A NOTE ON PRINCIPAL COMPONENT REGRESSION

EDWARD GREENBERG

Number 30

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

A NOTE ON PRINCIPAL COMPONENT REGRESSION

EDWARD GREENBERG

Number 30

September, 1972

A Note on Principal Component Regression

Edward Greenberg *

The purpose of this note is to provide an interpretation for the regression coefficients obtained when y is regressed on successive principal components of a matrix of independent variables, X. [1/] The approach is motivated by Goldberger's suggestion (1) that when multicollinearity is present the statistician should interpret the X matrix as a 'badly designed experiment' and attempt to find those linear combinations of parameters which can be estimated with reasonably good precision. A similar point was made by Silvey: "Relatively precise estimation is possible in the direction of the latent vectors of X'X corresponding to large latent roots; relatively imprecise estimation in these directions corresponding to small latent roots." (5)

Specifically, our approach is to find that linear combination of parameters which can be estimated with minimum variance; second, the linear combination orthogonal to the first with minimum variance, etc. When multicollinearity is a serious problem, after a few linear combinations the variance may become sufficiently large that the investigator decides not to estimate additional linear combinations.

I. Minimum variance linear combinations and principal components.

The variance of a linear combination, $c'\hat{\beta}$, is $\sigma^2 c'(X'X)^{-1}c$ {2.P126. }. We ignore the constant σ^2 and impose the normalizing rule $c'c=1$ to avoid the trivial solution $c=0$.

* Thanks to Steven Colby, Farouk El-Sheikh and Robert Parks.

Minimizing (1) $H = c'(X'X)^{-1}c - \mu(c'c - 1)$ we obtain

$$(2) \quad \frac{\delta H}{\delta C} = 2(X'X)^{-1}c - 2\mu c = 0 \quad \text{or}$$

$$(3) \quad (X'X)^{-1}c - \mu c = 0$$

$$(4) \quad (X'X)^{-1} - \mu I \quad c = 0$$

Hence, μ is a characteristic root of $(X'X)^{-1}$ and c is the associated characteristic vector. Further since

$$(5) \quad c'(X'X)^{-1}c = \mu c'c = \mu$$

we choose the smallest characteristic root, μ_1, and its associated characteristic vector, v_i. (in Silvey's notation).

This result can also be derived from Silvey's expression for the variance of $c'\hat{\beta}$:

$$\frac{\text{var}\,(c'\hat{\beta})}{\sigma^2} = \frac{\alpha_1^2}{\lambda_1} + \frac{\alpha_2^2}{\lambda_2} + \ldots\ldots \frac{\alpha_j^2}{\lambda_j} \quad \text{(Silvey P.}$$

where the α_i are the coefficients permitting the expression of c interms of the characteristic vectors, v_i. That is, $c = \alpha_1 v_1 + \alpha_2 v_2 + \ldots .\alpha_j v_j$, and $\Sigma\alpha_i^2 = 1$. The λ are the associated characteristic roots, $\lambda_1 > \lambda_2 > \ldots > \lambda_j$

To minimize the variance, choose $\alpha_1 = 1$, $\alpha_i = 0$, $i \neq 1$. obtaining the result that var $\frac{c'\hat{\beta}}{\sigma^2}$ is proportional to $1/\lambda_1$ the reciprocal of the largest characteritic root of X'X. This equals μ_1, the smallest characteristic root of $(X'X)^{-1}$:

The characteristic vector associated with the largest characteristic root of X'X is, of course, the vector of loadings of the first principal component of X. The c which minimizes the variance of c' is therefore the set of loadings associated with the first principal component, 2/

Continuing, the linear combination, orthogonal to v_1, having the smallest variance corresponds to the second principal component etc.

It is of interest to relate these linear combinations of β to regression on principal components. Consider the regression of y on the first principal components, Xv_1, $y = (Xv_1)\gamma_1$, say. The least squares estimator of γ_1, is

$$\hat{\gamma}_1 = (v_1' X'Xv_1)^{-1} v_1' X'y$$

$$= (\gamma_1 v_1' v_1)^{-1} v'X'y$$

$$= (1/\gamma_1)\, v_1' X'y$$

$$= v_1' (X'X)^{-1} X'y$$

$$= v_1' \hat{\beta}$$

Thus, the coefficient obtained when y is regressed on the first principal component of X'X may be interpreted as that linear combination of the elements of β which can be estimated with smallest variance, given the sample of independent variables. Since the components are orthogonal to each other, regressions of y on further principal components may be interpreted analagously.

II. Discussion and Illustrations.

Note that measuring X in standard units affects the linear combinations which will be obtained in a nontrivial way. This standardization has a convenient property, however; the characteristic root of a variable orthogonal to the others will equal one. For example, if each $X'X = I$, each $\lambda_i = 1$, and each v_i is of the form $(0,0,1,0,\ldots 0)$; thus each β_i is estimated with the same standard error, and no transformation is necessary.

If $X'X = \begin{pmatrix} 1 & r & 0 \\ r & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

the characteristic roots and vectors are

$$\begin{array}{ll} 1 + r, & 1/\sqrt{2}\ (1, 1, 0) \\ 1, & (0, 0, 1) \\ 1 - r, & 1/\sqrt{2}\ (1, -1, 0) \end{array}$$

The variances are therefore proportional to $1/(1+r)$, 1, and $1/(1-r)$, respectively. As r approaches 1, the linear combination $\beta_1 + \beta_2$ becomes relatively well estimated, and as r approaches -1, $\beta_1 - \beta_2$ is well estimated.

Since $\Sigma\lambda_i = n$ and all $\lambda_i \geq 0$, if some $\lambda_i = 1$ others will be less than one. This suggests that one should not drop principal components whose roots are greater than one, and that components with roots a little smaller should also be considered to avoid the danger that a variable which is highly correlated with y, but not with the other X's is dropped from the analysis. Two published studies using principal components appear to have avoided this danger: Stone (6) included a component with a root of .3859, and Meyer and Kraft (4) included one as low as .299.

Consider the data used by Malinvaud (3) to illustrate multicollinearity. He explains French imports from 1949 to 1959 as a function of gross domestic product (GDP), stock formation and consumption. Extracting principal components from these data, we obtain the following results:

i	λ_i	$1/\lambda_i$	v
1	2.0472	.4884	(-.692, -.143, -.708)
2	.9502	1.0524	(-.213, .977, .010)
3	.0026	384.6153	(-.690, -.158, .706)

It is clear that estimating more than two linear combinations of the β_i will yield very imprecise results: the variance of the third linear combination is some 800 times as large as the first, and over 350 times as large as the second 3/ . This impression is confirmed by the regression results obtained by Malinvaud: =

Independent variable	Coefficient	Standard Error	t-Statistic
X_1 GDP	-.078	.114	-.6842
X_2 Stock formation	.648	.187	3.4652 *
X_3 Consumption	.329	.164	2.0061

* Significant at 5% level

The results indicate that stock formation is highly significant; multicollinearity is suggested by the weak effects of GDP and consumption, the high correlation between them (.9973), and the 95% confidence ellipsoids displayed in FIGURES 1, 2 and 3. The ellipsoids reveal a wide range of values possible for the coefficients of GDP and consumption. The fact that the third characteristic root is close to 0 is a further indication that multicollinearity is present.

FIGURE 1

95% Confidence Ellipse for coefficients of X_1 and X_2

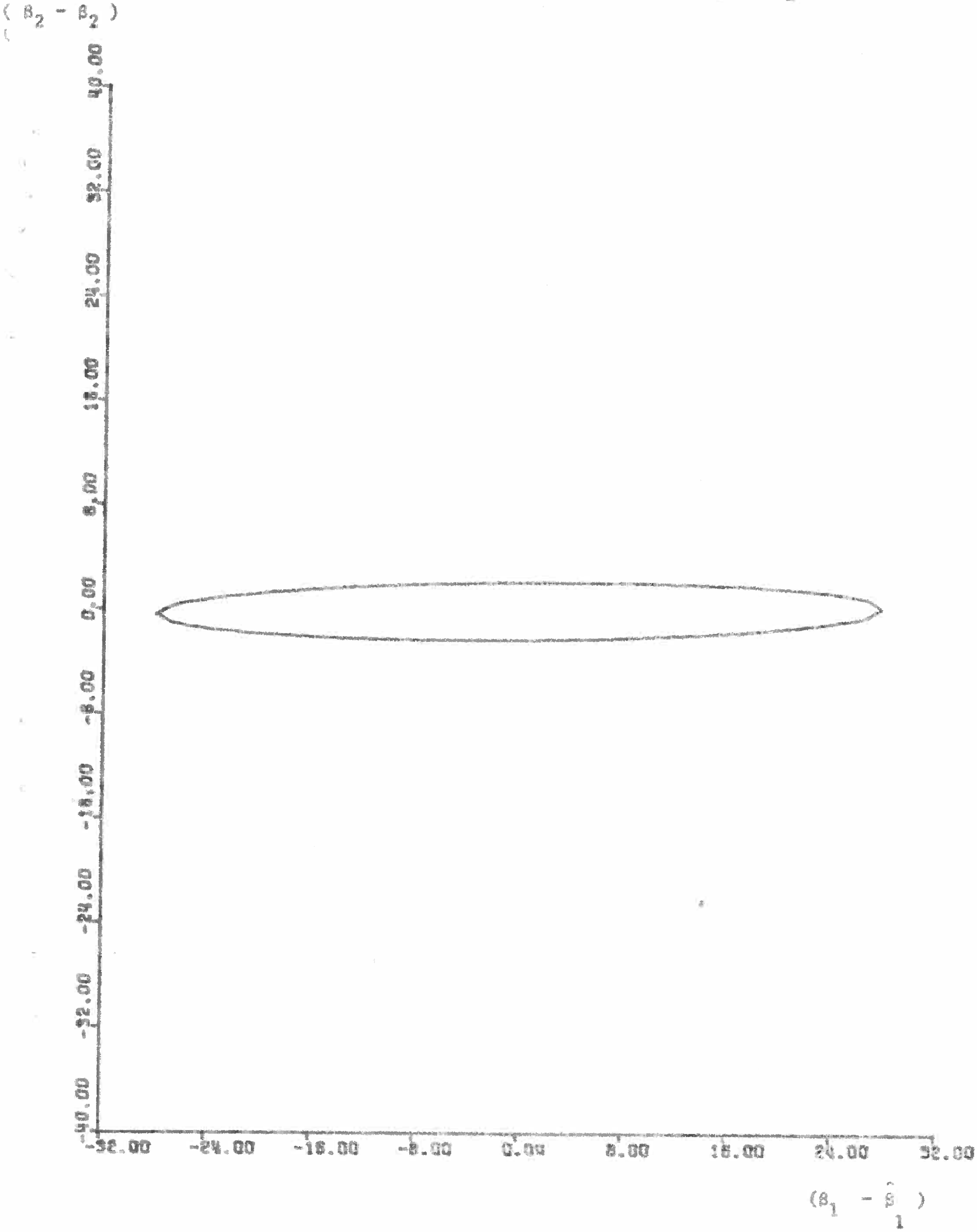
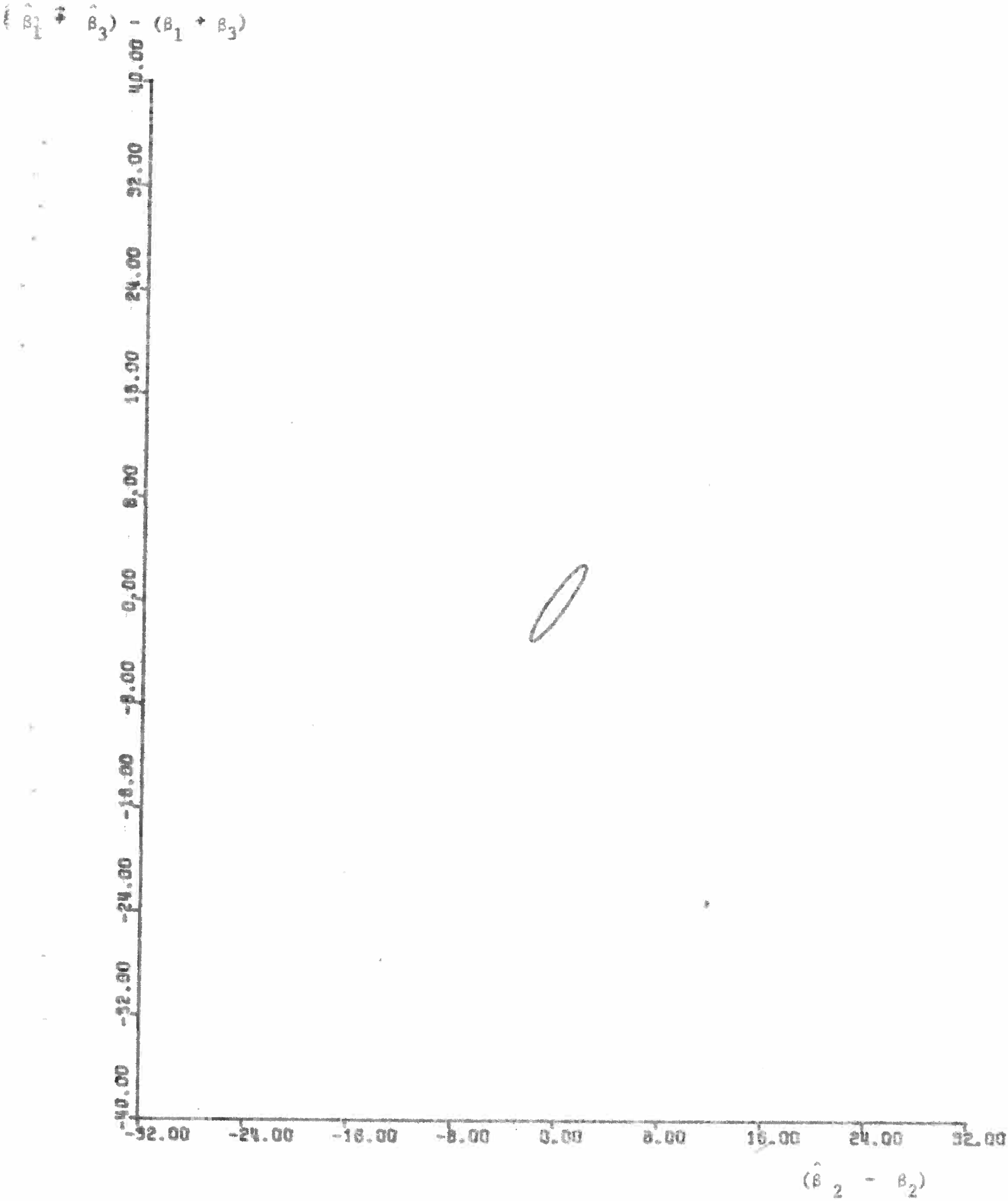


FIGURE 4

95% Confidence Ellipse for coefficients of $(X_1 + X_3)$ and X_2



III. Conclusions

The regression coefficient of the first principal component has been interpreted as that linear combination of the β_i which can be estimated with minimum variance. In that context, comparison of the characteristic roots with the diagonals of $(X'X)^{-1}$ (as in TABLE 1) provides information about how precisely various linear combinations of β can be estimated. Although linear combinations with low variances may have no 'natural' economic interpretation, it appears desirable to examine them when multicollinearity is viewed as a poorly-realized, unplanned experiment 4/. Examining the characteristic roots and vectors may be useful for both detecting multicollinearity indicated by large increases in the variances of successive linear combinations and for suggesting transformations which will permit relatively precise estimation of economically meaningful parameters.

TABLE 1

Independent Variable	Diagonal element of relevent (X'X) matrix
First principal component	.4884
Second principal component	1.0524
Stock formation (original regression)	1.07
GDP plus consumption, stock formation $X_1 + X_3$, X_2 (transformed equation)	1.3405
Consumption (original regression)	192.37
GDP (original regression)	193.28
Third principal component	384.615

Footnotes

1/ It is assumed that the elements of X are measured in standard deviation units; hence X'X is the correlation matrix.

2/ Analagous results may be obtained using the generlazed inverse of X'X if the matrix is singular.

3/ These and the following comparisons are in the terms of the diagonal elements in the $(X'X)^{-1}$ matrix. They must be multiplied by an estimated σ^2 to obtain the estimated variance. For any particular $\hat{\sigma}^2$, the relative values of the estimated variances are correct. In this example, $\hat{\sigma}_2$ does not differ greatly among the various regressions estimated.

4/ It is, of course, preferable to find additional data sources or other information to get around the multicollinearity problem.

References

1. Goldberger, A.S. Topics in Regression Analysis, New York: Macmillan, 1968.

2. Johnston, J. Econometric Methods, 2nd Edition, New York: McGraw-Hill, 1972.

3. Malinvaud, E. Statistical Methods of Econometrics, Chicago: Rand McNally, 1966.

4. Meyer, J. and Kraft, G. "The Evaluation of Statistical Costing Techniques as Applied in the Transportation Industry," American Economic Review, LI, (May 1961).

5. Silvey, S.D. "Multicollinearity and Imprecise Estimation," J. Royal Statistical Society, Series B, Vol. 31, 1969.

6. Stone, J.R.N. "The Analysis of Market Demand," J. Royal Statistical Society, Vol. CVIII, 1945.