Simultaneous Components Analysis

ten Berge, Jos M.F.; Kiers, Henk A.L.; van der Stel, Véronique

Published in:
Statistica Applicata

Publication date:
1992

Statistica Applicata Vol. 4, n. 4, 1992, 377-392

SIMULTANEOUS COMPONENTS ANALYSIS

Jos M.F. ten Berge, Henk A.L. Kiers and Véronique van der Stel

University of Groningen.

Simultaneous Components Analysis is a generalization of Principal Components Analysis for the situation where the same variables have been measured in two or more populations. A common matrix of component weights is used to ensure that the components are the same linear combinations of the same variables in each population. This common matrix of weights is determined such that the explained variance, summed over populations, is a maximum. The present paper deals with certain unknown properties of simultaneous components. The main result is that one always finds simultaneous components that explain a considerable amount of variance. Furthermore, it will be shown that neither using small samples, nor replacing optimal weights by simple weights, need to detract seriously from the usefulness of simultaneous components.

The research of H.A.L. Kiers has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

When the same set of variables has been administered to subjects from two or more different groups (populations, experimental conditions) it may be desirable to carry out some sort of generalized principal components analysis. Especially in cases where separate principal components analyses (PCA) in each group fail to reveal components (possibly rotated) with common interpretations across groups, a generalized PCA seems to be indicated.

Various generalizations of PCA have been developed, in response to whatever property of PCA is deemed essential. For instance, Flury (1984) has argued that PCA is essentially a method of diagonalizing a correlation matrix. Accordingly, he has developed a generalization of PCA, aimed at diagonalizing a set of correlation matrices simultaneously. Because most applications of PCA are aimed at goals other than diagonalizing the correlation matrix, this particular generalized PCA is unlikely to be of much use in practice.

A more promising approach has been taken by Levin (1966) who capitalized on the property that PCA produces a loading matrix A such that AA' resembles the correlation matrix of the variables. Accordingly, he suggested a generalized method based on carrying out a PCA on the average of the correlation matrices. A different way to arrive at essentially the same components as Levin's has been

offered by Rizzi and Vichi (1992). They capitalized on the fact that PCA can also be derived as a rotation method for the variables, producing components with maximal variance. Although they derived their method in the context of three–way analysis the adaption to data matrices from different samples is straight–forward. Yet another method has been developed by Millsap and Meredith (1988). They started from the position, endorsed by the present authors, that PCA is essentially a method that generates linear combinations (components) of the variables which explain as much variance as possible in these variables. Furthermore, they paired this position with the concept of stationary (invariant) component weights: If the same component weights are used in each group to define the components as linear combinations of the variables, and the variables are the same for each group, then the components have the same meaning in each group. The resulting method, called Simultaneous Components Analysis (SCA), generalizes PCA by computing one matrix of weights, defining components with the same meaning across groups, such that the total amount of variance explained by these components is a maximum. Millsap and Meredith (1988) have suggested a gradient method to find the optimal weights in SCA. This gradient method appeared to have certain disadvantages, that could be avoided by adopting an alternating least squares algorithm for SCA, see Kiers and Ten Berge (1989) and Kiers (1990).

A comparison of SCA and STATIS (L'Hermier des Plantes, 1976; Escoufier, 1980; Coppi, 1986) has been carried out by Kiers (1988, p. 28).

The purpose of the present paper is to give an overview of certain properties of SCA that may be of interest to practitioners. First, the main formulae of SCA will be reviewed, and properties that can be inferred from these formulae will be summarized. Next, the question, how much variance can be lost when SCA is used instead of a separate PCA for each group, will be addressed by mathematical means. Finally, some results from simulations will be reported. Specifically, it will be shown what sample sizes are needed to obtain reliable inferences with respect to explained variance and with respect to interpretations; how much explained variance is lost by using simultaneous components instead of separate principal components, and how much explained variance is lost when the optimal weights for SCA are replaced by simple values (1, –1, or 0).

1. THE MAIN FORMULAE OF SCA

Let $X_1,\ldots,X_m$ be data matrices of order $n_i$xk, obtained from m different groups of size n_i, i=1,...,m, where the same k variables have been administered. The mechanics of SCA can be explained most conveniently by first considering a simple alternative. This alternative is PCA on the pooled data matrix

$$X = \begin{pmatrix} X_1 \\ . \\ . \\ . \\ X_m \end{pmatrix} \tag{1}$$

of order nxk, where $n = \sum n_i$. In PCA on X, a kxr matrix B of component weights is determined such that the resulting nxr matrix XB=F of component scores yields the best 'reconstruction' of X by multiple linear regression. That is, the matrix B is determined such that the residual sum squares

$$\| X - FP' \|^2 = \| X - XBP' \|^2 = \sum_i \| X_i - X_i BP' \|^2 \tag{2}$$

has the smallest possible minimum as a function of P. Upon writing the optimal P' for arbitrary but fixed B as $P'=(B'X'XB)^{-1}B'X'X$ it is not difficult to show that the optimal B is obtained as the matrix containing the r principal eigenvectors of (X'X). Levin's generalized PCA (Levin, 1966) is essentially equivalent to the above 'PCA on X'.

Although PCA on X may seem to ignore the distinction between groups in that it yields components from X as columns of F, it can also be interpreted as a method that yields common components across the m groups. That is, if we partition F as

$$F = XB = \begin{pmatrix} X_1 \\ . \\ . \\ . \\ X_m \end{pmatrix} B = \begin{pmatrix} X_1 B \\ . \\ . \\ . \\ X_m B \end{pmatrix} \tag{3}$$

then it is clear that B defines the same linear combinations of the same variables in each group, thus providing simultaneous components across groups. At this

point it becomes easy to explain what SCA is about: SCA consists of an improved version of PCA on X. The improvement is twofold. First, instead of minimizing (2), we can do better by minimizing

$$f(B,P_1,\ldots,P_m)=\sum_i \left\| X_i - X_iBP_i' \right\|^2 \qquad (4)$$

which uses optimal linear regressions for each group separetely. In retrospect, (2) can be conceived of as (4) with the additional constraint that $P_1=\ldots=P_m$. Clearly, the minimum of (4) is smaller than the minimum of (2), even if one were to use the B which minimizes (2). Moreover, the very fact that that choice for B need not be the best choice for minimizing (4) offers a second possibility to reduce the residual sum of squares: We can find the B for which (4) has the smallest minimum as a function of $P_1,\ldots,P_m$. By minimizing (4) over $P_1,\ldots,P_m$ and B, SCA minimizes the unexplained variance, summed across groups.

Kiers and Ten Berge (1989) have shown how (4) can be minimized by an alternating least squares algorithm, updating B for fixed $P_1,\ldots,P_m$, and updating $P_1,\ldots,P_m$ for fixed B. In the present paper the computational details of this algorithm for SCA do not play any particular role. However, it is important to note that the optimal P_i, for fixed B, can be written as $P_i' = \left(B'X_i'X_iB\right)^{-1}\left(B'X_i'X_i\right)$. As a result, (4) can be expressed as a function of B as

$$\begin{aligned} f^*(B) &= \sum_i \left\| X_i - X_iB\left(B'X_i'X_iB\right)^{-1}B'X_i'X_i \right\|^2 \\ &= \sum_i \| X_i \|^2 - \sum_i \mathrm{tr}\left(B'\left(X_i'X_i\right)^2B \right)\left(B'X_i'X_iB\right)^{-1} \\ &= \sum_i \| X_i \|^2 - \sum_i \mathrm{tr}\left(B'C_i^2B\right)\left(B'C_iB\right)^{-1}, \end{aligned} \qquad (5)$$

where $C_i \equiv X_i'X_i$, i=1,...,m. Accordingly, minimizing $f^*(B)$ is tantamount to maximizing the function

$$g(B)=\sum_i \mathrm{tr}\left(B'C_i^2B\right)\left(B'C_iB\right)^{-1}. \qquad (6)$$

Whereas (4) represents SCA as a method for minimizing unexplained variance, (6) represents SCA equivalently as a method for maximizing the explained variance, summed across groups. It is important to note that SCA, like PCA, is insensitive to orthogonal or oblique rotations. That is, neither (4) nor (6) is affected by nonsingular transformations of B. In other words, SCA is, like PCA, concerned with finding an optimal subspace rather than r optimal directions. Contrary to PCA, however, solutions for different values of r are not nested in SCA. That is, a solution in r–1 dimensions can not be obtained from the r–dimensional SCA solution, as it can in PCA, but requires an independent run of the program.

Above, it was shown that 'PCA on X' can be conceived of as a constrained variant of SCA. Conversely, SCA can be conceived of as a constrained variant of 'PCA on $X_1,\ldots,X_m$' separately. That is, PCA on X_i minimizes $\left\| X_i - X_iB_iP_i' \right\|^2$, hence PCA on $X_1,\ldots,X_m$' minimizes $\sum_i \left\| X_i - X_iB_iP_i' \right\|^2$. It is clear from (4) that SCA is a constrained version of this, where it is required that $B_1=\ldots=B_m$. This point underlies the next section, where we shall address the question how much explained variance can be lost when the constraint $B_1=\ldots=B_m$ is imposed on 'PCA on $X_1,\ldots,X_m$'.

2. THE PRICE OF SIMULTANEOUS COMPONENTS

In the few years since SCA has become available, the amounts of variance explained by SCA have been surprisingly high compared to those explained by PCA on $X_1,\ldots,X_m$. For instance, if PCA on X_1 and X_2 explains 60% of variance, then SCA is likely to explain 57% or 58% of variance. This means that the price of simultaneous components tends to be low. In the present section we try to find an explanation for this remarkable finding. This will be done by searching for those (contrived) data which reveal the largest difference between the variance explained by PCA on X_1 and X_2 on the one hand, and by SCA on the other hand. We shall only consider the case m=2, and attempt to maximize the normalized difference (nd) in explained variance, defined as the non–negative ratio

$$\mathrm{nd}(C_1,C_2)=\frac{\sum_{i=1}^{r}\left(\lambda_i(C_1)+\mu_i(C_2)\right) - \sum_{j=1}^{2}\mathrm{tr}\left(B'C_j^2B\right)\left(B'C_jB\right)^{-1}}{\left(\mathrm{tr}\ C_1+\mathrm{tr}\ C_2\right)}, \qquad (7)$$

where λ_i and μ_i are i–th largest eigenvalues of C_1 and C_2, respectively, and B is the optimal kxr matrix of component weights for SCA.

Any attempt to maximize (7) in general is severely hampered by the absence of an explicit expression for the optimal B in SCA, or the amount of variance explained in SCA. Therefore, we have to adopt further simplifying restrictions. First, we shall only consider cases where r=1, because nd cannot be large when r is large (nd will be zero if r=k), and because this simplifies the expression for the SCA

solution, as will become evident below. In addition, it will be assumed that both C_1 and C_2 have the same eigenvectors and have the same eigenvalues, but in reverse order. This is a severe limitation, but it can be justified by the implication that the most important component of C_1 is the least important component of C_2, and vice versa. This seems to make it impossible for a simultaneous component to explain much variance, while PCA on X_1 and X_2 can still be very successful with r=1. Finally, all eigenvalues of C_1 and C_2 are taken equal except for the one that corresponds to the dominant component. Without loss of generality, the imposed limitations can be summarized by defining a C_1 and C_2 such that $C_1 = K\Lambda_1 K$ for some orthonormal kxk matrix K and some diagonal matrix Λ_1 containing the eigenvalues $\lambda, 1, \ldots, 1$, with $\lambda > 1$, and $C_2 = K\Lambda_2 K'$, with Λ_2 diagonal, containing the eigenvalues $1, 1, \ldots, 1, \lambda$ of C_2. In the sequel, this case will be referred to as the 'reversed eigenvalue case' because C_1 and C_2 differ only in that their eigenvalues are in opposite orders. Note that the columns of K are eigenvectors of both C_1 and C_2.

In the reversed eigenvalue case with r=1 the SCA solution can be written explicitly:

Result 1. In the reversed eigenvalue case with r=1 a globally optimal kx1 matrix B for SCA, written as **b**, is $\mathbf{k}_1+\mathbf{k}_k$, where k_i is the i–th column of K.

Proof. Using the property that $C_1 = K\Lambda_1 K'$ and $C_2 = K\Lambda_2 K'$ we can write the explained variance of SCA, see (6), as

$$g(\mathbf{b}) = \frac{\mathbf{b}'C_1^2\mathbf{b}}{\mathbf{b}'C_1\mathbf{b}} + \frac{\mathbf{b}'C_2^2\mathbf{b}}{\mathbf{b}'C_2\mathbf{b}} = \frac{\mathbf{x}'\Lambda_1^2\mathbf{x}}{\mathbf{x}'\Lambda_1\mathbf{x}} + \frac{\mathbf{x}'\Lambda_2^2\mathbf{x}}{\mathbf{x}'\Lambda_2\mathbf{x}}, \tag{8}$$

with $\mathbf{x}=K'\mathbf{b}$. Let **x** be partitioned as

$$\mathbf{x}' = \left(\pm\sqrt{a}, \beta', \pm\sqrt{c}\right), \tag{9}$$

where β is a k–2 vector, with $\beta'\beta = b$, and $a \geq 0$, and $c \geq 0$. Then using the specific form of the eigenvalues in Λ_1 and Λ_2 yields

$$g(\mathbf{b}) = \frac{\lambda^2 a + b + c}{\lambda a + b + c} + \frac{a + b + \lambda^2 c}{a + b + \lambda c}. \tag{10}$$

If we take $\mathbf{b} = \mathbf{k}_1 + \mathbf{k}_k$, then a=c=1 and β=0, yielding

$$g(\mathbf{k}_1 + \mathbf{k}_k) = \frac{2(\lambda^2+1)}{(\lambda+1)}. \tag{11}$$

The optimality of this **b** for SCA will now be proven by proving that

$$g(\mathbf{b}) \leq \frac{2(\lambda^2+1)}{(\lambda+1)}. \tag{12}$$

Starting from (10) we can write g(b) as

$$g(\mathbf{b}) = \frac{(\lambda^2 a + b + c)(a + b + \lambda c) + (\lambda a + b + c)(a + b + \lambda^2 c)}{(\lambda a + b + c)(a + b + \lambda c)} = \frac{2\lambda^3 ac + \lambda^2 u + \lambda u + 2v}{\lambda^2 ac + \lambda u + v}, \tag{13}$$

where $u \equiv (a^2 + ab + bc + c^2)$ and $v \equiv (a+b)(b+c)$. Tantamount to proving (12) is proving that

$$2(\lambda^2+1)(\lambda^2 ac + \lambda u + v) \geq (\lambda+1)(2\lambda^3 ac + \lambda^2 u + \lambda u + 2v), \tag{14}$$

which can be simplified to

$$\lambda^3(u - 2ac) + \lambda^2(2v - 2u + 2ac) + \lambda(u - 2v) \geq 0. \tag{15}$$

Dividing by $\lambda(\lambda-1)$ reduces (15) to

$$\lambda(u - 2ac) + (2v - u) \geq 0. \tag{16}$$

Clearly, for $\lambda > 1$ we have

$$\begin{aligned} &\lambda(u - 2ac) + (2v - u) \geq u - 2ac + 2v - u = 2v - 2ac \\ &= 2ab + 2ac + 2b^2 + 2bc - 2ac = 2b(a + b + c) \geq 0. \end{aligned} \tag{17}$$

Equality holds in (17) for $\lambda > 1$ if and only if u=2ac and b=0, that is, if and only if a=c and b=0.

Result 1 implies that the SCA solution for r=1 in the reversed eigenvalue case defines the simultaneous component as the sum of the first and last component of both C_1 and C_2. In addition, the associated maximum value of g(**b**) has been obtained explicitly. This allows us to specify the normalized difference in the case under consideration as a function of λ only. That is, in the reversed eigenvalue case with r=1 we have

$$nd(\lambda) = \frac{2\lambda - 2(\lambda^2+1)(\lambda+1)^{-1}}{2(\lambda + k - 1)} = \frac{\lambda - 1}{(\lambda+1)(\lambda+k-1)}. \tag{18}$$

We can now maximize nd in this special case over λ.

Result 2: In the reversed eigenvalue case with r=1 the maximum normalized difference is $\left(k + 2 + 2\sqrt{2k}\right)^{-1}$, and it is obtained for $\lambda = 1 + \sqrt{2k}$.

Proof: Differentiating nd with respect to λ, and setting the derivative to zero yields

$$\lambda^2 - 2\lambda - 2k + 1 = 0 \ , \tag{19}$$

which has only one positive root $\lambda = 1+\sqrt{2k}$. Note that the negative root is inappropriate as an eigenvalue of a covariance matrix. The associated maximum value of nd is

$$nd\left(1+\sqrt{2k}\right) = \frac{\sqrt{2k}}{\left(2+\sqrt{2k}\right)\left(k+\sqrt{2k}\right)} = \\ = \frac{\sqrt{2k}}{\left(4k+k\sqrt{2k}+2\sqrt{2k}\right)} = \frac{1}{\left(2\sqrt{2k}+k+2\right)} \ . \tag{20}$$

It is clear that this value is inversely related to k. It follows that the largest nd within the class of 'reversed eigenvalue cases' with r=1 occurs with k=2, where nd= 1/8 which corresponds to a gap of 12.5% variance explained between PCA on X_1, X_2 and SCA.

Result 2 has an important implication for SCA. This can be explained most easily if we turn to correlation matrices R_1 and R_2 instead of covariance or cross–products matrices C_1 and C_2. Obviously, matrices with eigenvalues $\lambda, 1, \ldots, 1$, cannot be correlation matrices, because their sum is not equal to the number of variabes. To remedy this, we multiply these eigenvalues by $k(\lambda + k - 1)^{-1}$, which renders their sum equal to k, but leaves nd unaffected. In the case k=2, where $\lambda = 3$, we obtain $\Lambda_1 = \begin{pmatrix} 1.5 & 0 \\ 0 & .5 \end{pmatrix}$ and $\Lambda_2 = \begin{pmatrix} .5 & 0 \\ 0 & 1.5 \end{pmatrix}$. Furthermore, correlation matrices of order 2x2 always have the same matrix of eigenvectors, that is, $K = \begin{pmatrix} \sqrt{.5} & \sqrt{.5} \\ \sqrt{.5} & -\sqrt{.5} \end{pmatrix}$. As a result, the pair of correlation matrices with reversed eigenvalues, for which nd is a maximum, is

$$R_1 = K\Lambda_1 K' = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} \text{ and } R_2 = K\Lambda_2 K' = \begin{pmatrix} 1 & -.5 \\ -.5 & 1 \end{pmatrix} . \tag{21}$$

The specific form of these correlation matrices may come as a surprise and merits a closer examination.

It is readily seen that, within the class of correlation matrices under consideration, the case where nd is a minimum is the limiting case in which

$\Lambda_1 = \Lambda_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, where nd=0 because R_1=R_2. In fact, it can be shown quite generally that, as soon as one correlation matrix in a pair is the kxk identity matrix, the SCA solution is the PCA solution of the other correlation matrix, and nd=0. For k=2 the other extreme would seem to be the case where $\Lambda_1 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ and $\Lambda_2 = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$, corresponding to $R_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ and $R_2 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Surprisingly, however, in this opposite case nd is also zero. To verify this, consider the SCA solution $\mathbf{b}=\mathbf{k}_1+\mathbf{k}_2$. Noting that $R_1^2\,\mathbf{k}_2 = R_1\mathbf{k}_2 = 0$ and that $R_2^2\mathbf{k}_1 = R_2\mathbf{k}_1 = 0$ we have

$$g(\mathbf{b}) = \frac{(\mathbf{k}_1+\mathbf{k}_2)'R_1^2(\mathbf{k}_1+\mathbf{k}_2)}{(\mathbf{k}_1+\mathbf{k}_2)'R_1(\mathbf{k}_1+\mathbf{k}_2)} + \frac{(\mathbf{k}_1+\mathbf{k}_2)'R_2^2(\mathbf{k}_1+\mathbf{k}_2)}{(\mathbf{k}_1+\mathbf{k}_2)'R_2(\mathbf{k}_1+\mathbf{k}_2)} = \\ = \frac{\mathbf{k}_1'R_1^2\mathbf{k}_1}{\mathbf{k}_1'R_1\mathbf{k}_1} + \frac{\mathbf{k}_2'R_2^2\mathbf{k}_2}{\mathbf{k}_2'R_2\mathbf{k}_2} = 4, \tag{22}$$

which equals the amount of variance explained by PCA on X_1 and X_2, that is 100%. The reason why applying $\mathbf{b}=\mathbf{k}_1+\mathbf{k}_2$ in SCA is just as effective as applying $\mathbf{k}_1$ to R_1 (PCA on X_1) and $\mathbf{k}_2$ to R_2 (PCA on X_2) rests in the property that $\mathbf{k}_1$ vanishes when premultiplied by R_2^2 or R_2, whereas $\mathbf{k}_2$ vanishes when premultiplied by R_1^2 or R_1, see (22).

At this point it becomes possible to understand why the maximum of nd for k=2 in the reversed eigenvalue case occurs when the correlations between the two variables involved are .5 and –.5, respectively, see (21). In the extreme case where these correlations are both 0, nd has a minimum (zero). However, it has just been shown that nd is also zero in the opposite extreme case, where the correlations are 1 and –1, respectively. The net result is that the maximum of nd is to be expected somewhere near the middle, where the correlations are .5 and –.5. respectively. The fact that this maximum occurs precisely at this 'middle' is a detail of merely esthetical value.

It can be argued that we have so far been concerned with limited cases only. Indeed, the maximal discrepancy of 12.5% explained variance between PCA on X_1, X_2 and SCA has only been established for the reversed eigenvalue case with k=2.

At this point, however, the relevance of this result can be broadened. One the one hand, it can be verified that a discrepancy as high as 12.5% will be observed for any even value of k, if we take k uncorrelated variables except that the correlations $r_{12}=r_{34}=r_{k-1,k}$ are .5 in R_1 and –.5 in R_2, respectively, using $r=\frac{1}{2}k$. One the other hand, there is strong evidence in support of the hypothesis that larger discrepancies than we have encountered are impossible. Specifically, we constructed 1700 pairs of random correlation matrices for k=2, 700 for k=3 and 700 for k=4, and recorded the maximal discrepancies in percentages of explained variance for r=1. These were 12.0, 8.7, and 6.8 for k=2,3,4, respectively, all falling short of the corresponding maxima for the reversed eigenvalue case, which are 12.5, 10.1 and 8.58, respectively, see (20). It seems safe to infer that the maxima we have derived are by no means limited to the reversed eigenvalue case, and that 12.5% is the maximum discrepancy quite generally.

It should be noted that all results obtained above, the alledged maximum discrepancy of 12.5% included, only pertain to average percentages over groups. This does by no means rule out the possibility of a huge discrepancy in explained variance between PCA and SCA for some specific X_i. For instance, if we have eight groups (m=8) then a discrepancy of 100% for X_1 is still compatible with an average of 12.5%, provided that the discrepancies for $X_2,...X_8$ are zero. However, if the number of groups is small, then each specific discrepancy in also small of necessity. For m=2, for instance, a discrepancy above 25% for X_1 or X_2 is already incompatible with an average discrepancy of 12.5%. From the practitioners point of view, this means that, for m=2, SCA is always rather succesful in terms of variance explained. Moreover, it should be remembered that the maximum discrepancy cases constructed above, such as the 'reversed eigenvalue case', are not likeky to occur in practice. For instance, consider the correlation matrices R_1 and R_2 of (21) once again. These matrices do yield a discrepancy as high as 12.5%, but are not highly realistic. That is, they might conceivably occur in situations where one variable measures membership of the communist party, and the other measures one's evaluation of a one–party system. The correlation between the two is likely to be positive in communist countries, and negative elsewhere. However, such sign reflections of correlations seem to be as exceptional as communist countries are nowadays. They can be spotted by inspection of the variable–components correlations of SCA. In the example under consideration, the simultaneous component is just the first (or the second) variable, with loading vectors $\begin{pmatrix}1\\.5\end{pmatrix}$ and $\begin{pmatrix}1\\-.5\end{pmatrix}$ which reflects heterogeneity of correlations across groups.

In general, it seems wise to inspect correlations between variables and components for each group separately. To demonstrate this even more clearly than we have just done, it is convenient to revisit the extreme k=2 case discussed above with correlations 1 and –1 between the variables in group 1 and 2, respectively. The correlations between the variables are as heterogeneous across groups as they can be. Nevertheless, components defined by **b**=(1 0)' in both groups account for the full 100% of variance, suggesting that the groups have much in common. However, the vectors of variable–component correlations in SCA are here (1 1)' and (1 –1)', respectively. Thus, by inspecting the variable–components correlations, previously ignored anomalies in one's data can easily be detected.

3. SOME RESULTS FROM SIMULATIONS

In addition to exploring the properties of SCA by mathematical means we have conducted a small simulation study. Not having available a variety of data sets to be analyzed with SCA, we have analyzed one particular, well–documented, data set in some detail. Specifically, we used data collected by Van Schuur (1984) that have also been described in Ten Berge (1986, Tables 3–4). The data consist of responses of 718 Italian and 1565 Danish party activists to 15 political measures on a five–point scale, ranging from 1 (very much against) to 5 (very much in favor). These data were treated as data for two populations (Italy and Denmark) from which random samples were drawn by means of SIMLIS (Boomsma, 1983, pp. 24–25). The sample sizes were 50, 75, 100, 125, 150, 175 and 200, respectively. For each sample size 14 pairs of correlations matrices were obtained, one from the Italian and one from the Danish population. Each pair was analyzed with SCA and PCA on X_1 and X_2, using two components throughout. The analyses were used to shed some light on the following questions:

1. What sample size is needed to obtain reliable results from SCA in terms of explained variance and in terms of interpretations of the components?
2. What discrepancies between SCA and PCA in terms of explained variance are to be expected?
3. How much variance explained is lost when the optimal weights in SCA are replaced by simple weights (1,0,–1)?

Each of these question will be addressed in a separate section.

4. SAMPLE SIZE AND SCA

For each of the seven sample sizes, the percentages of variance explained by two SCA components in Italian and Danish samples, averaged over 14 replications, were recorded and compared to the population values, see Table I.

Tab. I Percentages of variance explained by SCA in samples and in the population

	n=50	n=75	n=100	n=125	n=150	n=175	n=200	population
Italy	46.70	46.14	46.31	46.28	46.79	46.60	46.17	45.95
Denmark	44.10	44.67	43.46	43.69	43.73	43.17	43.08	42.60
Total	45.40	45.40	44.89	44.98	45.26	44.88	44.63	44.28

It can be seen from Table I that the explained variance has a small bias upward, which gradually vanishes with increasing sample size. Even in small samples, however, the bias is by no means dramatic.

In addition to percentages of explained variances, the stability of interpretations was also examined. This was done in the following manner. For each SCA solution the optimal weights were rotated according to the Varimax criterion. It has been pointed out above that the amount of variance explained by the components jointly is not affected by this operation. The same procedure was applied to the optimal weights in the population (Italy and Denmark combined). For each resulting component in every sample, the congruence (Tucker, 1951) between the rotated vector of weights and the corresponding vector in the population was evaluated. Average values (over 14 replications) are reported in Table II.

Tab. II Average congruence values between Varimax rotated weights of samples and populations

	n=50	n=75	n=100	n=125	n=150	n=175	n=200
Component 1	.97	.97	.98	.98	.99	.99	.99
Component 2	.94	.95	.96	.97	.98	.98	.98

In order to evaluate the results of Table II properly, it should be noted that congruence is a measure of proportionality, ranging from –1 to +1. A congruence value above .8 indicates strong similarity of the vectors involved. It is clear from Table 2 that all congruence averages are surprisingly high which implies that after Varimax rotation of the weights the interpretations of the population components are very well recovered in samples, even when their size is as small as 50.

In general, it can be concluded that the data under scrutiny reveal stable components, even when the sample size is small. Although there is a bias upward in the explained variance, the interpretations of the components in the samples match those of the populations remarkably well. On the other hand, this finding is likely to be contingent upon the number of components drawn: It is well–known that the major components from PCA tend to be highly stable across studies, whereas subsequent components may easily diverge.

5. THE PRICE OF SCA

Above, we have examined 'the price of SCA' in terms of explained variance at great length by mathematical means. It is also interesting to examine this price for the data under scrutiny, that is, to examine the difference in explained variance between PCA on X_1, X_2 and SCA, for the Italian and Danish subjects. Table III contains these differences, expressed in terms of percentages, for each sample size (averaged over 14 replications), and for the population.

Tab. III Average discrepancies in percentages of explained variance for PCA and SCA

	n=50	n=75	n=100	n=125	n=150	n=175	n=200	population
Italy	.36	.36	.32	.27	.22	.26	.25	.24
Denmark	.65	.59	.52	.35	.34	.36	.31	.26

In the population, the price of SCA is remarkably small. That is, SCA explains only one fourth of a percent less variance than PCA for Italy and Denmark separately. In small samples, the difference is biased upward. This is tantamount to saying that the sampling bias of explained variance for PCA is a little higher than that for SCA. Nevertheless, Table 3 demonstrates that the price of SCA can be close to zero, even in small samples. The present authors have obtained similar results with other data. In fact, that has been the primary motivation for the present study.

6. THE PRICE OF SIMPLE WEIGHTS

Having assessed 'the price of small samples' and 'the price of simultaneous components' above, we now turn to 'the price of simple weights'. That is, we examine how much variance explained by SCA is lost when the weights are 'simplified' to values 1, 0, or –1. Throughout, simple weights are obtained by the simple weights option of the SCA program, after Varimax rotation of the optimal weights. Although, as was mentioned above, the Varimax rotation does not affect the explained variance, replacing weights by simple values does. The question now is, what reduction of explained variance is incurred for the data under scrutiny. Table IV contains the discrepancies in percentages of explained variance between optimal and simple weights, for each sample size (averaged over 14 replications), and for the population.

Tab. IV Percentages of explained variance lost when simpe weights are used

	n=50	n=75	n=100	n=125	n=150	n=175	n=200	population
Italy	.73	.51	.60	.67	.73	.68	.63	.97
Denmark	.96	.70	1.48	1.11	1.23	1.08	1.15	.86

It is clear from Table IV that the loss incurred by using simple weights is quite

small. This is in agreement with the well–known phenomenon that the direction of a linear combination is only slightly affected when the weights of the variables are replaced by other values with tha same sign.

The price of simple weights, as reported in Table IV, was determined on samples. That is, we have reported how much explained variance is lost when the optimal weights in each replication are replaced by simple weights in that sample. However, we might also examine the appropriateness of sample weights used as estimates of the population weights. Accordingly, we have evaluated the percentages of variance explained in the population by the optimal and simple weights from each replication. The average percentages for each sample size and for the population are reported in Table V.

Tab. V Average percentages of variance explained in the population on the basis of optimal and simple weights from samples of size n

	n=50	n=75	n=100	n=125	n=150	n=175	n=200	population
It. optimal	45.27	45.29	45.57	45.61	45.75	45.79	45.79	45.95
It. simple	44.65	44.73	45.15	45.10	45.16	45.13	45.19	44.98
Dk. optimal	42.04	42.25	42.33	42.35	42.36	42.41	42.46	42.60
Dk. simple	41.13	41.42	41.28	41.40	41.28	41.37	41.37	41.74

Table V shows that simple weights from samples are consistently inferior to optimal weights from samples, used in the populations, but the differences are small. They are, in fact, in the same order of magnitude as the differences reported in Table V. It follows that the results of Table 4 are also representative for the situation where simple weights from samples are treated as estimates of simple weights for the populations.

7. DISCUSSION

The theoretical part of this study has shown that it is by no means easy to contrive data for which SCA is significantly less successful than PCA on X_1, X_2. In fact, the maximal discrepancy in explained variance that we have been able to come up with is 12.5%, which still implies that SCA will always yield components explaining a relatively large amount of variance compared to PCA on X_1, X_2. In the example of political measures the discrepancy appeared to be as small as .25 percent. Similar results have been obtained from other data. It seems safe to conclude that SCA should not be thought of as a method to determine whether or not simultaneous components 'exist', in the sense that they explain a considerable amount of variance. Instead, SCA is a method to examine the nature (interpretations, correlations) of the simultaneous components that we know to exist. This conclusion

ties in nicely with the general tendency of PCA to produce generalizable components across populations.

From the practitioners' point of view it is important to establish that simple weights hardly detract from the explained variance. In practice, components are used more often than not to determine which variables can be linearly combined, to measure constructs reliably. A small loss of explained variance will usually be more that compensated for by the gain in simplicity of those linear combinations.

REFERENCES

Boomsma, A., 1983, *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non–normality.* Doctoral dissertation, University of Groningen.

Coppi R., 1986, Analysis of three–way data matrices based on pairwise relation measures. *Proocedings in Computation Statistics,* Edited by F. De Antoni, N. Lauro and A. Rizzi. Physica–Verlag; Vienna; 129–139.

Escoufier, Y., 1980, Exploratory Data Analysis when data are matrices. *Recent development in statistical inference and data analysis,* edited by K. Matusita. North–Holland, Amsterdam, 45–53.

Flury, B.N., 1984, *Common Principal Components in k Groups.* J. Amer. Stat. Association, 79, 892–898.

Kiers, H.A.L., 1988, Comparison of "Anglo–Saxon" and "French" three–mode methods. *Statistique et Analyse des Données,* 13, 14–32.

Kiers, H.A.L. SCA., 1990, *A program for simultaneous components analysis of variables measured in two or more populations.* University of Groningen: IEC Progamma.

Kiers, H.A.L., Ten Berge J.M.F., 1989, *Alternating least squares algorithms for Simultaneous Components Analysis with equal component weight matrices in two or more populations.* Psychometrika, 54, 467–473.

Levin, J., 1966, *Simultaneous factor analysis of several Gramian matrices.* Psychometrika, 31, 413-419.

L'Hermier des Plantes, H., 1976, Structuration des tableaux à trois indices de la statistique, Thése, Université Montpellier II.

Millsap, R.E., Meredith, W., 1988,*Component analysis in cross–sectional and longitudinal data.* Psychometrika, 53, 123–134.

Rizzi, A., Vichi M. Relations between sets of variates of a three–way data set. *Statistique Applicate,* this issue.

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; 262–266.

Tucker, L.R., 1951, *A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984).* Washington, D.C.: Dept. of the Army.

Van Schuur, H., 1984, *Structure in political beliefs. A new model for stochastic unfolding with applications to European party activists.* Doctoral dissertation, University of Groningen.

RIASSUNTO

L'Analisi Simultanea delle Componenti è una generalizzazione dell'Analisi in Componenti Principali nel caso in cui le variabili siano rilevate in due o più popolazioni. In questo tipo di analisi si adopera una matrice comune di pesi, in modo tale che le componenti siano esprimibili nei termini di un'identica combinazione lineare delle stesse variabili. Tale matrice comune di pesi è determinata in modo tale che la varianza spiegata, risultante dalla somma delle varianze spiegate delle singole popolazioni, sia massima. L'articolo in questione tratta di alcune proprietà finora sconosciute delle componenti simultanee. Il risultato principale cui si perviene è che le componenti simultanee risultano capaci di spiegare un notevole ammontare della varianza. Inoltre, si dimostra che né l'uso di piccoli campioni né la sostituzione dei pesi ottimali con pesi elementari inficiano l'utilità delle componenti simultanee.

Statistica Applicata Vol. 4, n. 4, 1992

MULTIPLICATIVE MODELS FOR GENOTYPE–ENVIRONMENT INTERACTION IN PLANT BREEDING

F.A. van Eeuwijk

Centre for Plant Breeding and Reproduction Research (CPRO–DLO), P.O. Box 16, 6700 AA Wageningen, The Netherlands.

In plant breeding genotypes are evaluated under varying environmental circumstances. Genotype–environment interaction is said to occur when genotypes respond differently to changes in the environment. Insight into this phenomenon is essential for progress in plant breeding. Many techniques have been developed for its analysis, usually assuming the data to be arranged in a two–way table of genotypes by environments. In this paper multiplicative models for the interaction are presented, which have not yet found broad application in plant breeding. These models are compared to the popular regression on the environmental mean model.

1. INTRODUCTION

Plant breeding is the science devoted to the development of new varieties. Research is initiated in response to changed demands on the existing varieties, ranging from higher yield to broader resistance to pathogens. To comply with a new standard the genetical constitution of plants is manipulated. The genetical constitution, or genotype, only partly determines the observable characteristics. Properties of organisms always are functions of both genotype and environment. The environment includes everything that is not genetic. Macro–climatological factors are among the most influential environmental factors, and not open to human manipulation. This is the reason why plant breeders emphasize the genetical approach. When genotypes respond differently to a change in the environment the phenomenon of genotype–environment interaction is said to occur. Genotype–environment interaction is of major importance in plant breeding, because its consequence is that the relative merits of genotypes depend on the environmental circumstances. In its most extreme form it can mean that a genotype A has higher yield than a genotype B on location 1, whereas B outyields A on location 2. For each genotype plant breeders have to identify the most influential environmental factors, and determine their mode of action. This must be the basis for decisions on the adaptedness of newly developed genotypes to future commercial growing environments.