

## **Biplots in Reduced-Rank Regression**

CAJO J. F. TER BRAAK

DLO-Agricultural Mathematics Group and  
DLO-Institute for Forestry and Nature Research  
The Netherlands

CASPAR W. N. LOOMAN

Department of Public Health and Social Medicine  
Erasmus University  
The Netherlands

### *Summary*

Regression problems with a number of related response variables are typically analyzed by separate multiple regressions. This paper shows how these regressions can be visualized jointly in a biplot based on reduced-rank regression. Reduced-rank regression combines multiple regression and principal components analysis and can therefore be carried out with standard statistical packages. The proposed biplot highlights the major aspects of the regressions by displaying the least-squares approximation of fitted values, regression coefficients and associated *t*-ratios. The utility and interpretation of the reduced-rank regression biplot is demonstrated with an example using public health data that were previously analyzed by separate multiple regressions.

*Key words:* Biplot; Canonical correlation analysis; Multiple regression; Multivariate analysis; Public health data; Reduced-rank regression; Redundancy analysis; Regression coefficient; *t*-ratio.

### 1. Introduction

In comparison with multiple regression, multivariate regression is rarely used by applied statisticians. When a number of response variates is of interest, each response variate is usually analyzed in a separate multiple regression. This practice is justified by the Gauss-Markoff setup of regression theory. In this setup, estimation in multivariate regression reduces to a series of multiple regressions (e.g. RAO, 1973: section 8c2). This holds true also for maximum likelihood estimation (e.g. MARDIA et al., 1979: section 6.2). Things change if a restriction is imposed on the rank of the matrix of regression coefficients (ANDERSON, 1951, 1984). This yields a more parsimonious model in which response variates react to the regressor variables only through a restricted number of 'latent variables'. The latent variables can be estimated by canonical

variates. IZENMAN (1975) introduced the apt name reduced-rank regression. Despite further work by TSO (1981), DAVIES & TSO (1982), among others, and, more recently, by LEFKOVITCH (1986), ISRAËLS (1987), VELU et al. (1986) and VAN DER LEEDEN (1990), reduced-rank regression has found few applications. This is unfortunate. In our view, reduced-rank regression is particularly useful, because it offers the possibility to produce a plot that easily summarises the major aspects of the regressions. This possibility has so far not been exploited. In this paper we show how the biplot graphic display (GABRIEL, 1971; 1982) can help to visualize the reduced-rank model. The proposed biplot represents geometrically the fitted values of the regressions, the estimated regression coefficients and their associated *t*-ratios. The two-dimensional biplot is exact for the rank 2 model. For higher ranks, it forms a least-squares approximation. We demonstrate the utility of the reduced-rank regression biplot with an example using public health data that has previously been analyzed by separate multiple regressions (KUNST et al., 1990).

Reduced-rank regression can be carried out by standard computer programmes for multivariate analysis. There exist essentially two different methods of estimation. Depending on the assumptions about the errors, parameter estimates can be obtained from a canonical correlation analysis (ANDERSON, 1951; TSO, 1981) or from a principal component analysis of the fitted values for the response variables (RAO, 1964; DAVIES & TSO, 1982). In the latter case, reduced-rank regression is also called 'principal component analysis of instrumental variables' (RAO, 1964; ROBERTS & ESCOUFIER, 1976) and 'redundancy analysis' (VAN DEN WOLLENBERG, 1977; ISRAËLS, 1984; VAN DER BURG & DE LEEUW, 1990). TER BRAAK (1990) proposed the same biplots in the context of canonical correlation analysis.

In this paper we aim to make reduced-rank regression accessible to users of multiple regression by summarizing the essential theory of reduced-rank regression and of the biplot and by presenting an illustrative example. We discuss several aspects of biplots that so far received little attention, among which, how to display qualitative regressor variables in the biplot, how to focus the biplot analysis on the effects of a particular subset of regressors and how to scale biplots. Details are given about what can and what can not be learned from differently scaled biplots, in particular, biplots in correlation scaling and in distance scaling. The paper also presents a new interpretative aid for biplots of *t*-ratios, the so-called Van Dobben-circles.

## 2. Example: Description and Multiple Regressions

Our example is based on the study by KUNST et al. (1990). Their study contributes to the discussion about a finding that differences in life expectancy between occupational classes in Britain have widened during the postwar period

(e.g. HART, 1986). By using regional mortality data and a socio-economic index, Kunst et al. studied the socio-economic mortality differences in the Netherlands between 1950–1984. The data consist of regional standardized mortality ratios of different causes of death over four periods (1950–54, 1960–64, 1970–74 and 1980–84) in 39 Dutch regions together with a regional index for socio-economic status (SES) and two control variables (indices for urbanization and religion, URB and CAT) (Table 1). Standardized mortality ratios were used to correct for regional differences in the age structure of the population, and are calculated as the ratio of the observed number of deaths to the number of deaths that are expected from multiplying the national age-specific death rates to the regional age-specific population numbers (ARMITAGE, 1971: 388). The question of interest is how the total mortality ratio and the cause-specific mortality ratio are related to socio-economic status, and how these relationships change over time. Kunst et al. used multiple regression to answer these questions. In this study we present the major patterns of change in a single display. We confine the analysis to male mortality due to eleven major causes of death (Table 1). Apart from this simplification, we parallel the analysis in Kunst et al. The total male mortality ratio changed little in time and was only weakly related to SES, but the ratios of

Table 1

Full names and abbreviations of the response variables (standardized mortality ratios for eleven causes of death) and the regressor variables in the example data set.

Response Variables	Abbreviation
Stomach Cancer	StomCa
Colon Cancer	ColoCa
Lung Cancer	LungCa
Prostate Cancer	ProsCa
Diabetes mellitus	DiabMe
Ischaemic Heart	IschHea
Other Heart disease	OthHea
Arterial Disease	ArteDi
COLD	COLD
Traffic accidents	Traffic
Non-traffic accidents	NonTraf
Regressor Variables	Abbreviation
Period 1–4	$P_1 - P_4$
Socio-economic status	SES
in period 1–4	$SES_1 - SES_4$
Urbanization	URB
in period 1–4	$URB_1 - URB_4$
Religion	CAT
in period 1–4	$CAT_1 - CAT_4$

Table 2

Regression coefficients of Period and SES for 11 causes of death in the model Period + Period.(SES + URB + CAT). For SES, *t*-ratios are shown below the coefficients. Also given are the multiple correlation coefficient (R) of the causes of death, and the variance inflation factor (VIF) for regressors. For other abbreviations see Table 1.

Cause	Period				SES.Period				R
	1	2	3	4	1	2	3	4	
StomCa	0.09	0.03	-0.03	-0.10	-0.14	-0.11	-0.06	-0.20	0.91
					-1.59	-1.26	-0.72	-2.29	
ColoCa	-0.07	-0.03	0.03	0.07	0.16	0.54	0.28	0.00	0.75
					1.17	3.95	2.01	0.02	
LungCa	-0.10	-0.02	0.05	0.08	0.54	0.30	0.16	0.04	0.96
					9.91	5.61	2.95	0.73	
Prozca	-0.07	-0.01	0.02	0.06	0.22	0.17	0.16	0.18	0.70
					1.53	1.16	1.06	1.23	
DiabMe	-0.02	0.06	0.01	-0.05	0.32	0.21	-0.15	-0.05	0.68
					2.15	1.37	-0.97	-0.32	
IscHea	-0.12	0.01	0.07	0.04	0.28	0.05	-0.13	-0.22	0.95
					4.43	0.78	-2.12	-3.53	
OthHea	0.09	0.02	-0.08	-0.04	0.10	-0.10	-0.04	-0.01	0.82
					0.85	-0.89	-0.38	-0.12	
ArteDi	0.06	0.01	0.01	-0.07	-0.15	-0.20	0.10	0.20	0.79
					-1.15	-1.60	0.78	1.59	
COLD	-0.07	-0.02	0.04	0.06	-0.19	-0.14	-0.03	-0.12	0.77
					-1.41	-1.09	-0.25	-0.94	
Traffic	-0.04	0.05	0.07	-0.08	-0.09	-0.26	-0.57	-0.47	0.91
					-1.11	-3.04	-6.73	-5.55	
NonTraf	0.06	0.04	0.01	-0.10	-0.17	-0.07	0.14	0.03	0.81
					-1.44	-0.62	1.16	0.26	
VIF	1.50	1.50	1.50	1.50	1.48	1.48	1.48	1.48	

the respective causes of death changed much more and their relationship with SES changed markedly in time (see Figure 2 in Kunst et al.). We therefore focus on the cause-specific mortality ratios. The regression coefficients associated with Period and SES are given in Table 2 for each of the eleven causes of death. In the text, we use the abbreviated names of the variables (Table 1).

Details are as follows. In the regression analyses, KUNST et al. (1990) controlled for two regional variables, urbanization (URB) (expressed by the percentage of inhabitants living in large cities) and religion (CAT) (expressed by the percentage of Roman-Catholic inhabitants). Hence, for each cause of death and each period, the standardized mortality ratio (transformed to natural logarithms) was regressed on SES, URB and CAT. Average values for the entire study period were used for SES, URB and CAT, because their geographic pattern changed little during the study period (KUNST et al.). The same regression

Table 3

Analysis of variance table totalled across causes of death.

Source	df	SS (%)
Periods	3	64.6
+ URB + CAT	2	6.8
+ (URB + CAT) . Period	6	2.7
+ SES	1	2.4
+ SES . Period	3	1.6
residual	23	21.9

coefficients can be obtained by merging the data for the four periods into a data set of 156 units (39 regions times 4 periods) and carrying out a regression on the factor Period and the interaction terms of period with SES, URB and CAT. In terms of model formulae (McCULLAGH & NELDER, 1989), the regression model can conveniently be written as  $\text{Period} + \text{Period} \cdot (\text{SES} + \text{URB} + \text{CAT})$ . The term with the dot represents the effect of SES, URB and CAT within periods and thus includes their main effect. We will take this as our basic model. Kunst et al. used it to calculate an  $F$ -test on the interaction between Period and SES, although the validity of the test for repeated measurements is questionable. Table 2 presents the regression coefficients for Periods and SES. (Deletion of URB and CAT from the regression has little influence on the regression coefficients in Table 2). To make the regression coefficients comparable, the transformed mortality ratios for each cause of death were centred and scaled to unit variance, the sum of the four coefficients for Period was constrained to zero, and SES, URB and CAT were centred and scaled to unit variance within periods before the interaction terms were calculated. The four interaction variables so normalized are indexed by period number, e.g. for SES,  $\text{SES}_1 - \text{SES}_4$ . These linear transformations do neither influence the fit of the regressions nor of the reduced-rank regression, because the estimation method we use is scale-invariant. We produce in Figure 1 a graphical approximation of Table 2 that more easily conveys the overall patterns of change. Computations were carried out with CANOCO (TER BRAAK, 1991) and repeated in GENSTAT 5 (GENSTAT, 1987).

### 3. The Model for Reduced-Rank Regression

In matrix notation, the multivariate linear regression model is (e.g. MARDIA et al., 1979)

$$Y = XM + E \quad (1)$$

with  $X(n \times p)$  and  $Y(n \times q)$  matrices containing observations on  $n$  units for  $p$  regressor and  $q$  response variables, respectively,  $M$  the  $(p \times q)$  matrix of regression coefficients and  $E$  the  $(n \times q)$  matrix of random errors. The errors of different observations (rows of  $E$ ) are assumed to be independently and identically distributed with zero expectation and  $(q \times q)$  covariance matrix  $\Sigma$ . Imposing the restriction that the rank of  $M$  equals  $r$ , means that  $M$  can be factorized as

$$M = CB', \quad (2)$$

where  $C$  and  $B$  are  $(p \times r)$  and  $(q \times r)$  matrices. Insertion in (1) gives the reduced-rank regression model (IZENMAN, 1975; DAVIES & TSO, 1982)

$$Y = XCB' + E. \quad (3)$$

This model can also be written in the form of a factor analysis model (e.g. MARDIA et al., 1979)

$$Y = FB' + E, \quad (4)$$

by defining the  $(n \times r)$  matrix of factor scores

$$F = XC. \quad (5)$$

In other words, the reduced-rank regression model can also be derived from the fixed factor analysis model (4) by imposing the restriction that the factor scores are linear combinations of external regressor variables. In the terminology of factor analysis,  $B$  contains factor loadings. Model (4) is a proper factor analysis model only if  $\Sigma$  is diagonal.

#### 4. Biplots

The basis for the biplot graphic display (GABRIEL, 1971; 1982) of a  $(p \times q)$  matrix of rank  $r$  is its factorization in terms of the product of two matrices of order  $(p \times r)$  and  $(q \times r)$ , just as  $M$  is factorized into  $B$  and  $C$  in equation (2). A biplot is obtained by plotting the rows of  $B$  and the rows of  $C$  as vectors in an  $r$ -dimensional Cartesian coordinate system. For example, in the two-dimensional Figure 1 there are vectors for 11 response variables (causes of death) and four regressor variables  $SES_1$ – $SES_4$  for period 1–4. The matrix product  $CB'$  is represented in the plot by the scalar inner product between the vectors for the rows of  $B$  (response variables) and the vectors for the rows of  $C$  (regressor variables). Easy-to-use rules for reading the biplot follow from the algebraic identity for a scalar inner product between two vectors  $b$  and  $c$  (GREEN, 1976: section 3.3)

$$\begin{aligned} \langle b, c \rangle &= bc' = \|b\| \|c\| \cos(\theta) = \text{sign} \|c\| \| \text{projection of } b \text{ on } c \| = \\ &= \text{sign} \|b\| \| \text{projection of } c \text{ on } b \| \end{aligned} \quad (6)$$

where  $\|\cdot\|$  is the Euclidean length of a vector,  $\theta$  the angle between the vectors and  $\text{sign} = 1$  or  $-1$  depending on whether the angle is acute or obtuse. For example, projecting the vectors for SES on to the cause of death IscHea yields projection points in the natural order  $\text{SES}_1 - \text{SES}_4$ . Because the angle is acute for  $\text{SES}_1$ , the regression coefficient with  $\text{SES}_1$  is positive. The regression coefficient is smaller in period 2 and becomes negative for the periods 3 and 4. Moreover, the distances between the projection points are proportional to the corresponding differences in regression coefficients of the rank 2 model. For more rules and examples to read a biplot, see GABRIEL (1971, 1982), BRADU, GABRIEL (1978), SEBER (1984) or KRZANOWSKI (1988).

Not only  $\mathbf{M}$ , but also the expected values of the response variables  $E(\mathbf{Y})$  are factorized, namely by  $\mathbf{FB}'$  (equation 4). In Figure 1, one could therefore also plot the rows of  $\mathbf{F}$  to give vectors for each of the units. These vectors and the vectors for the rows of  $\mathbf{B}$  (response variables) form a biplot of the expected responses. The rules to read the plot are analogous to the biplot for  $\mathbf{M}$ . For greater similarity with the usual scatterplot, units are commonly indicated by points rather than by vectors. Because the biplot would become too crowded, we merely indicated, for each of the periods, the convex hull of the points.

## 5. Estimation

Following DAVIES & TSO (1982), we estimate the parameters of the reduced-rank model (3) by weighted least-squares. Maximum likelihood estimators for normal errors have been discussed for general  $\Sigma$  by TSO (1981) and for structured  $\Sigma$  by VAN DER LEEDEN (1990). The analysis by TSO (1981) leads to canonical correlation analysis. TER BRAAK (1990) describes how to construct the biplot of  $\mathbf{M}$  and  $E(\mathbf{Y})$  from standard computer output for this case. The least-squares analysis is instructive as it shows the relationships with multiple regression and principal components.

DAVIES & TSO (1982) showed that the error sum of squares allows the orthogonal decomposition

$$\|(\mathbf{Y} - \mathbf{XCB}')\Gamma\|^2 = \|(\mathbf{Y} - \hat{\mathbf{Y}}_p)\Gamma\|^2 + \|(\hat{\mathbf{Y}}_p - \mathbf{FB}')\Gamma\|^2 \quad (7)$$

with  $\hat{\mathbf{Y}}_p$  the matrix of fitted values obtained by separate multiple regressions (i.e. by multivariate regression with  $p$  regressors) and  $\Gamma$  a symmetric positive definite matrix. Statistical considerations lead to the choice  $\Gamma = \Sigma^{-\frac{1}{2}}$ , but for understanding what follows  $\Gamma$  can be ignored (i.e.  $\Gamma = \mathbf{I}$ ). The least-squares solution for any particular reduced-rank is obtained by minimizing the second term on the right-hand side of (7). As follows from the Eckhardt-Young theorem (ECKHARDT & YOUNG, 1936; DAVIES & TSO 1982) the minimum is obtained from the singular value decomposition (SVD)

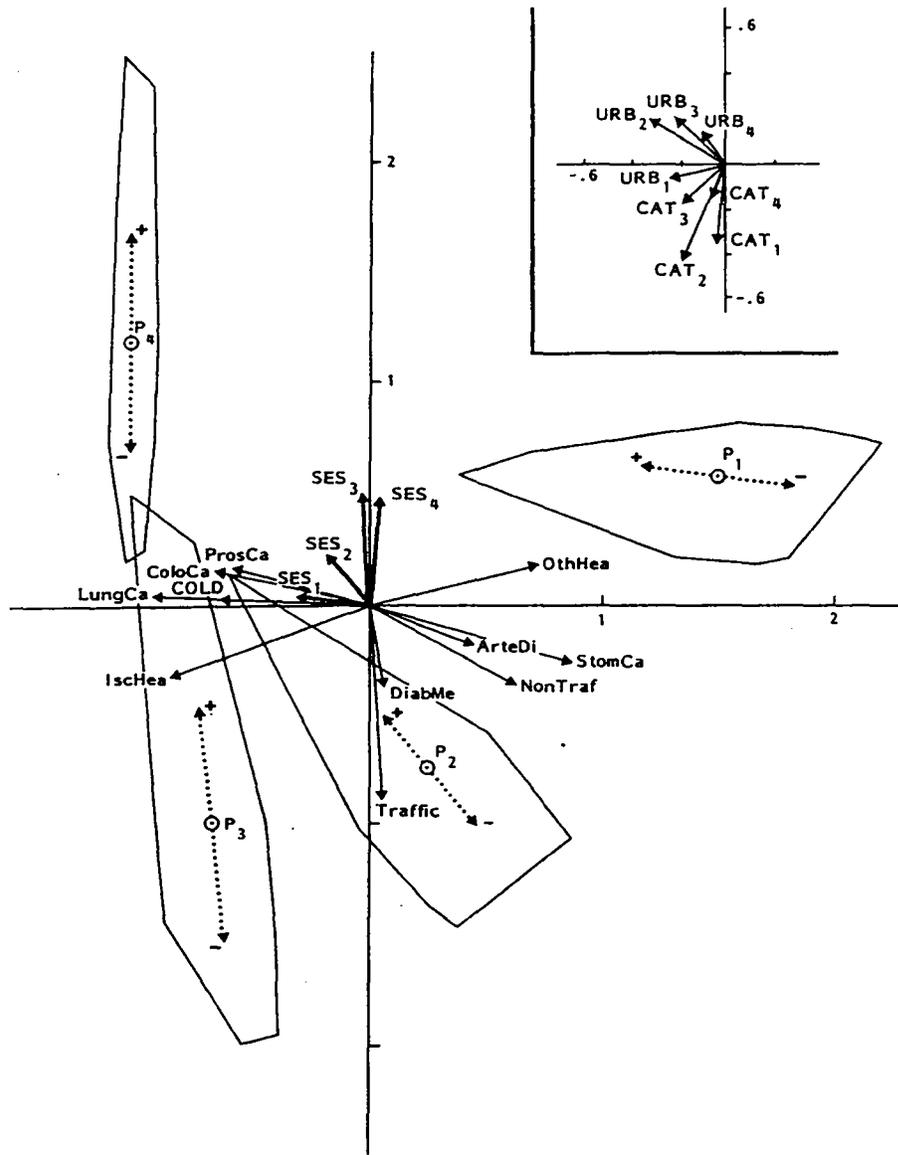


Fig. 1. Biplot of the regression coefficients in Table 2 based on reduced-rank regression of causes of death with model Period + Period . (SES + URB + CAT). The inset shows the vectors that approximate the coefficients of URB and CAT. The length of a vector for a cause of death is equal to its multiple correlation in the rank-2-regression model. The statistical units (regions) in each period are indicated by their convex hull. For further explanation, see Section 6.

$$\hat{Y}_p \Gamma = P \Lambda Q' \tag{8}$$

where  $P$  and  $Q$  are orthonormal matrices of order  $n \times t$  and  $q \times t$  with  $t = \min(n, q)$  containing the singular vectors, and  $\Lambda$  a diagonal matrix with the singular values on the diagonal in decreasing order ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_t \geq 0$ ). The minimum of the last term of (7) for the rank  $r$  model is  $\lambda_{r+1}^2 + \dots + \lambda_t^2$  and is attained by retaining the first  $r$  singular vectors, i.e. by setting (GABRIEL, 1971)

$$\hat{F} = n^{\frac{1}{2}} [P]_r \quad \text{and} \quad \hat{B} = n^{-\frac{1}{2}} \Gamma^{-1} [Q \Lambda]_r, \tag{9}$$

where  $[A]_r$  denotes the matrix consisting of the first  $r$  columns of the matrix  $A$ . (The factors involving  $n^{\frac{1}{2}}$  and the particular placement of  $\Lambda$  in (9) are scaling options explained in GABRIEL (1971). With (9),  $n^{-1} \hat{Y}' \hat{Y} = n^{-1} \hat{B}' \hat{F}' \hat{F} \hat{B}' = \hat{B}' \hat{B}'$ , so that innerproducts of rows of  $\hat{B}$  are 'covariances' among the fitted values in the rank  $r$  model; see also section 9). Finally, we obtain  $\hat{C}$  from (5) by regressing the columns of  $\hat{F}$  on  $X$ . If  $\Gamma = I$ , the solutions can thus be obtained by a principal components analysis of the fitted values  $\hat{Y}$ . Reduced-rank regression inherits from multiple regression the property that it is invariant for linear rescaling of the regressors, but from principal components analysis that is *not* invariant for linear rescaling of the response variables (JOLLIFFE, (1986). There are three methods to overcome the latter problem. (1) Standardize each response variable to zero mean and unit variance and use  $\Gamma = I$ . (2) Weight the response variables with the inverse of the error covariance matrix  $\hat{\Sigma}_p$ , estimated by multivariate regression, i.e. use  $\Gamma = \hat{\Sigma}_p^{-\frac{1}{2}}$ . (3) Weight each response variable with the inverse of its error variance ( $\hat{\sigma}_k^2$ ,  $k = 1 \dots q$ ), estimated by multiple regression. i.e. use  $\Gamma = \text{diag}(\hat{\sigma}_1^{-1}, \hat{\sigma}_2^{-1}, \dots, \hat{\sigma}_q^{-1})$ . Method (1) leads to redundancy analysis (VAN DEN WOLLENBERG, 1977), whereas method (2) leads to canonical correlation analysis as was first noted by VELU et al. (1986). Method (3), proposed by TER BRAAK (1990) as an intermediate between redundancy analysis and canonical correlation analysis, gives scale-invariance and is particularly attractive if method (2) is hazardous because of near-singular  $\Sigma$ , for example if  $n$  is less than or only marginally exceeds  $p + q$ . In our example we use method (3).

It follows from (8) and (9), that an alternative interpretation of reduced-rank regression is that it seeks a low-rank approximation of the usual full rank least-squares fit  $\hat{Y}_p$ .

It is instructive to see in which sense  $\hat{C} \hat{B}'$  approximates the matrix  $\hat{M}_p$  of regression coefficients as estimated by unrestricted least-squares. Note that from (8),  $\hat{Y}_p = X \hat{M}_p$  and  $P' P = I$ ,

$$\Gamma \hat{Y}_p' \hat{Y}_p \Gamma = \Gamma \hat{M}_p' X' X \hat{M}_p \Gamma = Q \Lambda^2 Q'. \tag{10}$$

Therefore,  $Q$  and  $\Lambda$  appear also in the singular value decomposition

$$(X' X)^{\frac{1}{2}} \hat{M}_p \Gamma = R \Lambda Q'. \tag{11}$$

This can be checked by noting that (10) follows from (11) because  $\mathbf{R}'\mathbf{R} = \mathbf{I}$ . Using (11), we are actually minimizing

$$\|(\mathbf{X}'\mathbf{X})^{\frac{1}{2}}(\hat{\mathbf{M}}_p - \mathbf{C}\mathbf{B}')\Gamma\|^2 \quad (12)$$

and the minimum for a rank  $r$  approximation is obtained by setting

$$\hat{\mathbf{C}} = n^{\frac{1}{2}}(\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}[\mathbf{R}]_r \quad \text{and} \quad \hat{\mathbf{B}} = n^{-\frac{1}{2}}\Gamma^{-1}[\mathbf{Q}\Lambda]_r. \quad (13)$$

Unsurprisingly,  $\hat{\mathbf{B}}$  is the same as in (9). In conclusion, reduced-rank regression with loss-function as in (7) yields a weighted least-squares approximation of  $\hat{\mathbf{M}}_p$  of the form  $\hat{\mathbf{C}}\hat{\mathbf{B}}'$ . The weight matrices in this approximation are  $\mathbf{X}'\mathbf{X}$  and  $\Gamma\Gamma$ . The estimated covariance matrix of the  $i$ -th column and  $j$ -th column of  $\hat{\mathbf{M}}_p$  is  $(\mathbf{X}'\mathbf{X})^{-1}(\hat{\Sigma}_p)_{ij}$  (ANDERSON, 1984: 291). So, with  $\Gamma = \hat{\Sigma}_p^{-\frac{1}{2}}$ , the weight matrices nicely factor the inverse of the covariance matrix of the estimated regression coefficients. VELU et al. (1986) derived formulae for the asymptotic variance of  $\hat{\mathbf{B}}$  and  $\hat{\mathbf{C}}$ .

## 6. Example: Reduced-Rank Regression

Reduced-rank regression was applied to the 11 causes of death with model formula  $\text{Period} + \text{Period} \cdot (\text{SES} + \text{URB} + \text{CAT})$ . In the estimation and subsequent summary statistics, causes of death were weighted inversely to their error variance. The reduced-rank models for rank 1, 2, 3 and 4 account 56, 71, 74 and 76% of the total sum of squares, respectively, whereas the full rank regression accounts for 78%. The rank 2 model, displayed in Figure 1, hence, accounts for 91% (71/78) of the unconstrained regression sum of squares.

The biplot of the rank 2 model (Figure 1) contains arrows for the causes of death and for regressors, namely,  $\text{SES}_{1-4}$  and, in an inset,  $\text{URB}_{1-4}$  and  $\text{CAT}_{1-4}$ . The periods are indicated by points ( $P_1-P_4$ ) rather than arrows; this way of display seems more natural because period is a qualitative variable, but this is a difference in display only (cf. GOWER, 1992). The positions of the regions in each period are indicated by their convex hull.

The lengths and directions of the arrows convey useful information. Recall that in the biplot based on (9) and (13), inner products between vectors for causes of death yield the 'covariances' among the fitted values in the reduced-rank model and that the total variance of each cause of death was scaled to 1. Therefore, the length of the vector for a cause of death is equal to its rank 2 multiple correlation. *StomCa*, for example, shows a high multiple correlation (ca. 0.9), whereas *DiabMe* has a low correlation (ca. 0.4). Causes of death with long arrows that point in the same direction, show strong positive correlation in their fitted values. Negative correlations are found among causes of death pointing in opposite directions. It is thus seen from Figure 1 that *ProsCa*, *ColoCa*, *COLD*

and LungCa follow a similar pattern with respect to the regressors, as do ArteDi, StomCa and NonTraf. The vector for a regressor indicates the shift of a statistical unit in the diagram when the value of the corresponding regressor increases by one unit of measurement. In our example, the unit of measurement for the quantitative regressors is one standard deviation. The shift obtained by increasing or decreasing SES<sub>j</sub> by one standard deviation for a notional region in period *j* that has average values for the other regressors is indicated in the figure by dotted lines labelled with a plus and a minus sign, respectively. To which changes in causes of death such a shift corresponds can be seen from the Figure because inner products of the shift vector with cause of death vectors yields the (rank 2) changes in cause-specific mortalities. From Figure 1 we thus see that the differences between periods are much bigger than the differences related to SES, URB and CAT. A quantitative decomposition is given in a multivariate analysis of variance table (Table 3). Each entry in Table 3 simply corresponds to an average (weighted by the inverse of the error variance) of the corresponding entries in the univariate analyses of variances.

Changes in time can easily be seen from Figure 1 by projecting the points for periods (P<sub>1</sub>–P<sub>4</sub>) on to a cause of death vector and looking at the order of the projection points. Schematically, causes of death with arrows pointing westwards increase (ProsCa, ColoCa, LungCa, COLD, IscHea), those with arrows pointing eastwards decrease in time (OthHea, StomCa, NonTraf, ArteDi), whereas causes of death with arrows pointing southwards first increase and later decrease (Traffic, DiabMe). But there is more detail in the Figure. On closer examination, IscHea is seen to increase from period 1 to 3 but to decrease thereafter, whereas OthHea of which the arrow points in the opposite direction, follows the opposite pattern. These inferences from Figure 1 agree with the changes in regression coefficients for periods in Table 2.

The relation with SES also changes in time, as can be seen by projecting the arrows for SES<sub>1</sub>–SES<sub>4</sub> on each cause of death vector. Schematically, causes of death with arrows pointing westwards start off with a positive regression coefficient which decreases to near zero (ProsCa, ColoCa, LungCa, COLD), those with arrows pointing southwards start with a regression coefficient near zero which later on becomes negative (Traffic, DiabMe), whereas causes of death with arrows pointing to the south-east have a near constant negative regression coefficient with SES (StomCa, NonTraf, ArteDi). The regression coefficient for IscHea is seen to decrease from positive to negative values and OthHea follows, again, the opposite pattern. These inferences do not agree in full with the regression coefficients for SES in Table 2. Most strikingly, the pattern displayed in Figure 1 for COLD, DiabMe, OthHea and ArteDi disagrees considerably with their regression coefficients in Table 2. Because the effects of Period dominate those SES (Table 3), the effects of SES are less faithfully displayed than those of Period. In section 8, we show how to display the effects of SES more precisely by eliminating the effects of the other regressors. Because the regression coefficients

associated with URB and CAT are nuisance parameters in the present analysis, we abstain from interpreting them. Their display in Figure 1 is not very precise.

An interesting aspect of Figure 1 is that the arrows for SES first point to the west and later move to the north, i.e. approximately to Period 4. By consequence, the points of regions with high SES in Periods 1–3 are close to (i.e. have a similar mortality pattern as) points of regions with low SES in Period 4. The figure suggests that regions with low SES lag behind.

### 7. Biplots of $t$ -Ratios

Standard output of a multiple regression analysis includes not only regression coefficients, but their associated  $t$ -ratio (estimate/estimated standard error). The biplot for  $\hat{M}_p$  can be enriched in a simple way to show approximate  $t$ -ratios as well (TER BRAAK, 1990). An example is Figure 2, which includes the vectors for causes of death and  $SES_1 - SES_4$  of Figure 1. The matrix of  $t$ -ratios  $T$ , say, and its approximation are

$$T = V^{-\frac{1}{2}} \hat{M}_p D^{-\frac{1}{2}} \approx (V^{-\frac{1}{2}} \hat{C}) (D^{-\frac{1}{2}} \hat{B})', \quad (14)$$

where  $V$  and  $D$  are diagonal matrices of orders  $p$  and  $q$  containing the diagonal elements of  $(X'X)^{-1}$  and  $\hat{\Sigma}_p$ , respectively. A biplot for  $T$  is therefore obtained from the biplot for  $\hat{M}_p$  by changing the lengths of the vectors in the latter. If the regressor variables are centred and scaled to unit length,  $V$  contains variance inflation factors (MONTGOMERY & PECK, 1982: section 8.4.2). With the regressor variables so scaled, the vector for any regressor variable must thus be divided by the square root of its variance inflation factor (a number  $\geq 1$ ) and will therefore be reduced in length. The reduction is indicated in Figure 2 by partly dashing the vectors for regressors. The higher the variance inflation factor the longer is the dashed part. Rather than changing the lengths of the vectors for the response variables, it is convenient to the mark position on each vector where a projection point of a regressor variable would precisely yield the critical  $t$ -ratio at a particular significance level, say 5%. The coordinates of the mark for the  $k$ -th response variable are  $\alpha_k \mathbf{b}_k$  with  $\alpha_k = (t_c \hat{\sigma}_k \|\mathbf{b}_k\|^{-2})$ ,  $t_c$  the critical  $t$ -ratio and  $\mathbf{b}_k'$  the  $k$ -th row of  $\hat{B}$ . In Figure 2, the line segment where the  $t$ -ratio is non-significant at the 5%-level is dashed. Notice that  $\alpha_k$  is sometimes greater than 1, for example for DiabMe in Figure 2.

From the enriched biplot it is easy to see which regressor variables have nonsignificant regression coefficients for any particular response variable. These are the regressors whose solid part projects into the dashed part of the response variable (or its mirror-image on the other side of the origin). Figure 3 (repro-

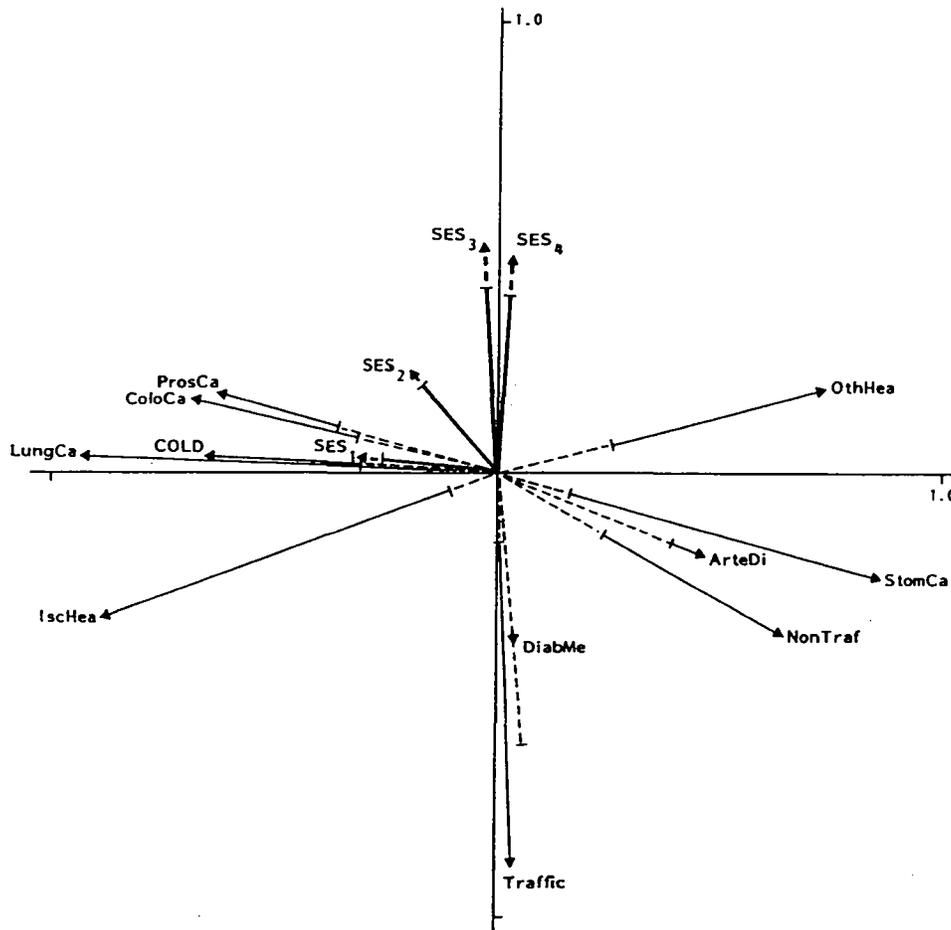


Fig. 2. Biplot of the coefficients and associated  $t$ -ratios for SES (Table 2). The positions where the lines change from solid to dashed are important for inferring  $t$ -ratios as explained in Section 7 and Figure 3.

duced from Figure 2) illustrates this for IscHea with dotted help-lines: the regressors with nonsignificant effect on IscHea lie in the region between these dotted lines. Thus, from Figure 3, SES<sub>2</sub> is nonsignificant, whereas SES<sub>1</sub> has a significant, positive effect on IscHea and SES<sub>3</sub> and SES<sub>4</sub> a (marginally) significant, negative effect. These inferences correspond to the ones in Table 2. Note, however, that the graphical test is not exact – even if the assumptions of a  $t$ -test hold true – because the biplot displays the observed  $t$ -ratios in Table 2 with some error.

It is also possible to indicate in which region the response variables lie that react significantly positive to a particular regressor. By elementary geometry, this

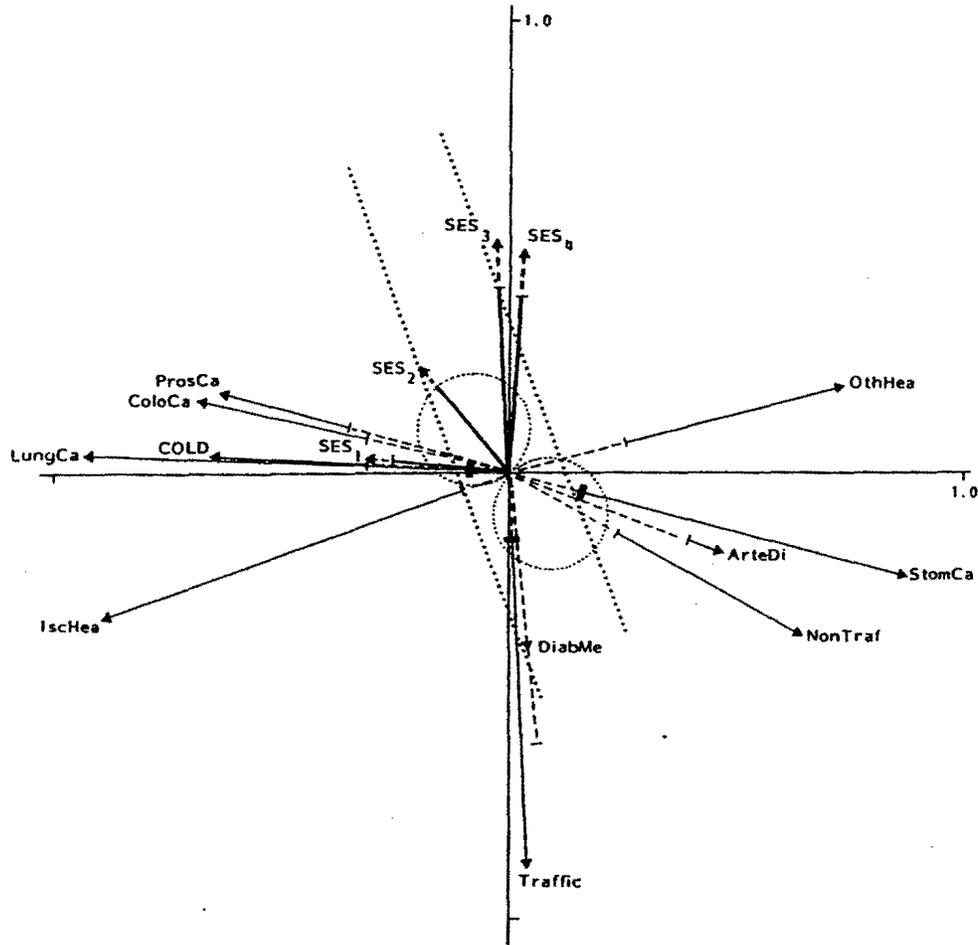


Fig. 3. Dotted lines added to Figure 2 for inferring the significance of regressors for IscHea. The solid lines for  $SES_4$  and  $SES_2$  reach outside the region enclosed by the dotted lines and are thus inferred to be significant. The dotted circles are Van Dobben-circles for inferring which regressands react significantly to  $SES_2$ . The causes of death that are inferred to react significantly are marked (LungCa, Traffic, StomCa).

region is a circle with as diameter the solid line-segment for that particular regressor variable. The response variables whose dashed line-segment lies wholly inside the circle have, in this approximation, a significant positive regression coefficient. This interpretative aid was first noted by H. Van Dobben. Figure 3 shows the Van Dobben-circle for  $SES_2$  and singles out LungCa as having a significant positive reaction to  $SES_2$ . By mirroring the circle in the tangent at the origin (see Figure 3), we find the region of response variables with significant, negative regression coefficients. From Figure 3, Traffic and StomCa are seen to

react significantly but negatively to  $SES_2$ . Comparison with Table 2 shows that the approximate graphical test missed the positive effect on ColoCa and falsely detected an effect on LungCa.

In the construction of the  $t$ -ratio biplot, the role of response variables and regressors can, in principle, be interchanged, i.e. the length of vectors for response variables can be changed and marks can be placed on vectors for regressor variables to indicate the critical  $t$ -ratio. This alternative  $t$ -ratio biplot has two disadvantages: (1) one can not derive the variance inflation factors from it and (2) it is more difficult to display  $t$ -ratios and regression coefficients in a single plot, because the response vectors may need to be magnified enormously.

## 8. Concomitant Regressors

Figure 2 focuses on the effects of particular regressors ( $SES_1 - SES_4$ ), but does not give a better approximation to their regression coefficients than Figure 1. The approximation can be improved by factoring the relevant part of  $\hat{M}_p$  only. For this purpose, we partition the regressors in two sets, and partition  $X$  likewise in  $X_1$  and  $X_2$  so as to contain concomitant regressors and regressors of interest, respectively. With  $M = (M_1 : M_2)$ , model (1) then becomes

$$Y = X_1 M_1 + X_2 M_2 + E \quad (15)$$

with  $M_1$  unconstrained and  $M_2$  constrained to rank  $r$ , i.e.  $M_2 = CB'$ . As ANDERSON (1951), DAVIES & TSO (1982) and VELU (1991) show, estimates for  $C$  and  $B$  are obtained by regressing both  $Y$  and  $X_2$  on  $X_1$  and applying the reduced-rank procedure to the residual matrices. Depending on the choices of  $\Gamma$ , this procedure is equivalent to partial redundancy analysis or partial canonical correlation (TER BRAAK, 1990). Figure 4 presents an example with  $SES_1 - SES_4$  as the only regressors of interest. The concomitant regressors are thus the Period dummy variables and the control variables and their interaction with Period. After fitting the concomitant variables, the terms  $SES + \text{Period} \cdot SES$  account for 4% of the original sum of squares (Table 3). The rank 2 model (Figure 4) displays 95% of this 4%. Figure 4 optimally approximates the regression coefficients and  $t$ -ratios for  $SES$  (Table 2) in two dimensions and is therefore better than the Figures 1 and 2 in this respect. For example, the circles drawn for assessing which causes of death are significantly related to  $SES_2$  now rightly contain ColoCa, LungCa and Traffic only.

In Section 6 we described on the basis of Figure 1 how the relationships with  $SES$  changed in time. Figure 4 tells largely the story, but without the discrepancies noted for Figure 1.

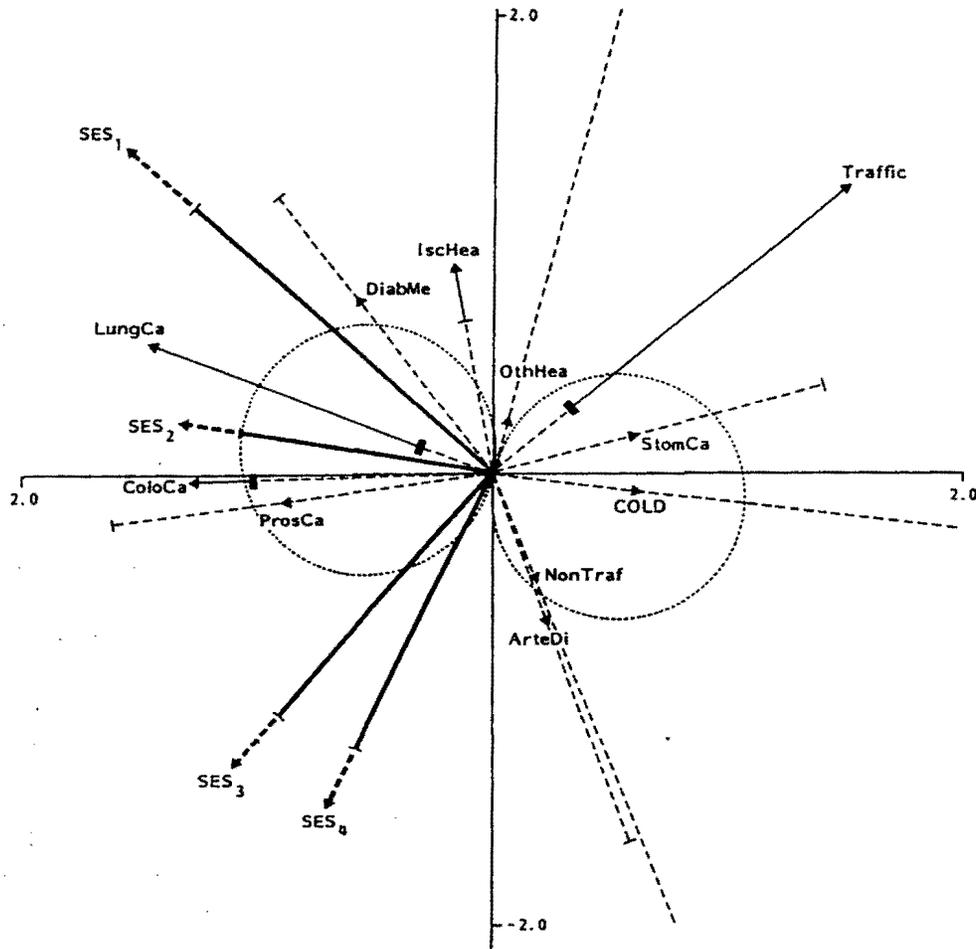


Fig. 4. Biplot of the coefficients and associated  $t$ -ratios for SES only (Table 2) based on reduced-rank regression with concomitant regressors. The dotted circles are for inferring which regressands react significantly to  $SES_2$ . The causes of death that are inferred to react significantly are marked (LungCa, Traffic, ColoCa). For clarity of display, the vectors for causes of death are magnified by a factor of 6. For further explanation, see Section 8.

### 9. Scaling of the Biplot

There are two alternative scalings of the biplot. In the first scaling, chosen in (9) and (13), the lengths and innerproducts of the arrows for response variables (rows of **B**) are the standard deviations and covariances, respectively, among the fitted values in the rank  $r$  model. In the second scaling, the squared length of the arrow for a regressor (rows of **C**) is the size of its effect, as measured by the weighted sum of squares of its rank  $r$  regression coefficients (the weights depend on  $\Gamma$ ).

This scaling also preserves the (weighted) Euclidean distance between the statistical units calculated from the rank  $r$  fitted values of the response variables (the weights depend again on  $\Gamma$ ). The second scaling is obtained by moving  $\Lambda$  in (9) and (13) from the equation for  $\hat{\mathbf{B}}$  to the equations for  $\hat{\mathbf{F}}$  and  $\hat{\mathbf{C}}$ , respectively, and deleting the factor involving  $n$ , i.e.

$$\hat{\mathbf{F}} = [\mathbf{P} \Lambda], \quad \hat{\mathbf{C}} = (\mathbf{X}' \mathbf{X})^{-\frac{1}{2}} [\mathbf{R} \Lambda], \quad \text{and} \quad \hat{\mathbf{B}} = \Gamma^{-1} [\mathbf{Q}], \quad (16)$$

This gives identical estimators for  $E(\mathbf{Y})$  and  $\mathbf{M}$  and thus the same error sum of squares as (9) and (12). Then, from (16),

$$\hat{\mathbf{Y}} \Gamma \Gamma \hat{\mathbf{Y}}' = \hat{\mathbf{F}} \hat{\mathbf{B}}' \Gamma \Gamma \hat{\mathbf{B}} \hat{\mathbf{F}}' = \hat{\mathbf{F}} \hat{\mathbf{F}}' \quad \text{and} \quad \hat{\mathbf{M}} \Gamma \Gamma \hat{\mathbf{M}}' = \hat{\mathbf{C}} \hat{\mathbf{B}}' \Gamma \Gamma \hat{\mathbf{B}} \hat{\mathbf{C}}' = \hat{\mathbf{C}} \hat{\mathbf{C}}' \quad (17)$$

so that, with  $\Gamma = \mathbf{I}$ , distances between units are Euclidean distances and, with  $\Gamma = \Sigma^{-\frac{1}{2}}$ , Mahalanobis distances, and the length of an arrow for a regressor is the norm of its rank  $r$  regression coefficients (with metric  $\Gamma \Gamma$ ). Use of scaling (16) instead of (9) would have been advantageous in Figure 4 for displaying the relative effect sizes of  $\text{SES}_1$ – $\text{SES}_4$ . The implied definition of effect size is, however, not satisfactory for these data because of the various standardizations employed in the original data and in the analysis.

TER BRAAK (1983, 1987) and GRENNACRE (1984) proposed the names 'covariance biplot' and 'distance biplot' for biplots in the first and second scaling, respectively. GABRIEL & ODOROFF (1990) term them 'Column Metric Preserving' and 'Row Metric Preserving'. If each of the quantitative variables is scaled to unit variance, the covariance biplot displays correlations instead of covariances and it would therefore more naturally to use the term 'correlation biplot'. However, this term is best reserved for plots that display correlations only. Such a correlation biplot is obtained by representing the regressors by their intra-set correlations with the canonical axes. The correlation biplot shows a weighted least-squares approximation of correlations between response variables and regressors (TER BRAAK, 1990, 1991). In Figures 1–4, by contrast, the regressors are represented by the canonical weights  $\mathbf{C}$ . As these plots display regression coefficients for the regressors and correlations for the (fitted) response variables, they are termed 'regression biplots in correlation scaling'.

There is another practical aspect in the scaling of biplots. In Figure 4, the vectors for causes of death as originally calculated by our computer programme were small in comparison to the vectors for SES. This came about because the additional fit due to  $\text{SES} + \text{SES.Period}$  is only 4% of the total variance (Table 3). For the purpose of display only, the vectors for causes of death were multiplied by six, as stated in the legend of the Figure. Regression coefficients inferred from innerproducts between cause of death vectors and SES vectors are therefore increased by a factor of six. Said differently, Figure 4 displays the rank 2 regression coefficients for causes of death that are each scaled to a variance of 36

(instead of 1). The magnification does not influence other aspects of the interpretation.

By contrast,  $t$ -ratios are scale independent and so is the biplot for  $t$ -ratios proposed in section 7. By construction, the coordinates for each response variable mark the critical  $t$ -ratio. The scale marks on the axes are not needed for reading the plot. Once all coordinates are calculated, the plot can therefore be uniformly rescaled at will. This was used in the Figures 2–4 to display the variance inflation factors as described in Section 7 despite the fact that the regressors were not standardized to unit sum of squares.

## 10. Discussion

Multiple regression models are often simplified by setting to zero regression coefficients of which the estimate is small in a statistical sense. In multivariate regression, a rival approach is to impose a restriction on the rank of the matrix of regression coefficients. This yields another proxy for the full-rank regression coefficients. The real gain of this approach lies in the dimension reduction that makes the biplot possible. Reduced-rank regression is useful because its biplot highlights the most important aspects of the regression.

DAVIES & TSO (1982) proposed reduced-rank regression in a context with many predictor variables compared to the number of statistical units, so that regression coefficients are underdetermined due to multicollinearity. However, reduced-rank regression is no solution to the multicollinearity problem. As noted below (9),  $\hat{C}$  is the solution of the multivariate regression of  $\hat{F}$  on  $X$ . Therefore, multicollinearity among the regressors in  $X$  destabilizes estimates of  $M$  and  $C$  in the same way. The variance inflation factors visible in the biplots of  $t$ -ratios from the dashed part of the regressor vectors (Figure 2) should warn the user against any multicollinearity.

In many observational studies, regression coefficients have little meaning because of multicollinearity among regressors. Multicollinearity invalidates the interpretation of a regression coefficient as the conditional effect of a regressor, given the values of the other regressors, and hence makes biplots of regression coefficients useless. Many users of multivariate techniques abstain therefore from interpreting regression coefficients altogether and focus on (biplots of) correlation coefficients (TER BRAAK, 1990; Rencher, 1988). By contrast, users of multiple regression seek solutions in procedures for subset selection. However, when analyzing a series of related response variables, the subsets of regressors tend to differ unpredictably among response variables. Multivariate forms of subset selection can be used to prevent this (e.g., ROBERT & ESCOUFIER, 1976; TER BRAAK, 1991). The effects of the selected regressors can subsequently be visualized and interpreted by reduced-rank regression. Although this combined approach does not solve any of the fundamental limitations of subset selection, it

may be helpful for exploring the structure in a regression problem beyond the structure contained in simple correlation coefficients. An alternative is to use biased estimation methods, such as partial least squares (AASTVEIT & MARTENS, 1986; HÖSKULDSSON, 1988) or ridge regression with rank restriction on the matrix of regression coefficients. The emphasis then necessarily shifts from interpreting regression coefficients to prediction.

In our illustrative example, there were no problems of multicollinearity. The biplot of reduced-rank regression helped to reveal a systematic relation between the increase or decrease of causes of death over time and their relation to SES (Figure 1). Simplified, Figure 1 shows that the high-class regions are trendsetters for 'modern' diseases, whereas the low-class regions lag behind in mortality pattern. Similar information could perhaps be presented more simply in optimally designed tables. We feel, however, that the biplot provides a good compromise between detail and overview.

As the theory and the example show, reduced-rank regression combines the attractive features of two popular statistical techniques: principal components analysis and multiple regression. In our view, reduced-rank regression should be added to the standard toolkit of applied statisticians.

#### Acknowledgements

The example was first presented at a workshop of the Dutch Society for Ordination and Classification (Autumn, 1989) devoted to the analysis of our example data. We are indebted to H. Van Dobben for pointing out that the *t*-ratio biplot could be enhanced with circles. We would like to thank A. Kunst for discussions and H. J. B. Birks, A. A. M. Jansen, B. Keen, M. Stapel and F. A. Van Eeuwijk for comments on the manuscript.

#### References

- AASTVEIT, A. H. & MARTENS, H., 1986: ANOVA interactions interpreted by partial least squares regression. *Biometrics* 42, 829–844.
- ANDERSON, T. W., 1951: Estimating linear restrictions on regression coefficients. *Ann. Math. Statist.* 22, 327–351.
- ANDERSON, T. W., 1984: *An introduction to multivariate statistical analysis* (2nd ed.). New York: Wiley.
- ARMITAGE, P., 1971: *Statistical methods in medical research*. Oxford: Blackwell Scientific Publications.
- BRADU, D. & GABRIEL, K. R., 1978: The biplot as a diagnostic tool for models of two-way tables. *Technometrics* 20, 47–67.
- DAVIES, P. T. & TSO, M. K.-S., 1982: Procedures for reduced-rank regression. *Appl. Statist.* 31, 244–255.

- ECKHART, C. & YOUNG, G., 1936: The approximation of one matrix by another of lower rank. *Psychometrika* 1, 211–218.
- GABRIEL, K. R., 1971: The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58, 453–467.
- GABRIEL, K. R., 1982: Biplot. In *Encyclopedia of Statistical Sciences*. Vol. 1, (ed. S. Kotz, and N. L. Johnson), pp. 263–271. New York: Wiley.
- GABRIEL, K. R. & ODOROFF, C. L., 1990: Biplots in biomedical research. *Statist. Med.* 9, 469–485.
- GENSTAT 5., 1987: GENSTAT 5 Reference manual. London: Clarendon Press.
- GOWER, J. C., 1992: Generalised biplots. *Biometrika* 79, 475–493.
- GREEN, P. E., 1976: *Mathematical Tools for Applied Multivariate Analysis*. New York: Academic Press.
- GREENACRE, M. J., 1984: *Theory and applications of correspondence analysis*. London: Academic Press.
- HART, N., 1986: Inequalities in health: the individual versus the environment. *J. Roy. Statist. Soc. A* 149, 228–246.
- HÖSKULDSSON, A., 1988: PLS regression methods. *J. Chemomet.* 2, 211–228.
- ISRAËLS, A. Z., 1984: Redundancy analysis for qualitative variables. *Psychometrika* 49, 331–346.
- IZENMAN, A. J., 1975: Reduced-rank regression for the multivariate linear model. *J. Mult. Anal.* 5, 248–264.
- JOLLIFFE, I. T., 1986: *Principal component analysis*. New York: Springer Verlag.
- KRZANOWSKI, W. J., 1988: *Principles of multivariate analysis*. Oxford: Clarendon Press.
- KUNST, A. E., LOOMAN, C. W. N. & MACKENBACH, J. P., 1990: Socio-economic mortality differences in the Netherlands in 1950–1984: a regional study of cause-specific mortality. *Soc. Sci. Med.* 31, 141–152.
- LEFKOVITCH, L. P., 1986: Linear predictivity: an alternative for MANOVA and multivariate multiple regression. *Biom. J.* 28, 771–781.
- MARDIA, K. V., KENT, J. T. & BIBBY, J. M., 1979: *Multivariate analysis*. London: Academic Press.
- MCCULLAGH, P. & NELDER, J. A., 1989: *Generalized linear models* (second edition). London: Chapman and Hall.
- MONTGOMERY, D. C. & PECK, E. A., 1982: *Introduction to linear regression analysis*. New York: Wiley.
- RAO, C. R., 1964: The use and interpretation of principal component analysis in applied research. *Sankhya A* 26, 329–358.
- RENCHER, A. C., 1988: On the use of correlations to interpret canonical functions. *Biometrika* 75, 363–365.
- ROBERT, P. & ESCOUFIER, Y., 1976: A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Appl. Statist.* 25, 257–265.
- SEBER, G. A. F., 1984: *Multivariate observations*. New York: Wiley.
- TER BRAAK, C. J. F., 1983: Principal components biplots and alpha and beta diversity. *Ecology* 64, 454–462.
- TER BRAAK, C. J. F., 1987: Ordination. In *Data analysis in community and landscape ecology*, (eds R. H. G. Jongman, C. J. F. Ter Braak, and O. F. R. Van Tongeren), pp. 91–173. Wageningen: Pudoc.
- TER BRAAK, C. J. F., 1990: Interpreting canonical correlation analysis through biplots of structural correlations and weights. *Psychometrika* 55, 519–531.
- TER BRAAK, C. J. F., 1991: *CANOCO – a FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis, principal components analysis and redundancy analysis (version 3.1)*. Wageningen: Agricultural Mathematics Group.
- TSO, M. K.-S., 1981: Reduced-rank regression and canonical analysis. *J. Roy. Statist. Soc. B* 43, 183–189.

- VAN DEN WOLLENBERG, A. L., 1977: Redundancy analysis. An alternative for canonical correlation analysis. *Psychometrika* 42, 207–219.
- VAN DER BURG, E. & DE LEEUW, J., 1990: Non-linear redundancy analysis. *Brit. J. Math. Stat. Psych.* 43, 217–230.
- VAN DER LEEDEN, R., 1990: *Reduced rank regression with structured residuals*. Leiden: DSWO Press.
- VELU, R. P., REINSEL, G. C. & WICHERN, D. W., 1986: Reduced rank models for multiple time series. *Biometrika* 73, 105–118.
- VELU, R. P., 1991: Reduced rank models with two sets of regressors. *Appl. Statist.* 40, 159–170.

Received, Aug. 1993

Accepted, Sept. 1993

Dr. C. J. F. TER BRAAK  
DLO-Agricultural Mathematics Group  
Box 100  
6700 AC Wageningen  
The Netherlands