By J. F. Affleck-Graves MSc (Cape Town), C. G. Troskie,[1] MSc (Pretoria), PhD (Unisa)
University of Cape Town, A. H. Money, PhD (Cape Town)

A principal component index subject to constraints

ABSTRACT

The application of principal components in the construction of stock market indices has not been enthusiastically received by the investment community. One of the main reasons for this is that a principal component approach often results in the allocation of negative weights to some of the securities. It is shown that by a simple restatement of the problem, this disadvantage can be easily overcome. In addition, extra constraints can be imposed on the weights assigned to the different securities if so desired.

1 INTRODUCTION

The use of principal component analysis in the construction of economic and stock market indices is not new. Theil[2] discussed their use in economic indices, while Feeney and Hester[3] gave a detailed account of how this idea could be applied in the construction of stock market indices. Troskie[4] showed how these ideas could be used on The Johannesburg Stock Exchange, and presented some examples of such indices.

The concept of an index based on principal components is intuitively appealing. If an index is designed to measure movement in the market, then it will be most sensitive (and hence most informative) if the weights are assigned in such a way that the index has a maximum variance over all linear combinations of the stocks to be included in the index. But such a combination is simply the largest principal component, which can be obtained very easily.

Unfortunately, as far as the construction of stock market indices is concerned, the method of principal components has one important disadvantage. Some of the securities may be assigned negative weights which implies a negative holding in these particular companies. This is conceptually unacceptable to the majority of investors who argue that negative weights result in an artificial index which cannot be compared to individual portfolio performance[5].

In this paper it will be shown that by a simple restatement of the principal component problem, this important disadvantage of negative weights can be easily overcome. In addition, there are further benefits to the designer of the stock market index in the form of a facility to impose additional constraints.

2 PRINCIPAL COMPONENT ANALYSIS

In this section a brief summary of the concept of principal components is given – for a detailed mathematical exposition, the reader is referred to Anderson[6].

The problem of principal components can be briefly stated as follows: Given a set of p variables $R(= \begin{bmatrix} R_1 \\ \cdot \\ \cdot \\ \cdot \\ R_p \end{bmatrix})$ with covariance matrix Σ, find the linear combination $a = X'R$ such that the variance of a (i.e. $x'\Sigma x$) is a maximum. To do this, one extracts from Σ its characteristic roots $(\lambda_1, \ldots, \lambda_p)$ and vectors $(h_{(1)}, \ldots, h_{(p)})$; that is, one solves

$$|\Sigma - \lambda_i I| = 0$$

$$(\Sigma - \lambda_i I)h_{(i)} = 0$$

$$\text{subject to } h'_{(i)}\, h_{(i)} = 1$$

$$\text{for } \quad i = 1, 2, \ldots, p$$

The characteristic vector corresponding to the largest characteristic root gives the weights of the linear combination with maximum variance. Moreover, the variance of this linear combination is given by the largest characteristic root.

The technique of principal components has an additional feature. Not only do we have a component which has maximum variance, but it is possible (very simply) to obtain a second component, which has maximum variance subject to the condition that it is uncorrelated (orthogonal) with the first component. The weights of this component are given by the characteristic vector of Σ corresponding to the second largest characteristic root, which also gives the variance of this component. In fact, p such components can be constructed (if Σ is of full rank) – with weights given by the p characteristic vectors of Σ – which have the useful property of being mutually orthogonal.

Now, in general, Σ is not known and must be estimated by

$$S = \sum_{i=1}^{N} (Z - \bar{Z})(Z_i - \bar{Z})'/N-1$$

where Z_i is the vector of the ith observation on each of the p variables, $R_i, \ldots, R_p$. The characteristic roots and vectors extracted from S will be estimates of the corresponding roots and vectors of Σ.

Thus, each estimated characteristic vector and characteristic root pair provides an estimate of a principal component and its associated variance.

3 PRINCIPAL COMPONENT ANALYSIS WITH POSITIVE WEIGHTINGS

The problem of principal components can be stated in terms of a Mathematical Programming Problem as follows:

$$\text{Max } x'\Sigma x$$
$$\text{subject to } \sum_{i=1}^{p} x_i^2 = 1 \qquad (1)$$
$$x_i \text{ unrestricted in sign.}$$

Clearly, if one wishes to restrict the X_i to be positive then one can simply restate the problem as

$$\text{Max } x'\Sigma x$$
$$\text{S.T. } \Sigma x_i^2 = 1 \qquad (2)$$
$$x_i \geq 0 \text{ for } i = 1,2,\ldots,p$$

This then is a quadratic programming problem (since the objective function has terms of the form $x^2_i\delta^2_i$ and $x_ix_j\delta_{ij}$) subject to a quadratic constraint ($\Sigma x^2_i = 1$) and the constraint that the x_i's be positive or zero. This problem can be solved using any of the methods for solving either general nonlinear programming problems (e.g. the Flexible Tolerance Method described in Himmelblau[7], or quadratic programming problems subject to quadratic constraints (e.g. Barron[8]).

Unfortunately, the extension of this idea of positive weightings to the second, third and other components is not possible in general. Clearly, if the vector X is strictly greater than zero (that is, no element of X equals zero) then it is not possible to find another vector Y also strictly greater than zero such that X and Y are orthogonal ($\Sigma x_iy_i = 0$). However, given a first component which is positive, it is possible to find an orthogonal second component if one relaxes the positive restriction on the second component. The problem can then be formulated as follows:

Find the first component (positive) by solving the Quadratic Programming Problem (2) above. Then, solve the following Quadratic Programming Problem:

$$\text{Max } y'\Sigma y$$
$$\text{S.T. } \sum_{i=1}^{p} y_i^2 = 1$$
$$\sum_{i=1}^{p} y_i x_i = 0 \qquad (3)$$
$$y_i \text{ unrestricted in sign.}$$

The additional constraint in the above probem ($\Sigma_i y x_i = 0$) is merely a linear constraint since the weights x_i are known from the solution to (2).

Clearly, this idea can be easily extended to obtain all the other components, provided one is prepared to forsake the nonnegativity constraint on all but the first component. It can be argued that this is reasonable, since it is often hoped that the first component will provide some overall description of the market, while subsequent factors will describe other aspects of the market (e.g. contrasting various sectors etc.). In general then, the r^{th} component ($r \geq 2$) can be found as follows:

$$\text{Max } z'\Sigma z$$
$$\text{S.T. } \sum_{i=1}^{p} z_i^2 = 1$$
$$\sum_{i=1}^{p} w_{ij} z_i = 0 \quad j = 1,2,\ldots,r-1 \qquad (4)$$
$$z_i \text{ unrestricted in sign}$$

where w_{ij} is the weight assigned to the i^{th} variable by the j^{th} component.

In this way, the problem of principal component analysis subject to the restriction that the weights of the first component are all positive or zero, can be solved as shown above, and as many of the remaining components as desired may still be computed.

4 EXAMPLES

In order to illustrate the concepts presented in the previous section, several examples are presented below. All of these examples are based on basically the same problem, namely to construct an index from ten securities quoted on The Johannesburg Stock Exchange; five from the Coal sector and five from the Gold-Witwatersrand and Others sector (Table 1). The covariance matrix was estimated using the weekly closing prices[9] of these ten securities for the period 4 January 1974 to 20 February 1976.

The first principal component was determined using both the traditional approach (hereafter referred to as the P-Component) and the proposed approach which restricts the weights to be nonnegative[10] (hereafter referred to as the G-Component). In addition the second component was found using both methods[11]. The results are presented in Table 1 below.

Table 1

	P-Components		G-Components	
Security	First	Second	First	Second
Apex	–0,1034	0,5006	0,0000	0,4691
Clydesdale	–0,0043	0,0577	0,0001	0,0375
Tavistock	–0,2045	0,8084	0,0000	0,8620
Trans Natal	–0,0010	0,0629	0,0004	0,1032
Welgedacht	–0,0203	0,1208	0,0027	0,1391
Durban Deep	0,6784	0,1389	0,6918	0,0234
ERPM	0,5871	0,1610	0,6137	–0,0155
Grootvlei	0,1277	0,0430	0,1279	–0,0145
Marievale	0,1801	–0,0906	0,1768	–0,0640
S.A. Lands	0,3058	0,1370	0,3117	0,0193
Variance	1 641 570	50 670	1 556 792	133 051
% variation exp	94,61%	2,92%	89,73%	7,67%

The above example illustrates the manner in which a principal component analysis can be performed subject to the additional constraint that the first component be nonnegative. It is interesting to note that the first G-Component explains only 5% less of the variation than the first P-Component (which will obviously always explain more). Moreover, when the first two components are combined the two methods describe almost exactly the same percentage of the variation (97,53% compared to 97,40%)[12].

One of the main advantages of the Flexible Tolerance Method is that additional constraints can be easily included in the analysis. For example, in the above 10 security problem, the constraint that no security in the first component be given a weight of more than 0,5 can be included by formulating the problem as follows.

$$\begin{aligned} &\text{Max } x' Sx \\ &\text{S.T. } \sum_{i=1}^{10} x_i^2 = 1 \qquad (5) \\ &\quad 0 \leq x_i \leq 0{,}5 \quad i = 1,2,\ldots,10 \end{aligned}$$

Alternatively it might be desirable that the two groups of securities (Coal and Gold) have equal weight. For illustrative purposes, it could be required that the sum of the squares of the weights assigned to the Coal securities equal the sum of the squares of the weights assigned to the Gold securities. The formulation then becomes:

$$\begin{aligned} &\text{Max } x' Sx \\ &\text{S.T. } \Sigma x_i^2 = 1 \\ &\quad \sum_{i=1}^{5} x_i^2 - \sum_{j=6}^{10} x_j^2 = 0 \qquad (6) \\ &\quad x_i \geq 0 \quad i = 1,2,\ldots,10 \end{aligned}$$

Finally, it might be desired to combine the above two cases, which leads to the following formulation:

$$\begin{aligned} &\text{Max } x' Sx \\ &\text{S.T. } \Sigma x_i^2 = 1 \qquad (7) \\ &\quad \sum_{i=1}^{5} x_i^2 - \sum_{j=6}^{10} x_j^2 = 0 \\ &\quad 0 \leq x_i = 0{,}5 \quad i = 1,2,\ldots,10 \end{aligned}$$

The first component was found for each of the above three problems, using the Flexible Tolerance Method and the results are presented in Table 2 below.

Table 2

Security	Problem (5)	Problem (6)	Problem (7)
Apex	0,0000	0,0000	0,0002
Clydesdale	0,0003	0,6758	0,5000
Tavistock	0,0000	0,0000	0,0000
Trans Natal	0,0044	0,2080	0,5000
Welgedacht	0,0076	0,0003	0,0023
Durban Deep	0,5000	0,4966	0,4807
ERPM	0,4998	0,4293	0,4293
Grootvlei	0,3093	0,0811	0,1151
Marievale	0,3939	0,1277	0,1513
S.A. Lands	0,4994	0,2148	0,2202
Variance	1 321 490	772 306	772 075
% variation exp	76,16%	44,51%	44,50%

It should be noted that the example presented in this section indicates that when no additional constraints were imposed there was a difference of approximately 5% in the variance of the first P-Component and the first G-Component. From experience it appears that this is often approximately the order of the difference when only the nonnegative weights restriction is imposed. In addition it is obvious from mathematical or geometric considerations that the second G-Component will always have variance at least as large as the second P-Component, and thus the percentage variation explained by the first two P-Components will usually be close to that explained by the first two G-Components.

However, when additional constraints are imposed it is not easy to predict the effect on the G-component as far as the percentage variation explained is concerned. It is also not easy to assess the relative performance of the first two components (compared with the first two P-Components) as this will be very dependent on the nature of the constraints imposed and whether in fact they are imposed on both components or only the first.

It is also interesting to note that the above technique can be used if it is required to obtain a traditional first component (i.e. weights unrestricted in sign) subject to the imposing of additional constraints. For example, if an unrestricted in sign first component is desired subject to the condition that no security have weight more than 0,5 the problem can be formulated as:

$$\begin{aligned} &\text{Max } x' Sx \\ &\text{S.T. } \Sigma x_i^2 = 1 \\ &\quad -0{,}5 \leq x_i \leq 0{,}5 \quad i = 1,2,\ldots N \end{aligned}$$

5 PRACTICAL EXAMPLE

Since the Flexible Tolerance Method is an iterative search procedure, Himmelblau (1972) recommends that the analysis be repeated using several different starting solutions. As this results in somewhat more computation than the traditional principal component analysis it is suggested that in practice the traditional first principal component be found. If it is nonnegative and satisfies all of the additional restrictions imposed then these weights should be used. However if the constraints are not satisfied then the Flexible Tolerance Method should be used to find the index which is the most volatile subject to the imposed constraints. In practice this will often be an iterative procedure as is shown in the following example.

A principal component type index was constructed for the Chemical sector of the JSE using four securities quoted in this sector. The covariance matrix was estimated using weekly closing prices for the period 5 December 1969 to 29 December 1972 and the first principal component derived from this matrix is listed in the second column of Table 3 below:

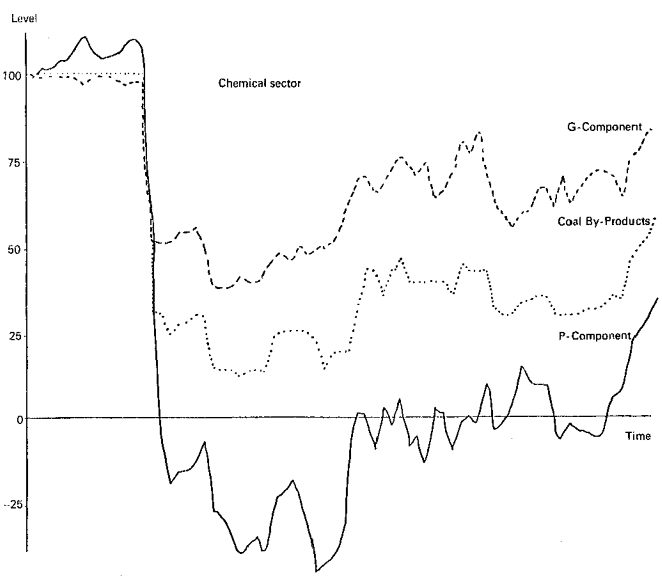
Table 3

Security	P-Component	G-Component	Constrained G-Component
AECI	–0,1459	0,0000	0,1414
Coal By-Products	0,9330	0,9994	0,7000
Fedmis	–0,3278	0,0000	0,0000
Sentrachem	0,0267	0,0346	0,7000

From Table 3 it can be seen that two of the weights of the P-Component are negative. The graph of the index based on these weights (initialised to 100) is presented in Figure 1 below for the period 5 December 1969 to 20 February 1976. As can be seen the index "goes negative" from April 1971 until February 1973, after which it oscillates between positive and negative weights until August 1975. Clearly, this is a most unsatisfactory situation.

In this paper it has been suggested that if the weights of the first principal component are negative, then the flexible tolerance method (or some alternative method) should be used to restrict all the weights to nonnegative values. Doing this, the weights listed in the third column of Table 3 (G-Component) were obtained.

Figure 1



On examining Table 3 it is obvious that an index based on the weights of the G-Component would consist, for all practical purposes, of a single security – Coal By-Products. This is almost certainly an unsatisfactory situation, especially as Coal By-Products is a relatively unstable security – historically it is liable to sudden rather violent price movements (see Figure 1). Thus, it is apparent that an upper bound should be imposed on the weight assigned to any one security. Only four securities were considered in this example, and the sum of the squares of the weights must equal one. Thus, it might be desirable that the maximum weight allocated to any one security be limited to 0,7. Including these additional constraints ($X_i \leq 0,7$ for i = 1;2;3; and 4) the problem was resolved and the weights obtained are presented in the last column (Constrained G-Component) of Table 3. The graph of an index based on these weights is presented in Figure 1.

As can be seen from Figure 1, both of the indices using the Flexible Tolerance Method avoid the problem of "going negative". In addition, the restricted problem ($X_i \leq 0,7$) is less affected than the other indices by the sudden drop in Coal By-Products in April 1971.

6 CONCLUSIONS

In this paper, the problem of principal component analysis, when it is desired to impose various additional constraints on the weights, has been examined. A method of solving such problems has been indicated, and various practical examples presented. Such a technique, it is felt, will have many applications in analyses where principal components are used, as the method gives great flexibility to the researcher. It must be emphasised that the technique is not limited to use in stock market problems. However, the latter do lend themselves very well to such applications, and this technique could prove most useful in this field of research.

It is proposed that the following procedure be adopted when constructing stock market indices based on principal components: initially, the first principal component should be found. If all the weights are nonnegative and any other necessary constraints are met, then these weights should be used to construct the index. If, however, the weights are not all non-negative, or if some additional required constraints are not met, then the flexible tolerance method (or an alternative method) should be used, and will result in an index which is the most volatile subject to the imposed constraints.

References

1 The authors wish to thank the Council for Scientific and Industrial Research for their financial support.

2 Theil, H. Best Linear Index Numbers of Prices and Quantities. *Econometrica,* 28 (1960), 464-480.

3 Feeney, George J. and Donald D. Hester. Stock Market Indices: A Principal Components Analysis. In Hester, D. D. and Tobin, J.: *Risk Aversion and Portfolio Choice.* New York, Wiley, 1967.

4 Troskie, C. G. Principal Components and its Applications. Unpublished paper read at the Conference of the S.A. Statistical Association, October 1970.

5 Although a negative weighting can be considered as a short holding in the particular security, it remains an unacceptable concept to a large body of investors.

6 Anderson, T. W. *An Introduction to Multivariate Statistical Analysis.* New York, Wiley, 1958.

7 Himmelblau, D. M. *Applied Nonlinear Programming.* New York. McGraw-Hill, 1972.

8 Barron, David, P. Quadratic Programming with Quadratic Constraints, *Naval Research Logistics Quarterly,* 21 (1972), 253-260.

9 It should be noted that there is no unanimity on whether stock market indices should be based on price or return. The examples presented in this section deal with indices based on price but the methods are equally applicable if an index based on return is preferred.

10 The Flexible Tolerance Method (Himmelblau (1972)) was used to determine the G-Components. This method was used as the algorithm was readily available on the computer. It is not suggested that this is the most suitable method for use in general but it does appear from numerous empirical examples performed by the authors that the method provides satisfactory results for the stock market indices problem.

11 Although the second G-Component must be allowed to have negative weights it will not, in general, be identical to the second P-Component as both must be orthogonal to their respective first components.

12 Clearly the total percentage variation of the first two P-Components must be greater than or equal to that of the first two G-Components. It is not suggested that the totals will always be as close as indicated in this example.