South African Statist. J. (1987) 21, *155-171*

THE WEIGHTED COVARIANCE BIPLOT - AN APPLICATION

G.D.I. Barr: Department of Mathematical Statistics, University of Cape Town, Private Bag, Rondebosch, 7700 Republic of South Africa.

B.S. Kantor: Department of Economics, University of Cape Town, Private Bag, Rondebosch, 7700 Republic of South Africa.

L.G. Underhill: Department of Mathematical Statistics, University of Cape Town, Private Bag, Rondebosch, 7700 Republic of South Africa.

Key words: Covariance biplot; graphical display; price variability; relative prices; weighted covariance biplot.

Summary: This paper demonstrates the application of the covariance biplot technique to the analysis of relative price movements in South Africa over the period 1960-1983. In this application it is appropriate to use weighted values of the data and the paper provides a theoretical analysis of the covariance biplot procedure for the weighted case. It is demonstrated how the biplot analysis of relative price movements gives far greater insights into relative price behaviour and variability than previous approaches.

---

1. INTRODUCTION

The term *rate of inflation* is widely applied in the analysis of economic performance. Although it is applied to individual items the term is more generally used to refer to the change in price of a weighted average of a selected set or basket of goods and services. The weights used will vary according to the relative

STMA 1987 SUBJECT CLASSIFICATION: 06:110(14:070)

importance of the items, for example they may be proportional to the consumption expenditure on each item or the production value of each item. Thus the rate of inflation as measured by changes in the consumer price index is based on a different set of items and weightings to that measured by changes in the production price index.

An important issue that has received less attention is the measurement of changes in the relative prices of items within the basket. Increased variability in relative prices, in iteself, leads to greater uncertainty for both producers and consumers in the determination of almost all aspects of their economic activity. For example, a change in the price of a primary input *relative* to the price of the finished product has more impact on the producer than an equal change in the price of both items. This paper is concerned with the measurement and assessment of relative price changes.

In order to examine the variability of relative prices for some set of n goods the price indices P_{it} for goods $i = 1,\ldots,n$ are recorded for m time points $t = 1,\ldots,m$ (usually years). In this type of application it would be usual for m to be greater than n. In the analysis below we will make this assumption.

An increment of the price index of good i over year t is defined. In the study below the logarithmic increment is used so that the resulting data can be interpreted as inflation rates. Thus we may define:

$$x_{it} = \log P_{it} - \log P_{i,t-1}.$$

x_{it} is then the rate of price inflation for good i over year t.

The average inflation rate over the set of goods considered at some time point t is the weighted average

$$\bar{x}_{\cdot t} = \sum_{i=1}^{n} w_i x_{it}$$

where the weights w_i, $i = 1,\ldots,n$ which sum up to 1, are the same for each year and equal the average expenditure share on the i-th good over the m years.

The matrix $X = [x_{it}]$ contains all the information about price variation:

- each column captures the variation in the rate of inflation over the goods at a given time,
- each row captures the variation in the rate of inflation of a particular good over time.

In order to measure relative price variability previous authors (Parks 1978, Theil 1967, Bleijer 1981, 1983 and Cuikerman 1983) have focussed on the scalar measure

$$VP_t = \sum_{i=1}^{n} w_i (x_{it} - \bar{x}_{\cdot t})^2 .$$

This formula, however, captures only a small portion of the relative price information available in matrix X. In particular, measures such as VP_t give no indication of how relative price variability is constituted at a given point in time. Quite different columns of X could give rise to the same value of VP_t. More information can be obtained by considering the matrix X in some higher dimension.

In this paper we present a useful graphical display technique for viewing the movement of relative prices through time. The biplot procedure used here requires a generalization of the standard method (Gabriel 1971, 1972, 1981, Greenacre and Underhill 1982, Greenacre 1984: Appendix A, Gower 1984) in which the observations are weighted. Greenacre (1984, p. 348) has commented that the usefulness of the biplot on weighted data has not been explored. In this application the use of weights is both natural and necessary for the purposes of interpretation.

The biplot of relative price movements will capture a large portion of the interpretable information on relative price contained in matrix X in a two dimensional graphical display. In particular one can track the movement of the relative price vector through time in terms of its across-category variance VP_t, its correlation across years and the magnitudes of its components (relative to the average inflation rate). This allows one to identify which price categories are giving rise to changes in the variability of relative prices at any point in time.

The weighted covariance biplot

The covariance biplot represents an $n \times m$ data matrix as two clouds of points, n points representing the rows and m points representing the columns. For visual inspection one projects these points onto a low dimensional subspace in such a way that as much as possible of the variability of the data matrix is preserved.

Summary statistics of the X matrix when weights are applied

The matrix X has n rows which correspond to price categories (of goods) and m columns which correspond to time points (with $m > n$).

Let $\Sigma = [\sigma_{tt'}], t,t' = 1,2,\ldots,m$ be the weighted variance-covariance matrix defined by

$$\Sigma = (X - 1w^T X)^T \Omega (X - 1w^T X)$$

where $w = (w_1,w_2,\dots,w_n)^T$ and $\Omega = \mathrm{diag}(w_1,w_2,\dots,w_n)$.

Apart from the weighted averages of the columns of X the quantities of interest for the purposes of interpretation are contained in the matrix Σ. We now list these quantities and later show how they can be obtained from the biplot.

(i) The weighted average of column t

(1.1) $$\bar{x}_{\cdot t} = \Sigma_i w_i x_{it}.$$

This equals the weighted average of the price changes at time point t, i.e. the inflation rate at time point t.

(ii) The weighted variance of column t

(1.2) $$\sigma_{tt} = \Sigma_i w_i (x_{it}-\bar{x}_{\cdot t})^2.$$

This value equals VP_t, the usual measure of relative price variability at time point t.

(iii) The weighted covariance between columns t and t'

(1.3) $$\sigma_{tt'} = \Sigma_i w_i (x_{it}-\bar{x}_{\cdot t})(x_{it'}-\bar{x}_{\cdot t'}).$$

(iv) The weighted correlation between columns t and t'

(1.4) $$\rho_{tt'} = \frac{\sigma_{tt'}}{\sqrt{\sigma_{tt}\sigma_{t't'}}}.$$

This correlation coefficient adjusts for the fact that some categories are relatively more important in a basket of goods and services than others. This quantity will thus give an indication of the closeness of the relative price constitution at two time points adjusted for the different weightings of the categories.

For completeness we mention two more quantities which are easy to extract from the biplot although in this application they have limited economic interpretation.

(v) The weighted variance of the difference between columns t and t' equals

(1.5) $$\Sigma_i w_i[(x_{it}-\bar{x}_{.t}) - (x_{it'}-\bar{x}_{.t'})]^2.$$

(vi) If we let $Y^{(i)}$ be the i-th row of $Y = X - 1w^TX$ and Σ^{-1} is the Moore-Penrose generalised inverse of Σ then the Mahalanobis distance between two rows $Y^{(i)}$ and $Y^{(i')}$ is

(1.6) $$(Y^{(i)} - Y^{(i')})\Sigma^{-1}(Y^{(i)} - Y^{(i')})^T.$$

The weighted SVD and the perfect biplot

We now show how the quantities of interest listed in (ii)-(vi) above are obtained from the biplot.

The singular value decomposition of $Z = \Omega^{\frac{1}{2}}Y$ is of the form

$$Z = UD_\alpha V^t \quad \text{where } D_\alpha = \text{diag}(\alpha_1,\ldots,\alpha_r),\ r = \text{rank}(Z),$$
$$\alpha_1 \geq \ldots \geq \alpha_r > 0$$
$$U^tU = V^tV = I_r,$$

where r equals n-1 (since column means have been subtracted and $m > n$).

This leads to the decomposition

(2.1) $$Y = FG^T$$

where

$$F_{(n\times r)} = \Omega^{-\frac{1}{2}}U$$
$$G_{(m\times r)} = VD_\alpha.$$

The *perfect* biplot given by (2.1) is the set of $n + m$ points in r-dimensional Euclidean space comprising

n row points, f_i (each is a row of F)

m column points, g_t (each is a row of G).

The geometrical interpretation of these points is in terms of the distances of each point from the origin and the cosines of the angles which pairs of points subtend at the origin.

(i) Each element of Y, y_{it} equals the inner product of f_i and g_t i.e. the product of the distance of f_i and g_t from the origin and the cosine of the angle subtended by f_i and g_t at the origin,

i.e. $y_{it} = f_i \cdot g_t \equiv \|f_i\|\|g_t\|\cos\theta_{it}$.

Thus the point f_i represents the vector of across-time inflation rates in category i and the point g_t represents the vector of across-category inflation rates at time t. Thus the inner product of f_i and g_t gives an indication of how close the inflation rate of price category i was to the weighted mean of column t (the general inflation rate at time t). One corollary to this is that if the inflation rates of two price categories show similar behaviour through time they are plotted close together. In addition note that the sum of the projections of all the g_t (year points) on a particular f_i will be zero.

(ii) (a) It follows from (2.1) that

$$GG^T = VD_\alpha D_\alpha V^T = Z^T Z$$
$$= \Omega^{½}(X-1w^TX)^T(X-1w^TX)\Omega^{½}.$$

Therefore $\|g_t\|^2 = \Sigma_i w_i(x_{it}-\bar{x}_{.t})^2$. Thus the squared distance of g_t from the origin (or squared length of

g_t) is equal to the weighted variance σ_{tt} of column t of X (cf. (1.2)).

(b) Furthermore

$$g_t \bullet g_{t'} = \Sigma_i w_i (x_{it} - \bar{x}_{.t})(x_{it'} - \bar{x}_{.t'}).$$

Thus the inner product of g_t and $g_{t'}$ equals the weighted covariance $\sigma_{tt'}$ between columns t and t' of X (cf. (1.3)).

(c) The cosine of the angle between g_t and $g_{t'}$,

$$\cos(\theta_{tt'}) = \frac{g_t \bullet g_{t'}}{\|g_t\| \|g_{t'}\|} .$$

$$= \frac{\Sigma_i w_i (x_{it} - \bar{x}_{.t})(x_{it'} - \bar{x}_{.t'})}{\sqrt{\Sigma_i w_i (x_{it} - \bar{x}_{.t})^2 \Sigma_i w_i (x_{it'} - \bar{x}_{.t'})^2}} .$$

Thus the angle between g_t and $g_{t'}$ on the biplot equals the weighted correlation between columns t and t' of X (cf. (1.4)).

Thus if the relative price structure at two points is highly correlated, g_t and $g_{t'}$ will lie in the same direction from the origin, if they are negatively correlated they will lie on opposite sides of the origin and if the correlation between the two variables is close to zero, they will tend to lie at right angles to each other.

(d) Denoting the t-th column of Z by $Z_{(t)}$ we have

$$\|g_t - g_{t'}\|^2 = (Z_{(t)} - Z_{(t')})^T (Z_{(t)} - Z_{(t')})$$

$$= \Sigma_i w_i [(x_{it} - \bar{x}_{.t}) - (x_{it'} - \bar{x}_{.t'})]^2 .$$

Thus the square of the distance between g_t and $g_{t'}$ is equal to the weighted variance of the difference between columns t and t' of X (cf. (1.5)).

(e) From (2.1) it follows that

$$F = \Omega^{-1/2}U = \Omega^{-1/2}ZVD_\alpha^{-1}$$

and consequently

$$FF^t = \Omega^{-1/2}Z\Sigma^{-1}Z^T\Omega^{-1/2}.$$

Thus

$$f_i f_{i'}^T = Y^{(i)}\Sigma^{-1}Y^{(i')T}$$

where $Y^{(i)}$ denotes the i-th row of Y. Therefore the distance between f_i and $f_{i'}$ is

$$\|f_i - f_{i'}\|^2 = (Y^{(i)} - Y^{(i')})\Sigma^{-1}(Y^{(i)} - Y^{(i')})^T.$$

which is the Mahalanobis distance between $Y^{(i)}$ and $Y^{(i')}$ (cf. (1.6)).

The approximate biplot

For the practical implementation of the biplot one considers an approximation of (2.1), where the first p columns of F and the first p columns of G are retained. This gives an approximation of Y, and a corresponding approximate biplot of $n + m$ points in p-dimensional space.

These first p columns correspond to the p largest singular values $\alpha_k, (k=1,\dots,p)$ and the quality of the approximation is measured by

$$q_p = 100 \sum_{k=1}^{p} \alpha_k^2 / \sum_{k=1}^{r} \alpha_k^2.$$

For a two dimensional plot we take $p=2$. All the properties of the perfect biplot described above hold approximately here.

2. APPLICATION OF THE COVARIANCE BIPLOT TO RELATIVE PRICE MOVEMENTS

The covariance biplot procedure was applied to the analysis of relative price movements for the data described above. In this application the various price categories are treated as the rows and the time points as the columns. The weights (w_i) applied in the analysis are those of the consumption expenditure shares of the various price categories in the analysis adjusted so as to sum to unity (table I). When these weights are incorporated into the analysis the column means are the weighted means of the x_{it} namely $\bar{x}_{.t}$ (the general inflation rate).

The quality of the biplot in two dimensions is nearly 90% (table III) indicating that the distortion in the two-dimensional display should be acceptably small. Nevertheless, each interpretation made from the display needs to be confirmed by inspection of the original data matrix.

Considering first the column (time) points in figure 1, note that their length approximates the weighted standard deviation of that column and thus has a direct interpretation since it is proportional to the square root of VP_t. These distances are in general agreement with the computed VP_t's of table I. The cosines of the angles between time points in figure 1 approximate the correlation between the relative price structures at different points in time. The two-dimensional display thus gives a simultaneous representation of the degree of relative price

TABLE I. x_{it} for five subcategories of goods and services for South Africa, $x_{.t}$ and VP_t for period 1960–1983 inclusive. The time interval over which x_{it} is computed is 12 months.

Year	Weight	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
Housing	0,2982	3,24	3,45	2,03	2,29	4,05	5,69	3,51	5,12	4,56	6,30	5,16	6,35
Food	0,3899	1,43	-0,41	3,01	-0,89	8,78	3,68	2,78	1,87	3,63	0,73	5,94	5,03
Clothing/ Footwear	0,0917	0,93	1,23	0,91	1,10	3,42	2,09	1,68	0,65	0,00	6,45	2,89	6,33
Furniture/ Equipment	0,1346	-0,30	0,20	-0,79	-0,60	1,00	1,28	0,88	0,10	1,54	0,38	2,07	3,21
Motor vehicles	0,0856	0,43	2,97	2,58	2,51	0,49	1,57	4,74	4,35	-0,54	5,64	2,00	8,32
$x_{.t}$		1,61	1,26	1,98	0,57	5,12	3,63	2,81	2,70	2,94	3,29	4,57	5,58
VP_t		1,46	2,92	1,59	2,25	9,97	2,60	1,09	3,56	3,06	7,80	2,50	1,75

Year	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
Housing	7,89	6,10	11,53	11,79	9,63	7,04	8,50	11,83	9,10	16,03	15,87	16,23
Food	10,86	12,66	18,95	6,48	8,72	11,32	12,58	13,34	24,99	10,66	10,45	9,95
Clothing/ Footwear	6,21	6,63	14,07	7,81	10,61	9,13	9,71	5,70	13,41	12,80	14,15	7,76
Furniture/ Equipment	6,38	9,84	14,96	7,92	9,57	8,26	10,19	6,03	9,60	14,70	12,01	7,65
Motor vehicles	9,33	2,70	9,39	15,34	15,78	10,58	11,53	8,51	13,78	14,04	10,48	10,69
$x_{.t}$	8,81	8,92	14,93	9,14	9,88	9,37	10,69	10,80	16,16	13,29	12,62	11,38
VP_t	3,33	11,73	12,43	8,52	3,59	3,41	3,00	8,72	52,29	5,28	5,65	10,92

TABLE II. P_{it} for five subcategories of goods and services for South Africa. Observations are for the months of January from 1960 to 1984. Each index has been scaled so that the average price for 1960 = 100.
Source: South African Reserve Bank Quarterly Bulletins.

Year	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
Housing	91,0	94,0	97,3	99,3	101,6	105,8	112,0	116,0	122,1	127,8	136,1	143,3	152,7
Food	97,1	98,5	98,1	101,1	100,2	109,4	113,5	116,7	118,9	123,3	124,2	131,8	138,6
Clothing/ Footwear	96,4	97,3	98,5	99,4	100,5	104.0	106,2	108,0	108,7	108,7	116,0	119,4	127,2
Furniture/ Equipment	101,4	101,1	101,3	100,5	99,9	100,9	102,2	103,1	103,2	104,8	105,2	107,4	110,9
Motor vehicles	92,6	93,0	95,8	98,3	100,8	101,3	102.9	107,9	112,7	112,1	118,6	121,0	131,5

Year	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
Housing	165,2	175,6	197,1	221,8	244,2	262,0	285,2	321,0	351,6	412,7	483,7	568,9
Food	154,5	175,4	211,9	226.1	246,7	276,3	313,3	358,0	459,7	511,4	567.7	627,1
Clothing/ Footwear	135,4	144,6	166,5	180,0	200,2	219,3	241,7	255,9	292,6	332,5	383,1	414,0
Furniture/ Equipment	118,2	130,4	151,5	164,0	180,4	196,0	217,0	230,5	253,7	293,9	331,4	357,7
Motor vehicles	144,4	148,3	162,9	189.9	222,4	247,2	277.4	302,1	346,7	398,9	443,0	493,0

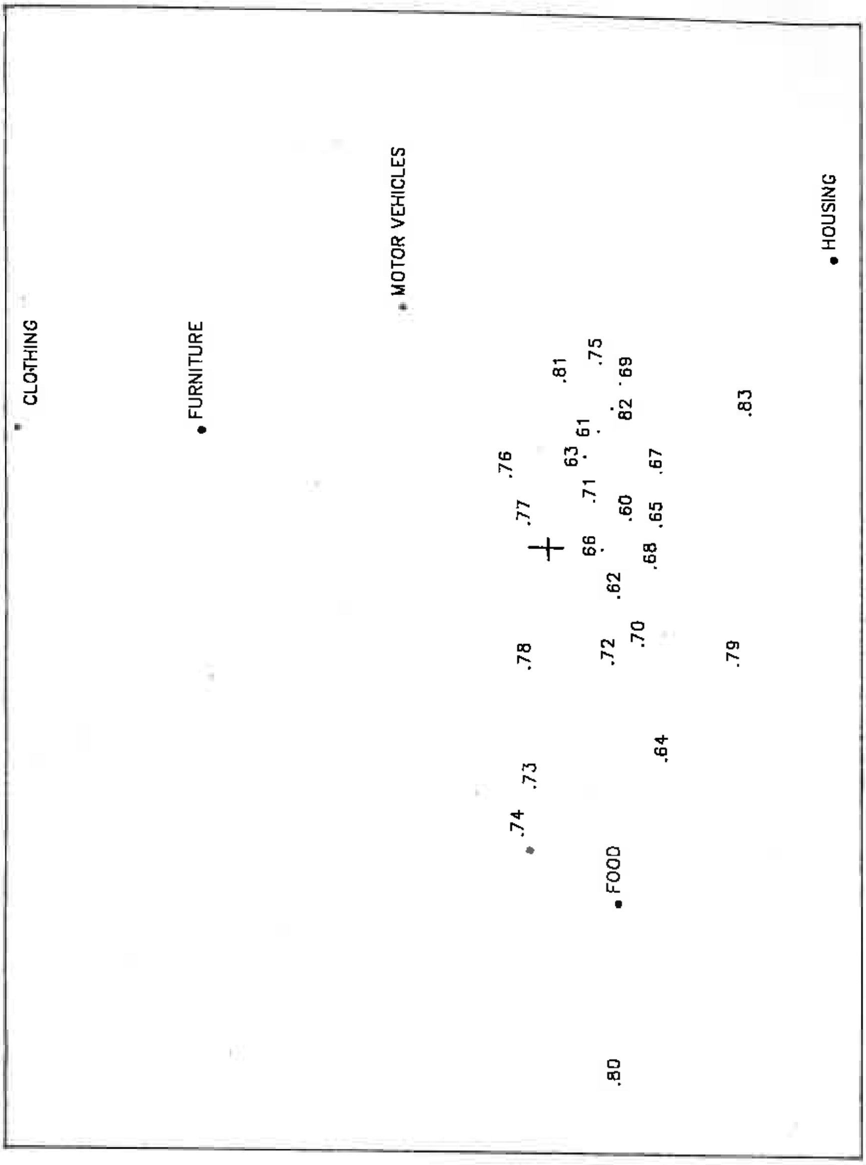
CLOTHING
FURNITURE
MOTOR VEHICLES
HOUSING
FOOD
.81
.75
.69
.82
.61
.63
.83
.76
.71
.67
.77
.60
.65
.66
.68
.62
.78
.72
.70
.79
.64
.73
.74
.80

FIGURE 1

variability and the closeness of the constitution of this variability at different points in time. Thus points for 1983 and 1973 have similar levels of relative price variability but very different compositions of that variability.

Turning to the interpretation of the row points, the categories of consumer prices, we see that the categories furniture, clothing and motor vehicles, which are close to each other, behaved similarly over time. The position of food and housing indicates the very different behaviour of food and housing prices over time *vis-a-vis* each other and the group comprising furniture, clothing and motor vehicles.

What is of particular interest in this application is that the biplot interpretation enables one to consider the price category sources of the VP_t value and in which direction they act. Thus, for example, the year 1980 lies in the same direction as food but subtends an angle of greater than 180 degrees with the other four price categories. This implies that food price changes were above the average price change in 1980 and that all other categories were below the average price for 1980. The configuration of the five price categories relative to the point representing the year 1980, in fact indicates that the large value of VP_t was due primarily to the relatively large increase in the food price in 1980. Price variability in 1983 was, in contrast, high because of the relatively large price changes of housing and the relatively low price changes of furniture and clothing. In this case it is seen that price changes in food and motor vehicles were close to the average and made a small contribution to relative price variability in 1983.

TABLE III: The quality of the biplot approximation for increasing p.

p	q_p(%)
1	*69,751*
2	*88,612*
3	*97,780*
4	*100,000*

CONCLUSION

In this paper we have demonstrated an extension of the covariance biplot to an example where it is natural to consider weighted values of the data both for average value and variance computations. It is seen how the technique allows a number of data features to be displayed simultaneously within a single plot and thus represents a compact and powerful tool for exploratory data analysis.

ACKNOWLEDGEMENT

G.D.I. Barr and B.S. Kantor acknowledge support from the Human Sciences Research Council.

L.G. Underhill acknowledges support from the Foundation for Research Development of the Council for Scientific and Industrial Research.

REFERENCES

BLEIJER, M.I. (1981). The dispersion of relative commodity prices under very rapid inflation. *Journal of Development Economics*, vol. 9, pp. 347-356.

——— (1983). On the anatomy of inflation. The variability of relative commodity prices in Argentina. *Journal of Money, Credit and Banking*, vol. 15, no. 4, pp. 469-482.

CUIKERMAN, A. (1983). Relative price variability and inflation: A survey and further results. *Carnegie Rochester Conference Series on Public Policy*, 19, (1983), North Holland, pp. 103-158.

GABRIEL, K.R. (1971). The biplot-graphic displace of matrices with application to principal component analysis. *Biometrika*, 58, pp. 453-467.

GABRIEL, K.R. (1972). Analysis of meteorological data by means of canonical decomposition and biplots. *Journal of Applied Meteorology*, 11, pp. 1071-1077.

GABRIEL, K.R. (1981). Biplot displays of multivariate matrices for inspection of data and diagnosis. In Barnett, V. (ed.), *Interpreting multivariate data* (147-174). Wiley, Chichester, U.K.

GOWER, JOHN C. (1984). Multivariate analysis: ordination, multi-dimensional scaling and allied topics, in *Handbook of applicable mathematics*, (Chief Ed.) W. Lederman, vol. VI, Statistics, Part B, (Ed.) E. Lloyd, Wiley, Chichester.

GREENACRE, M.J. (1984). *The theory and applications of correspondence analysis*. London, Academic Press.

GREENACRE, M.J. & UNDERHILL, L.G. (1982). Scaling a data matrix in a low dimensional Euclidean space. In *Topics in Applied Multivariate Analysis*. D. Hawkins (ed.), Cambridge: Cambridge University Press.

PARKS, R.W. (1978). Inflation and relative price variability. *Journal of Political Economy*, vol. 86, pp. 79-96.

THEIL, H. (1967). *Economics and information theory*. Chicago: Rand McNally.

Manuscript received 1986.10.30, first revision 1987.04.15, second revision 1987.06.15.