▼ Journal of Research (Science), Bahauddin Zakariya University, Multan, Pakistan.
Vol.15, No.1, June 2004, pp. 33-39 ISSN 1021-1012

ESTIMATION AND ANALYSIS OF REGRESSION COEFFICIENTS WHEN EXPLANATORY VARIABLES ARE CORRELATED

G R Pasha[1], Muhammad Akbar Ali Shah[2] and Ghosia[3]
[1]Department of Statistics, Bahauddin Zakariya University, Multan.
[2]Department of Statistics, Islamia University, Bahawalpur.
[3]Institute of Statistics, Punjab University, Lahore.

Abstract: In this paper an unconventional method of the principal components regression is adopted for the solution of multicollinearity and an attempt is made to show that by using this technique, some fairly precise estimates of the co-efficient are obtained. This attractive property of the principal components regression makes it superior to the OLS method while dealing with the multicollinear data. Comparison between the variance of the OLS- estimates and the principal components regression on income and consumption is made.

Keywords: Latent roots, least square, multicollinearity, principal components, regression.

INTRODUCTION

Principal components regression is a method of inspecting the sample data on design matrix for directions of variability and using this information to reduce the dimensionality of the estimation problem. The reduction in dimensionality is achieved by imposing exact linear constraints that are sample specific but have certain minimum variance properties, (See Greenberg [1975], Fomby and Hill [1978], and Johnson *et al*. [1973]). The method of principal components regression has received considerable attention in recent years as a method for dealing with ill-conditioned data (See Farebrother [1972], Fomby and Hill [1978], Hill *et al*. [1977], Johnson *et al*. [1973], and Massey [1975]). In ill-conditioned linear regression problems, in which the regressors are nearly collinear, the use of ordinary least squares (OLS) is generally to be avoided owing to its poor performance, such as large means square errors (MSE). In such problems, biased parameter estimators may provide much lower MSE values than the unbiased OLS estimator. Attempts to find reasonably accurate approximations of the minimum MSE parameter estimator have been reported in the literature but apparently they were only successful to a limited degree [Soderstrom and Stoica 1995].

There are various methods of estimating the parameter vector β. These include OlS, ridge regression (RR), principal components regression (PCR) and partial least squares regression [Butler and Denham 2000, Frank and Friedman 1993, Helland 1990, Helland and Almoy 1994].

For the k-variable regression involving explanatory variables $X_1,X_2,\ldots..,X_k$ an exact linear relationship is said to exist if the following condition is satisfied: $\lambda_1 X_1+ \lambda_2 X_2+,\ldots\ldots\ldots,\lambda_k X_k=0$ (1)

where $\lambda_1,\lambda_2,.....,\lambda_k$ are constants such that not all of them are zero simultaneously. The term multicollinearity is used in a broader sense to include the case of perfect multicollinearity, as in Eq.(1) , as well as the case where the X variables are highly interrelated but not perfectly so as follows:

$$\lambda_1 X_1+ \lambda_2 X_2+,\ldots\ldots\ldots,\lambda_k X_k+\in_i=0 \qquad (2)$$

where $\in_k$ is a stochastic error term. To see the difference between perfect and less than perfect multicollinearity, assume that $\lambda_2\neq0$ then Eq.(1) may be written as

$$X_{2i}= -(\lambda_1 /\lambda_2)X_{1i} - (\lambda_3 /(\lambda_2)X_{3i} -,\ldots\ldots\ldots,- (\lambda_k /\lambda_2)X_{ki} \qquad (3)$$

Which shows how X_2 is exactly linearly related to other variables or how it can be derived from a linear combination of other X variables. In this situation, the co-efficient of correlation between the variable X_2 and the linear combination on the right hand side of Eq.(3) is bound to be unity.

Similarly if $\lambda_2\neq0$, Eq.(2) can be written as

$$X_{2i}= -(\lambda_1 /\lambda_2)X_{1i} - (\lambda_3 /(\lambda_2)X_{3i} -,\ldots\ldots\ldots,- (\lambda_k /\lambda_2)X_{ki} –(1/\lambda_2)\in_I \qquad (4)$$

Which shows that X_2 is not an exact linear combination of other X's because it is also determined by the stochastic error term $\in_i$.

THE METHOD OF PRINCIPAL COMPONENTS

The aim of the method of the principal components is the construction out of a set of variables, X_j's , j=1,2,….,k of new variables z_i called principal components, which are linear combinations of the X's.

$$z_{1t} = a_{11}x_{1t}+a_{21}x_{2t}+,\ldots\ldots,+a_{k1}x_{kt}; \qquad t = 1,2,\ldots,n$$
$$z_{2t} = a_{12}x_{1t}+a_{22}x_{2t}+,\ldots\ldots,+a_{k2}x_{kt}$$
$$:$$
$$:$$
$$z_{kt} = a_{1k}x_{1t}+a_{2k}x_{2t}+,\ldots\ldots,+a_{kk}x_{kt}$$

Here a's are called loadings which are chosen, so constructed principal components satisfy two conditions;

a) The principal components are uncorrelated.
b) The first principal component z_1 absorbs and accounts for the maximum possible proportion of the total variation in the set of all X's, the second principal component absorbs the maximum of the remaining variation in the X's (after allowing for the variation accounted for the first principal component) and so on.

TEST FOR THE SIGNIFICANCE OF THE LOADINGS

The loadings are in fact similar to correlation co-efficients. This test does not take into account the number of variables, X's in the set, and the order of extraction of the principal components. Burt and Banks [1947] have suggested the following adjustment to the standard error of the correlation co-efficients in order to obtain the standard errors of the loadings

$$S(a_{ij}) = \{S(r_{xj\ xm})\}(k/k+I-i)^{1/2}$$

Where k=number of X's in the set. I=subscript of Z, i.e., the order of its extraction (the position of Z in the extraction process). The Burt-Banks formula, clearly takes into account both the number of X's and the order of extraction of the Z's.

BARLETT'S CRITERIA FOR THE NUMBER OF PRINCIPAL COMPONENTS TO BE EXTRACTED

Assume the latent roots are computed for k variables $\lambda_1,\lambda_2,\ldots\ldots,\lambda_k$ and that the first r roots $\lambda_1,\lambda_2,\ldots\ldots,\lambda_r$, (for r<k) seem both sufficiently large and sufficiently different to be retained. The question then whether the remaining (k-r) roots are sufficiently alike for one to conclude that the associated Z's should be retained in the analysis. Bartlett [1954] has suggested that the quantity

$$\chi_c^2 = n\, I_n \{(\lambda_{r+1},\lambda_{r+2},\ldots\ldots,\lambda_k)^{-1}(\lambda_{r+1},\lambda_{r+2},\ldots\ldots,\lambda_k/(k-r))^{k-r}\}$$

has a χ^2-distribution (approximately) with v=1/2(k-r-1)(k-r-2) degrees of freedom. The null hypothesis have assume equality of the excluded latent roots, i.e.

$$H_0 = \lambda_{r+1} = \lambda_{r+2} = \ldots. = \lambda_r$$

If $\chi_c^2 > \chi^2_{(1-\alpha,v)}$, we reject the null hypothesis, that is we accept that the excluded latent roots are not equal; hence, we should include additional Z's in our analysis.

PRINCIPAL COMPONENTS REGRESSION

Let the model under consideration be

$$\begin{aligned} Y &= XAA'\beta + e \\ &= XA\theta + e \qquad \therefore A'\beta = \theta \\ &= Z\theta + e \qquad \therefore \quad Z = XA \end{aligned} \tag{5}$$

where A = $(a_1,\ldots.,a_k)$ is a kxk matrix whose columns (a_i) are orthogonal characteristic vectors of X'X ordered to correspond to the relative magnitudes of the characteristic roots of the positive definite matrix X'X and Z = $(z_1,\ldots.,z_k)$ is the nxk matrix o principal components. Accordingly $z_i = Xa_i$ is called the ith principal component, where $z_i'z = \lambda_l$ is the ith largest characteristic root of X'X.

The principal components estimator of β is obtained by deleting one or more of the variables z_i applying ordinary parameter space. Assume for the moment that Z has been partitioned into two parts z_1 the z_i to be retained, and z_2 the z_i to be deleted. This partitioning imposes an identical partitioning on A. Thus Eq.(5) becomes

$$\begin{aligned} Y &= XA_1\theta_1 + XA_2\theta_2 + e \\ &\quad Z_1\theta_1 + 'z_2\theta_2 + e \end{aligned} \tag{6}$$

Where X = $\{A_1:A_2\} = \{z_1:z_2\}$ $\Rightarrow \theta\hat{}_1 = (z_1'z_1)^{-1}z_1'Y$ the LS estimator of θ_1 with z_2 omitted from equation (5.2) can be easily obtained. Specifically, $\theta\hat{}_1$ is unbiased due to the orthogonality of z_1 and z_2. Its variance covariance matrix is given by

$$V(\hat{\theta}_1) = \sigma^2(z_1'z_1)^{-1}$$

Since $\beta = A_1\theta_1 + A_2\theta_2$. Omitting the components in z_2 means that θ_2 has implicitly been set equal to zero. Hence $A_2\theta_2 = 0$ and the principal components estimator of β is

$$\hat{\beta} = A_1\hat{\theta}_1 = A\hat{\theta}$$

where $\hat{\theta} = (\hat{\theta}_1, \hat{\theta})'$ with 0 a null vector of conformable dimension.

EXAMPLE

Klein and Goldberger [1964] attempted to fit the following regression model to the United States economy:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + e_i$$

where Y=consumption, X_1=wage income, X_2=non-wage, non-farm income X_3 = farm income. The data [Klein and Goldberger 1964] are presented in Table 1. The results of the data are summarized in Tables 2 and 3.

Table 1:

Y	X_1	X_2	X_3
62.8	43.41	17.10	3.96
65.0	46.44	18.65	5.48
63.9	44.35	17.09	4.37
67.5	47.82	19.28	4.51
71.3	51.02	23.24	4.88
76.6	58.71	28.11	6.37
86.3	87.69	30.29	8.96
95.7	76.73	28.26	9.76
98.3	75.91	27.91	9.31
100.3	77.62	32.30	9.85
103.2	78.02	31.39	7.21
108.9	83.57	35.61	7.39
108.5	90.59	37.58	7.98
111.4	95.47	35.17	7.42

Table 2:

Predictors	Coefficients	Variances	T-value
Constant	81.703	46.854	2.73
Wage income	0.380	0.097	1.22
Non-wage, non-farm income	1.418	0.519	1.97
Farm income	0.533	1.960	0.38

Table 3: Analysis of Variance.

Source of Variation	Degree of Freedom	Sum of Square	Mean Square	F-value
Regression	3	4151.2	1383.7	36.68
Error	10	367.2	36.7	
Total	13	4518.4		

$$\text{Correlation matrix} = \begin{vmatrix} 1 & 0.9431 & 0.1069 \\ & 1 & 0.7371 \\ & & 1 \end{vmatrix}$$

We see that coefficient of determination $R^2 = 91.9\%$ is highly significant but in contrast all the β's are insignificant. The computations reveal that the cause of multicollinearity lies mainly in the intercorrelation between X_1 and X_2. Now it has been proved that the data are highly multicollinear.

The principal component analysis based on correlation method yields the summary Table 4.

Table 4: Coefficients for Principal Components (Correlations coefficients in parentheses).

Variable	a_1	($r_{z1x'j}$)	a_2	($r_{z2x'j}$)	a_3	($r_{z3x'j}$)
X_1	0.5973	(0.9747)	-0.2264	(-0.1208)	0.7587	(0.1651)
X_2	0.5813	(0.9486)	-0.5181	(0.2764)	-0.6268	(-0.1364)
X_3	0.5526	(0.9018)	0.8248	(0.4401)	-0.1775	(-0.0386)
Var(λ_i)	2.6634		0.2847		0.04738	
% of variation	89 (approx.)		9 (approx.)		2 (approx.)	

So the three principal components are:

$Z_1 = 0.5973\ x'_1 + 0.5813x'_2 + 0.5526\ x'_3$;

$Z_2 = -0.2264\ x'_1 – 0.5181x'_2 + 0.8248\ x'_3$ and

$Z_3 = 0.7587\ x'_1 – 0.1364x'_2 – 0.0386\ x'_3$.

If we retain all these three components we will get the estimate similar to the OLS estimates. Now we decide the number of principal components to be retained in the analysis. Table 4 shows the co-efficient of correlation between the first principal component Z_1 and X_1, X_2 and X_3 are quite large. In this way, the correlations between Z_2 and X_1, X_2 and X_3 are also, to some extent, reasonable but the relationships between Z_3 and X_1, X_2 and X_3 are not very strong. It means that the first two principal components are sufficient to describe the maximum variation in X's.

We see that all the coefficients (loadings) of the first principal component are significant. Only third co-efficient of the second principal component is significant and not even a single co-efficient of the third principal component is significant. Now we statistically conclude the number of principal components to be retained. Let we test H_o: $\lambda_2=\lambda_3$; $\chi^2_c=10.007$; $\chi^2_{(0.95,2)} = 5.99$. We can well reject the null hypothesis of the equality of the second and the third principal component. We cannot exclude the second and third component at the same time, and retain the first two components in the analysis. The above discussion reveals that the third principal component is not beneficent for the analysis. The first principal component explains 89% of the total sample variance. The first two principal components collectively explain 98% of the total sample variance. Consequently, sample variation is summarized very well by two principal components. Third principal component is not beneficent for the analysis. So, we estimate the regression co-efficients by assuming that $a'_3\beta=0$. If it is true the new estimates will be more precise than the OLS estimates. Let we test H_o: $a'_3\beta=0$ i.e. $0.7587\beta_1-0.6268\beta_2-0.1775\beta_3=0$; $F_c= 0.8048$; $F_{(0.98,1,10)}= 4.96$ the hypothesis is rejected. So, the estimate of β's by using principal components regression (which is now equivalent to the restricted least square estimates) are

$$\beta^* = \begin{vmatrix} 0.6545 \\ 0.8403 \\ -0.170 \end{vmatrix}, \quad V(\beta^*)= \begin{vmatrix} 4.004578194\times10^{-1} & 6.291876605\times10^{-4} & 0.014897414 \\ & 0.103504743 & -0.362882850 \\ & & 1.345366520 \end{vmatrix}$$

The fitted model using the principal components regression is

$\hat{Y}_i = 0.6545X_1+0.8403X_2-0.1700X_3$ with $V(\beta^*_1)= 4.004578194x10^{-1}$

$V(\beta^*_2)= 0.103504743$; $V(\beta^*_3)= 1.34536652$

On comparing $V(\beta\hat{})$ and $V(\beta^*)$, we come to know that by using the principal components regression, there is a fall of 31.3% in the variance of $\beta\hat{}_3$, 80.05% in the variance of $\beta\hat{}_2$ and 95.8% in the variance of $\beta\hat{}_1$.

CONCLUSIONS

The principal components and their co-efficients (loadings) are obtained by using the correlation matrix of the regressors. All the loading of the third principal component are insignificant. Moreover, the correlation between original variables X's (standardized) and the third principal component are insignificant. We, therefore, exclude the third principal component from the analysis and retain only the first two components. The model is re-estimated using the first two principal components and explicitly assuming that $a_3\beta=0$. There is a slight decrease of 0.7% in the co-efficient of determination due to the use of principal components regression. Looking at the results, we observe that the principal components regression technique provides the best estimates of the co-efficients of the population regression function, in particular when the sample data are suffering from multicollinearity. If the original variables are uncorrelated then there is no use of the principal components analysis. Multicollinearity, if present among the regressors, seriously affect the property of minimum variance of the OLS estimates. If the purpose is just of the forecasting or prediction, then the existence of multicollinearity does not harm any more but if the aim is to get the precise estimates, some alternative ways should be adopted. Of many other solutions to the incidence of multicollinearity, the principal components regression is better due to its advantages.

References

Bartlet, (**1954**)

Butler, N.A. Denham, M.C. (**2000**) "The peculiar shrinkage properties of partial least squares regresson", *J.R. Statist. Soc*., B 62(3), 585-593.

Burt, C. and Banks, C. (**1947**) "A factor analysis of body measurement for British adult males", *Ann. Eugen.*, 13, 238-256.

Fareebrother (**1972**) "Principal component estimators and minimum mean squares error criteria in regression analysis", *Rev. of Econ. and Statistics*, 54, 332-336.

Fomby, T.B. and Hill, R.C. (**1978**) "Multicollinearity and the value of a priori information", *Comm. in Statist.*, A 8, 477-486.

Frank, I., Friedman, J. (**1993**) "A statistical view of some chemometrics regression tools", *Technometrics*, 35, 109-135.

Greenberg, E. (**1975**) "Minimum variance properties of principal components regression", *JASA* 70, 194-197.

Helland, I. (**1990**) “Partial least squares regression and statistical models” *Scan. J. Statist.*, 17, 581-607.

Helland, I. and Almoy, T. (**1994**) “Comparison of prediction methods when only a few components are relevant”, *JASA* 89, 583-591.

Hill, P.C., Fomby, T.B. and Johnson, S.R. (**1977**) “Components selection norms for principal components regression”, *Comm. in Statist*. A 6, 309-333.

Johnson, S.R., Remier, S.C. and Rothrock, T.P. (**1973**) “Principal components and the problem of multicollinearity”, *Metroeconomica* 25, 306-317.

Klein, L.R. and Goldberg, A.S. (**1964**) “An Economic Model of the U.S.A.”, North Holland Publishing Company, Amsterdam, p. 131.

Massey, W.F. (**1975**) “Principal components regression in explanatory statistical research”, *JASA* 60, 234-256.

Soderstrom, T. and Stoica, P. (**1995**) “Emitter waveform estimation in array signal processing”, *Int. J. Control*, 61, 965-80.