constraints on $V$. The set of vectors $V$ which satisfy the constraints constitute a closed convex set. Furthermore, the unconstrained function (17) is bounded. Therefore a maximum of (17) under the constraints $CV = 0$ must also exist, but may not be unique.

**Identification**

The constraints required to implement stationarity/invariance are insufficient to identify $W$. In (16), $W$ can be replaced by any transformation $W^* = W^T$, where $T$ is any $m \times m$ nonsingular matrix, leaving the function value unchanged. A minimum of $m^2$ further constraints are necessary for identification. The constraints can be implemented within a constrained optimization algorithm.

Let $H$ be a fixed $n \times m$ matrix of full column rank. One way to identify $W$ is to require that $H^T W = U$, where $U$ is a fixed $m \times m$ nonsingular matrix. Assuming $H^T W$ to be nonsingular, these constraints identify $W$ since $T$ is uniquely defined as $T = (H^T W)^{-1} U$, for some initial $W_0$. As a special case, let $H$ be an $n \times m$ matrix, $m$ of whose rows would form an $m \times m$ identity matrix, the remaining rows being null. The constraints then fix the elements of selected rows in $W$ to be equal to corresponding rows in $U$. This special case will be denoted the simple equality case in what follows.
The identification constraints are trivial in the sense that given a matrix \( W \) that maximizes (16), a matrix \( T \) may be selected to identify \( W^* \) without altering the value of (16). But in practice, \( W \) is unknown, and the identification constraints are specified at the start of the maximization procedure. Our optimization algorithm will converge only for fully identified solutions. Different initial identifications need not yield identical maximum values for (16) after optimization. In this sense, the identification constraints are not trivial in practice, and some care must be taken in choosing \( H \) and \( U \).

The constraints in the simple equality case can be implemented by substitution, and do not require a fully constrained optimization algorithm. This follows since the Jacobian of the constraint equations \( H^TW = U \) is nonzero for nonsingular \( H \), and the conditions of the Implicit Function Theorem are satisfied (Avriel, 1976). Recalling the form of \( H \), we permute rows of \( H \), and corresponding rows and columns of \( \Sigma_{i(k)} \), so that \( H = (I \mid 0) \).

Then the constraints identify \( W \) by

\[
W = (U' \mid V),
\]

where \( U \) is fixed and the \((n - m) \times m\) matrix \( V \) is variable. The identified function to be maximized is

\[
F(W; G) = \sum_{k=1}^{r} \sum_{i=1}^{p} \text{tr} \left\{ (U' \mid V) \Sigma_{i(k)} G_{i(k)} \Sigma_{i(k)} \right\} \left( \begin{array}{c|c} U & \mid V \end{array} \right)
\]

\[
\times \left( (U' \mid V) \Sigma_{i(k)} \left( \begin{array}{c|c} U & \mid V \end{array} \right)^{-1} \right).
\]

The gradient matrix \( Q(W) \) is

\[
Q(W) = 2 \sum_{k=1}^{r} \sum_{i=1}^{p} \{ \Sigma_{i(k)} G_{i(k)} \Sigma_{i(k)} W [W^T \Sigma_{i(k)} W]^{-1} \}
\]

\[
- \Sigma_{i(k)} W [W^T \Sigma_{i(k)} W]^{-1}[W^T \Sigma_{i(k)} G_{i(k)} \Sigma_{i(k)} W][W^T \Sigma_{i(k)} W]^{-1} \}.
\]

The gradient matrix can be supplied to any iterative program that maximizes \( F \) employing gradient information. Note that \( U \) is held fixed and only \( V \) varies over iterative steps. The program we employ to perform the maximization in this paper is IMSL (1982) subroutine ZXCGR, a conjugate gradient algorithm. In theory, this algorithm will converge at least as quickly as a steepest descent algorithm, and usually more quickly (Avriel, 1976). In practice, we have found the algorithm to achieve swift convergence even for large problems.

**Start Values**

The function to be maximized in (19) can be shown to have local maxima. It is important that the start values for the optimization algorithm be well chosen. The following method for selecting start values has been found to work well in practice. Let \( W_0 \) be the matrix of start values for \( W \). Define the following matrices.

\[
\Sigma^* = \sum_{i=1}^{p} \sum_{k=1}^{r} \Sigma_{i(k)},
\]

\[
G^* = \sum_{i=1}^{p} \sum_{k=1}^{r} G_{i(k)} = AA',
\]

\[
\Gamma = A^T \Sigma^* A,
\]
### Table 1

Rescaled Covariances and Variances.*

Cohort 1 Covariances in Upper Triangle. Cohort 2 in Lower.

<table>
<thead>
<tr>
<th></th>
<th>VM1</th>
<th>NF1</th>
<th>LS1</th>
<th>NS1</th>
<th>SR1</th>
<th>VM2</th>
<th>NF2</th>
<th>LS2</th>
<th>NS2</th>
<th>SR2</th>
<th>VM3</th>
<th>NF3</th>
<th>LS3</th>
<th>NS3</th>
<th>SR3</th>
<th>VM4</th>
<th>NF4</th>
<th>LS4</th>
<th>NS4</th>
<th>SR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>VM1</td>
<td>.20</td>
<td>.28</td>
<td>.28</td>
<td>.16</td>
<td>.12</td>
<td>.33</td>
<td>.33</td>
<td>.22</td>
<td>.20</td>
<td>.31</td>
<td>.16</td>
<td>.14</td>
<td>.48</td>
<td>.29</td>
<td>.18</td>
<td>.32</td>
<td>.23</td>
<td>.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF1</td>
<td>.55</td>
<td>.38</td>
<td>.31</td>
<td>.27</td>
<td>.36</td>
<td>.35</td>
<td>.47</td>
<td>.39</td>
<td>.35</td>
<td>.40</td>
<td>.37</td>
<td>.43</td>
<td>.54</td>
<td>.31</td>
<td>.37</td>
<td>.44</td>
<td>.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS1</td>
<td>.40</td>
<td>.59</td>
<td>.42</td>
<td>.31</td>
<td>.34</td>
<td>.33</td>
<td>.48</td>
<td>.56</td>
<td>.42</td>
<td>.45</td>
<td>.38</td>
<td>.56</td>
<td>.42</td>
<td>.41</td>
<td>.40</td>
<td>.52</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS1</td>
<td>.61</td>
<td>.53</td>
<td>.49</td>
<td>.25</td>
<td>.20</td>
<td>.34</td>
<td>.33</td>
<td>.38</td>
<td>.52</td>
<td>.23</td>
<td>.25</td>
<td>.53</td>
<td>.40</td>
<td>.31</td>
<td>.49</td>
<td>.35</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR1</td>
<td>.47</td>
<td>.63</td>
<td>.65</td>
<td>.53</td>
<td>.26</td>
<td>.33</td>
<td>.29</td>
<td>.33</td>
<td>.40</td>
<td>.18</td>
<td>.42</td>
<td>.28</td>
<td>.21</td>
<td>.29</td>
<td>.35</td>
<td>.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VM2</td>
<td>.43</td>
<td>.39</td>
<td>.44</td>
<td>.51</td>
<td>.40</td>
<td>.23</td>
<td>.41</td>
<td>.26</td>
<td>.28</td>
<td>.34</td>
<td>.57</td>
<td>.33</td>
<td>.45</td>
<td>.24</td>
<td>.29</td>
<td>.42</td>
<td>.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF2</td>
<td>.87</td>
<td>.66</td>
<td>.50</td>
<td>.75</td>
<td>.62</td>
<td>.49</td>
<td>.39</td>
<td>.26</td>
<td>.48</td>
<td>.30</td>
<td>.20</td>
<td>.71</td>
<td>.40</td>
<td>.20</td>
<td>.42</td>
<td>.35</td>
<td>.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS2</td>
<td>.70</td>
<td>.83</td>
<td>.55</td>
<td>.66</td>
<td>.67</td>
<td>.51</td>
<td>.84</td>
<td>.48</td>
<td>.45</td>
<td>.51</td>
<td>.43</td>
<td>.47</td>
<td>.60</td>
<td>.38</td>
<td>.42</td>
<td>.53</td>
<td>.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS2</td>
<td>.43</td>
<td>.51</td>
<td>.74</td>
<td>.60</td>
<td>.54</td>
<td>.45</td>
<td>.56</td>
<td>.58</td>
<td>.35</td>
<td>.49</td>
<td>.46</td>
<td>.48</td>
<td>.55</td>
<td>.56</td>
<td>.37</td>
<td>.59</td>
<td>.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR2</td>
<td>.63</td>
<td>.58</td>
<td>.49</td>
<td>.75</td>
<td>.59</td>
<td>.46</td>
<td>.79</td>
<td>.71</td>
<td>.61</td>
<td>.20</td>
<td>.29</td>
<td>.65</td>
<td>.49</td>
<td>.29</td>
<td>.56</td>
<td>.35</td>
<td>.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VM3</td>
<td>.37</td>
<td>.48</td>
<td>.55</td>
<td>.44</td>
<td>.56</td>
<td>.38</td>
<td>.49</td>
<td>.63</td>
<td>.53</td>
<td>.43</td>
<td>.50</td>
<td>.37</td>
<td>.47</td>
<td>.42</td>
<td>.34</td>
<td>.61</td>
<td>.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF3</td>
<td>.46</td>
<td>.50</td>
<td>.56</td>
<td>.48</td>
<td>.78</td>
<td>.51</td>
<td>.60</td>
<td>.52</td>
<td>.51</td>
<td>.50</td>
<td>.32</td>
<td>.56</td>
<td>.38</td>
<td>.33</td>
<td>.58</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS3</td>
<td>.93</td>
<td>.74</td>
<td>.69</td>
<td>.74</td>
<td>.76</td>
<td>.64</td>
<td>.13</td>
<td>.86</td>
<td>.66</td>
<td>.83</td>
<td>.55</td>
<td>.68</td>
<td>.60</td>
<td>.40</td>
<td>.65</td>
<td>.57</td>
<td>.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS3</td>
<td>.61</td>
<td>.80</td>
<td>.63</td>
<td>.60</td>
<td>.74</td>
<td>.46</td>
<td>.73</td>
<td>.01</td>
<td>.65</td>
<td>.71</td>
<td>.63</td>
<td>.56</td>
<td>.89</td>
<td>.51</td>
<td>.57</td>
<td>.69</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR3</td>
<td>.58</td>
<td>.59</td>
<td>.56</td>
<td>.71</td>
<td>.65</td>
<td>.80</td>
<td>.73</td>
<td>.67</td>
<td>.88</td>
<td>.54</td>
<td>.55</td>
<td>.96</td>
<td>.79</td>
<td>.66</td>
<td>.56</td>
<td>.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Variance for Cohort 1.

<table>
<thead>
<tr>
<th>Occasion 1</th>
<th>Occasion 2</th>
<th>Occasion 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>VM</td>
<td>.39</td>
<td>.69</td>
</tr>
<tr>
<td>NF</td>
<td>.67</td>
<td>.79</td>
</tr>
<tr>
<td>LS</td>
<td>1.07</td>
<td>.91</td>
</tr>
<tr>
<td>WG</td>
<td>.68</td>
<td>.95</td>
</tr>
<tr>
<td>NS</td>
<td>.64</td>
<td>.88</td>
</tr>
<tr>
<td>SR</td>
<td>.84</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Variance for Cohort 2.

<table>
<thead>
<tr>
<th>Occasion 1</th>
<th>Occasion 2</th>
<th>Occasion 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>VM</td>
<td>.90</td>
<td>1.22</td>
</tr>
<tr>
<td>NF</td>
<td>.95</td>
<td>1.23</td>
</tr>
<tr>
<td>LS</td>
<td>1.20</td>
<td>.92</td>
</tr>
<tr>
<td>WG</td>
<td>1.07</td>
<td>1.12</td>
</tr>
<tr>
<td>NS</td>
<td>1.10</td>
<td>1.01</td>
</tr>
<tr>
<td>SR</td>
<td>.87</td>
<td>1.20</td>
</tr>
</tbody>
</table>

* Key to abbreviations:

- VM = Verbal Meaning
- NF = Number Facility
- LS = Letter Series
- WG = Word Groupings
- NS = Number Series
- SR = Spatial Relations

Number following prefix indicates occasion.

where $AA'$ is some factorization of $G^*$. Let the spectral representation of $\Gamma$ be $\Gamma = QAQ'$, where $Q$ is an $n \times n$ matrix of eigenvectors and $\Lambda$ is an $n \times n$ diagonal matrix of eigenvalues in descending order. Next calculate a matrix $W_1 = AQ_mT$, with $Q_m$ being the $n \times m$ submatrix containing the first $m$ columns of $Q$, and $T$ being an $m \times m$ nonsingular matrix. The latter will be chosen to rotate $AQ_m$ to identify its structure as $T = (H' AQ_m)^{-1} U$. For $G^* = I$, this method of calculating start values is identical to Levin's (1966) method of analysis.
Table 2

Optimal Compositing Weight Matrix \( W \).

<table>
<thead>
<tr>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>VH</td>
<td>1.00*</td>
</tr>
<tr>
<td>NF</td>
<td>0.39</td>
</tr>
<tr>
<td>LS</td>
<td>0.05</td>
</tr>
<tr>
<td>WG</td>
<td>0.93</td>
</tr>
<tr>
<td>NS</td>
<td>0.00*</td>
</tr>
<tr>
<td>SR</td>
<td>-0.44</td>
</tr>
</tbody>
</table>

Choices for the Weight Matrices

Weightings can be used to adjust the metric of the risk function across occasions and groups, and might be particularly useful when the variances of the observed variables vary widely over variables, occasions, or groups. Let

\[
\Sigma = (p, r)^{-1} \sum_{k=1}^{r} \sum_{i=1}^{p} \Sigma_{(ik)}
\]

be the average of the within-occasion covariance matrices over groups and occasions. Alternatively, the average could be taken within groups but over occasions, leading to separate averages within each group: \( \Sigma_{(i)} \). The converse is also possible, yielding averages across groups within specific occasions: \( \Sigma_{(i)} \). These alternatives lead to four choices for \( G_{(ik)} \)

\[
G_{(ik)} = D^{-1}(\Sigma),
\]

\[
G_{(ik)} = D^{-1}(\Sigma_{(i)}),
\]

\[
G_{(ik)} = D^{-1}(\Sigma_{(i)}),
\]

\[
G_{(ik)} = D^{-1}(\Sigma_{(ik)}),
\]

where \( D(\cdot) \) is a diagonal matrix containing the diagonal elements of \( \cdot \). In (25), identical weightings are applied in each occasion and group, while (26) through (28) give weightings which vary over groups or occasions, or both. We note that choosing \( G_{(ik)} \) as in (28) leads to a simultaneous component analysis in the correlation metric within each group and occasion. An analysis using (28) can be shown to be "scale-free", yielding rotationally equivalent solutions under nonsingular transformations of the observed variables (Meredith & Millsap, 1985).

Other choices for \( G_{(ik)} \) arise naturally from image theory (Guttman, 1953). The matrix

\[
\Psi_{(ik)} = D^{-1}(\Sigma_{(ik)})
\]

is a diagonal matrix of the anti-image variances for the variables in the \( k \)-th group at the \( i \)-th occasion. Setting \( G_{(ik)} = \Psi_{(ik)}^{-2} \) weights each variable in inverse proportion to its within-occasion and group anti-image variance. The matrix

\[
\Phi_{(ik)} = \Psi_{(ik)}^{-2} \Sigma_{(ik)}^{-1} \Psi_{(ik)}^{-2}
\]

is the anti-image covariance matrix for the variables in the \( k \)-th group at the \( i \)-th occasion. Setting \( G_{(ik)} = \Phi_{(ik)}^{-2} \) weights the variables by the inverse of the anti-image covariance
### Table 3

**Cohort 1: Standardized Structure Matrix**

<table>
<thead>
<tr>
<th></th>
<th>Occasion 1</th>
<th></th>
<th>Occasion 2</th>
<th></th>
<th>Occasion 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comp1</td>
<td>Comp2</td>
<td>Comp1</td>
<td>Comp2</td>
<td>Comp1</td>
</tr>
<tr>
<td>VM1</td>
<td>.82</td>
<td>.45</td>
<td>.62</td>
<td>.37</td>
<td>.68</td>
</tr>
<tr>
<td>NF1</td>
<td>.55</td>
<td>.72</td>
<td>.51</td>
<td>.63</td>
<td>.52</td>
</tr>
<tr>
<td>LS1</td>
<td>.52</td>
<td>.74</td>
<td>.44</td>
<td>.57</td>
<td>.47</td>
</tr>
<tr>
<td>WG1</td>
<td>.87</td>
<td>.52</td>
<td>.62</td>
<td>.46</td>
<td>.66</td>
</tr>
<tr>
<td>NS1</td>
<td>.37</td>
<td>.69</td>
<td>.52</td>
<td>.48</td>
<td>.45</td>
</tr>
<tr>
<td>SR1</td>
<td>.07</td>
<td>.79</td>
<td>.26</td>
<td>.58</td>
<td>.30</td>
</tr>
<tr>
<td>VM2</td>
<td>.62</td>
<td>.54</td>
<td>.87</td>
<td>.43</td>
<td>.71</td>
</tr>
<tr>
<td>NF2</td>
<td>.47</td>
<td>.71</td>
<td>.63</td>
<td>.76</td>
<td>.53</td>
</tr>
<tr>
<td>LS2</td>
<td>.48</td>
<td>.57</td>
<td>.38</td>
<td>.79</td>
<td>.47</td>
</tr>
<tr>
<td>WG2</td>
<td>.63</td>
<td>.52</td>
<td>.87</td>
<td>.42</td>
<td>.65</td>
</tr>
<tr>
<td>NS2</td>
<td>.32</td>
<td>.63</td>
<td>.31</td>
<td>.82</td>
<td>.39</td>
</tr>
<tr>
<td>SR2</td>
<td>.21</td>
<td>.58</td>
<td>.13</td>
<td>.84</td>
<td>.28</td>
</tr>
<tr>
<td>VM3</td>
<td>.68</td>
<td>.58</td>
<td>.74</td>
<td>.47</td>
<td>.89</td>
</tr>
<tr>
<td>NF3</td>
<td>.49</td>
<td>.58</td>
<td>.49</td>
<td>.66</td>
<td>.65</td>
</tr>
<tr>
<td>LS3</td>
<td>.45</td>
<td>.51</td>
<td>.34</td>
<td>.67</td>
<td>.52</td>
</tr>
<tr>
<td>WG3</td>
<td>.61</td>
<td>.51</td>
<td>.59</td>
<td>.48</td>
<td>.86</td>
</tr>
<tr>
<td>NS3</td>
<td>.39</td>
<td>.61</td>
<td>.36</td>
<td>.71</td>
<td>.56</td>
</tr>
<tr>
<td>SR3</td>
<td>.10</td>
<td>.51</td>
<td>.15</td>
<td>.63</td>
<td>.10</td>
</tr>
</tbody>
</table>

The matrix. Both choices for $G_{ik}$ follow the principle of weighting each variable in inverse proportion to the "error variance" in that variable, with the "error variance" defined in terms of the regression of each variable on the other variables measured concurrently. Such choices for $G_{ik}$ yield "scale-free" solutions (Meredith & Millsap, 1985).

**Component Pattern and Structure Matrices**

Traditionally, the component pattern and component structure matrices are scrutinized in interpreting the component solution. The structure matrix $S$ contains the covariances (or correlations) between the observed variables and the component scores. The pattern matrix $P$ contains the regression weights for deriving the observed variables from the component scores. In the multiple group/occasion case, the pattern and structure matrices will ordinarily vary over groups and occasions. To define the pattern and structure matrices, we require the following matrices and vectors

$$X_{ik} = [X_{1ik}, X_{2ik}, \ldots, X_{pik}]$$  \hspace{1cm} (30)

$$D_w = \begin{bmatrix} W & 0 & \cdots & 0 \\ 0 & W & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W \end{bmatrix} = I \otimes W,$$  \hspace{1cm} (31)
Table 4
Cohort 2: Standardized Structure Matrix.

<table>
<thead>
<tr>
<th></th>
<th>Occasion 1</th>
<th></th>
<th>Occasion 2</th>
<th></th>
<th>Occasion 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comp1</td>
<td>Comp2</td>
<td>Comp1</td>
<td>Comp2</td>
<td>Comp1</td>
<td>Comp2</td>
</tr>
<tr>
<td>VM1</td>
<td>.88</td>
<td>.61</td>
<td>.79</td>
<td>.60</td>
<td>.74</td>
<td>.55</td>
</tr>
<tr>
<td>NF1</td>
<td>.73</td>
<td>.77</td>
<td>.66</td>
<td>.68</td>
<td>.66</td>
<td>.64</td>
</tr>
<tr>
<td>LS1</td>
<td>.48</td>
<td>.81</td>
<td>.42</td>
<td>.65</td>
<td>.51</td>
<td>.71</td>
</tr>
<tr>
<td>WG1</td>
<td>.86</td>
<td>.64</td>
<td>.70</td>
<td>.65</td>
<td>.63</td>
<td>.53</td>
</tr>
<tr>
<td>NS1</td>
<td>.57</td>
<td>.81</td>
<td>.58</td>
<td>.63</td>
<td>.63</td>
<td>.63</td>
</tr>
<tr>
<td>SR1</td>
<td>.41</td>
<td>.77</td>
<td>.41</td>
<td>.71</td>
<td>.47</td>
<td>.61</td>
</tr>
<tr>
<td>VM2</td>
<td>.82</td>
<td>.63</td>
<td>.91</td>
<td>.62</td>
<td>.80</td>
<td>.57</td>
</tr>
<tr>
<td>NF2</td>
<td>.72</td>
<td>.70</td>
<td>.76</td>
<td>.77</td>
<td>.72</td>
<td>.65</td>
</tr>
<tr>
<td>LS2</td>
<td>.59</td>
<td>.72</td>
<td>.59</td>
<td>.79</td>
<td>.64</td>
<td>.75</td>
</tr>
<tr>
<td>WG2</td>
<td>.72</td>
<td>.62</td>
<td>.87</td>
<td>.63</td>
<td>.74</td>
<td>.60</td>
</tr>
<tr>
<td>NS2</td>
<td>.46</td>
<td>.61</td>
<td>.45</td>
<td>.79</td>
<td>.49</td>
<td>.67</td>
</tr>
<tr>
<td>SR2</td>
<td>.43</td>
<td>.69</td>
<td>.31</td>
<td>.83</td>
<td>.43</td>
<td>.71</td>
</tr>
<tr>
<td>VM3</td>
<td>.73</td>
<td>.70</td>
<td>.76</td>
<td>.65</td>
<td>.90</td>
<td>.70</td>
</tr>
<tr>
<td>NF3</td>
<td>.65</td>
<td>.71</td>
<td>.68</td>
<td>.73</td>
<td>.78</td>
<td>.74</td>
</tr>
<tr>
<td>LS3</td>
<td>.48</td>
<td>.68</td>
<td>.44</td>
<td>.69</td>
<td>.58</td>
<td>.82</td>
</tr>
<tr>
<td>WG3</td>
<td>.65</td>
<td>.64</td>
<td>.74</td>
<td>.67</td>
<td>.89</td>
<td>.68</td>
</tr>
<tr>
<td>NS3</td>
<td>.48</td>
<td>.62</td>
<td>.53</td>
<td>.69</td>
<td>.63</td>
<td>.82</td>
</tr>
<tr>
<td>SR3</td>
<td>.34</td>
<td>.61</td>
<td>.28</td>
<td>.66</td>
<td>.35</td>
<td>.77</td>
</tr>
</tbody>
</table>

\[ Z'_{(k)} = [Z'_{1(k)}, Z'_{2(k)}, \ldots, Z'_{p(k)}] = X'_w D_w, \]  
\[ \Sigma_{(k)} = \begin{bmatrix} 
\Sigma_{11(k)} & \Sigma_{12(k)} & \cdots & \Sigma_{1p(k)} \\
\Sigma_{21(k)} & \Sigma_{22(k)} & \cdots & \Sigma_{2p(k)} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{p1(k)} & \Sigma_{p2(k)} & \cdots & \Sigma_{pp(k)} 
\end{bmatrix}. \]  

The supervector $X'_w$ in (30) contains the $n$ observed random variables measured on each of $p$ occasions in the $k$-th group. The block diagonal matrix $D_w$ in (31) is the Kronecker product of a $p \times p$ identity matrix $I$ and the compositing matrix $W$. The supervector $Z'_{(k)}$ in (32) contains the $m$ component scores in each of $p$ occasions within the $k$-th group. The supermatrix $\Sigma_{(k)}$ in (33) contains the within-occasion and cross-occasion covariances within the $k$-th group.

A component structure supermatrix $S_{(k)}$ for the $k$-th group can be defined

\[ S_{(k)} = \Sigma_{(k)} D_w = \begin{bmatrix} 
\Sigma_{11(k)} W & \Sigma_{12(k)} W & \cdots & \Sigma_{1p(k)} W \\
\Sigma_{21(k)} W & \Sigma_{22(k)} W & \cdots & \Sigma_{2p(k)} W \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{p1(k)} W & \Sigma_{p2(k)} W & \cdots & \Sigma_{pp(k)} W 
\end{bmatrix}. \]  

The block diagonal submatrices $\Sigma_{ii(k)} W$, $i = 1, \ldots, p$, are the covariance matrices between...
Table 5
Component Correlations and Standard Deviations.

Cohort 1 Correlations in Upper Triangle, Cohort 2 in Lower.

<table>
<thead>
<tr>
<th>Occasion</th>
<th>Occasion 1</th>
<th>Occasion 2</th>
<th>Occasion 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.46</td>
<td>.70</td>
<td>.74</td>
</tr>
<tr>
<td>2</td>
<td>.67</td>
<td>.56</td>
<td>.76</td>
</tr>
<tr>
<td>3</td>
<td>.85</td>
<td>.63</td>
<td>.40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Occasion</th>
<th>Occasion 4</th>
<th>Occasion 5</th>
<th>Occasion 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>.66</td>
<td>.85</td>
<td>.62</td>
</tr>
<tr>
<td>5</td>
<td>.77</td>
<td>.71</td>
<td>.84</td>
</tr>
<tr>
<td>6</td>
<td>.59</td>
<td>.82</td>
<td>.58</td>
</tr>
</tbody>
</table>

Component Standard Deviations.

<table>
<thead>
<tr>
<th>Occasion</th>
<th>Occasion 1</th>
<th>Occasion 2</th>
<th>Occasion 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort 1</td>
<td>1.35</td>
<td>2.61</td>
<td>1.69</td>
</tr>
<tr>
<td>Cohort 2</td>
<td>1.80</td>
<td>3.12</td>
<td>2.07</td>
</tr>
</tbody>
</table>

the observed variables and the component scores within the i-th occasion for group k. These submatrices could be averaged over occasions if a single representative structure matrix is desired. The off-diagonal submatrices $\Sigma_{i(k)}W$, $i \neq j$, are the covariance matrices between the observed variables measured in occasion i and the component scores measured in occasion j.

Occasion-specific pattern matrices $P_{i(k)}$, $i = 1, \ldots, p$, can be calculated within the k-th group as

$$P_{i(k)} = \Sigma_{i(k)}W(W^T\Sigma_{i(k)}W)^{-1}.$$  \hfill (35)

The occasion-specific pattern matrices only consider within-occasion information in the regression of $X_{(k)}$ on $Z_{(k)}$. If an “average” pattern matrix is desired, the pattern matrices in (35) could be directly averaged, or an average could be calculated from the block diagonal elements of

$$P_{i(k)} = \Sigma_{i(k)}D_w(D_w^T\Sigma_{i(k)}D_w)^{-1}.$$  \hfill (36)

This pattern supermatrix may not be informative, and will not exist unless $D_w\Sigma_{i(k)}D_w$ is nonsingular.

Finally, the component covariance and correlation matrices are fundamental. For the k-th group, these are calculated as

$$\Phi_{(k)} = D_w^T\Sigma_{(k)}D_w,$$

and the correlations can be obtained by rescaling (37) with the reciprocal square roots of its diagonal elements.
Table 6
Standardized Pattern Matrices.

<table>
<thead>
<tr>
<th></th>
<th>Cohort 1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Occasion 1</td>
<td>Occasion 2</td>
<td>Occasion 3</td>
</tr>
<tr>
<td>VM</td>
<td>.78</td>
<td>.09</td>
<td>.83</td>
</tr>
<tr>
<td>NF</td>
<td>.28</td>
<td>.59</td>
<td>.39</td>
</tr>
<tr>
<td>LS</td>
<td>.22</td>
<td>.64</td>
<td>.08</td>
</tr>
<tr>
<td>WG</td>
<td>.80</td>
<td>.15</td>
<td>.84</td>
</tr>
<tr>
<td>NS</td>
<td>.06</td>
<td>.66</td>
<td>-.03</td>
</tr>
<tr>
<td>SR</td>
<td>-.37</td>
<td>.96</td>
<td>-.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Cohort 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Occasion 1</td>
<td>Occasion 2</td>
<td>Occasion 3</td>
</tr>
<tr>
<td>VM</td>
<td>.87</td>
<td>.03</td>
<td>.85</td>
</tr>
<tr>
<td>NF</td>
<td>.39</td>
<td>.51</td>
<td>.45</td>
</tr>
<tr>
<td>LS</td>
<td>-.10</td>
<td>.88</td>
<td>.16</td>
</tr>
<tr>
<td>WG</td>
<td>.79</td>
<td>.11</td>
<td>.78</td>
</tr>
<tr>
<td>NS</td>
<td>.04</td>
<td>.79</td>
<td>-.05</td>
</tr>
<tr>
<td>SR</td>
<td>-.18</td>
<td>.89</td>
<td>-.32</td>
</tr>
</tbody>
</table>

Estimation

One would implement our procedures by substituting for $\Sigma_{(k)}$ (and hence $\Sigma_{u(k)}$) the corresponding unbiased estimators $\hat{\Sigma}_{(k)}$. Furthermore, the $G_{(ik)}$ matrices could be estimated using diagonal elements of either $\hat{\Sigma}_{(k)}$ or $\hat{\Sigma}_{u(k)}^{-1}$ in the first two cases we have discussed for the choice of weights. In general, the estimator of $W$ derived by using $\hat{\Sigma}_{(k)}$ in place of $\Sigma_{(k)}$ will be consistent. Under normality assumptions, one would obtain maximum likelihood estimates of $W$ by substituting $\hat{\Sigma}_{(k)}$ for $\Sigma_{(k)}$ and requiring $G_{(k)} = I$, $G_{(ik)} = D^{-1}(\hat{\Sigma}_{u(k)})$, or $G_{(ik)} = D(\hat{\Sigma}_{u(k)}^{-1})$. Maximum likelihood estimates are possible given appropriate identification constraints and given that the function being maximized has a unique global maximum. The maximization function in (16) must be weighted by sample sizes to yield the maximum likelihood estimates. Other choices of weight matrices require careful consideration to determine the properties of the corresponding estimators of $W$. These results follow from standard theorems (Cramer, 1945; Rao, 1973).

Measures of Adequacy

The adequacy of the solution can be evaluated by calculating the risk functions for each occasion and group as given in (15). Assuming a weighted analysis, a measure of "variance explained" can be obtained by calculating the ratio of (15), summed over groups and occasions, to the same risk with $W = 0$, and subtracting this ratio from one. Two
practical issues arise in connection with the evaluation of estimated risk. One involves the
units of measurement for the original variables. The other depends on the adequacy of a
simple (weighted or unweighted) component analysis of the variables within each occasion
and group. We could give up the requirement of stationarity and invariance for $W$, and
solve for a series of $W_{ab}$ matrices. The natural comparison would involve a subtraction of
the unconstrained risk from the constrained risk. This could be divided by a risk calcul-
ated with $W = 0$ to provide a scaling of the risk in absolute units.

An Example

The data we use are taken from Nesselroade and Baltes (1974), and consist of six
subtest scores from the Primary Mental Abilities Test (Thurstone & Thurstone, 1962). We
use only the scores for male subjects in the two youngest cohorts. Each subject was
measured on three occasions; grades 7, 8, and 9 for the first group and grades 8, 9, and 10
for the second group. Further details can be found in Nesselroade and Baltes, and
Thurstone and Thurstone (1962). The sample sizes are 118 and 123, respectively. We
calculated an $18 \times 18$ covariance matrix in each group. Prior to analysis the data were
rescaled by the square root of the average variance for each variable, the average being
computed over occasions and groups. A weighted analysis was then performed using the
matrix $G_{ab}$ in (28).

Table 1 gives the rescaled covariance matrices for both groups. The younger group is
above the main diagonal. The $W$ matrix is in table 2. Note that $W$ was identified by fixing
the starred elements in $W$, so that the first component corresponds in a rough way to
crystallized intelligence and the second to fluid intelligence. We set $m = 2$, and no further
rotations were performed. The measure of adequacy for this solution, obtained by divid-
ing (15), summed over groups and occasions, by the same risk with $W = 0$ and subtract-
ing this ratio from one, is .70. Tables 3 and 4 give the standardized structure matrices for
the two groups. Note that the elements in these tables are correlation coefficients. Table 5
provides the correlations of the components within and across occasions for both cohorts,
and the component standard deviations. The younger cohort is above the main diagonal.

References

Statistical Libraries.
Millsap, R. E. (June, 1986) Component vs factor analytic approaches to longitudinal data. Paper presented at
the annual meeting of the Psychometric Society, Toronto, Ontario, Canada.
Monographs of the Society for Research in Child Development (Serial No. 154).
Pearson, K. (1901) On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 50,
157–175.
ten Berge, J. M. F. (1986) Rotation to perfect congruence and the cross-validation of component weights across
populations. Multivariate Behavioral Research, 21, 41–64.
Thurstone, L. L., & Thurstone, T. G. (1962) SRA Primary mental abilities. Chicago: Science Research Associa-
tes.

Manuscript received 1/30/86
Final version received 3/10/87