International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

An Alternative Approach to Estimate the Coefficients for Non-Orhogonal Data

B. Saikia[1], R. Singh[2]

[1]Research Scholar, Department of Statistics, North Eastern Hill University, Permanent Campus, Mawkynroh-Umshiing, Shillong 793022

[2] Associate Professor, Department of Statistics, North Eastern Hill University, Permanent Campus, Mawkynroh-Umshiing, Shillong 793022

Abstract: ***Multicollinearity can cause serious problem in estimation and prediction, increasing the variance of least squares estimators of the regression coefficients and tending to produce least square estimates that are too large in absolute value. This paper shows how the dynamic programming algorithm can be used to estimate the regression coefficient. It also shows how the algorithm can be used to calculate various statistical quantities like the coefficient of determination*** R^2***, the F-statistics and the t-statistics.***

Keywords: Dynamic programming, Least Square, Multicollinearity

1. Introduction

Multiple linear regression is a widely used statistical technique which allows estimating parameters that describe the distribution of a dependent variable with the help of a number of explanatory variables. The least squares (LS) solution gives stable estimates and useful results in case of independent coefficients. We often come across cases where the explanatory variables are nearly collinear. This condition is called multicollinearity now-a-day and is one of the most often encountered in econometrics dealing with several explanatory variables. The major problem with multicollinearity is that it leads to estimates with inflated variances in the estimation of regression coefficients and thus unacceptably large prediction intervals. High estimated variances (and therefore high estimated standard errors) also mean small observed test statistics. That is, the researchers will accept too many null hypotheses. Estimates of standard errors and parameters tend to be sensitive to change in the data and the specification of the model. In addition, the LS estimates are usually inflated with wrong signs – though they remain the best linear unbiased estimates (BLUE). It is noted that, if the aim of the researchers is to generate forecasts and if it is assumed that the multicollinearity problem will not be different for the forecast period, then multicollinearity is considered not to be a problem at all. This is because multicollinearity will not affect the forecasts of a model but only the estimation of the coefficients (Koutsoyiannis, 1977). In order to detect the presence of collinear variables many diagnostics have been proposed in the literatures, for instance, the condition number, variance inflation factor, variance decomposition proportion and others. Various methods exist in the literatures to tackle the problem of multicollinearity such as principal component regression (PCR), ridge regression (RR), partial least squares regression (PLSR), generalized inverse regression (GIR) and others. Saikia and Singh (2014) demonstrated a comparative study of the two methods, namely PCR and RR to tackle the problem of the multicollinearity and concluded about the superiority of PCR than RR. The property of being non-orthogonal PCs makes PCR superior than RR.

Kalaba, Natsuyama and Ueno (1999) introduced a new dynamic programming (DP) approach to the least square problems. This algorithm relied heavily on knowing the rank of the given matrix and the columns which are linearly independent. Later the algorithm has been developed in such a way that it overcomes the mentioned restrictions. The DP algorithm is seemed to be helpful to estimate the coefficients even in the presence of multicollinearity. This establishment constitutes of following two functions: the square of the length of the current discrepancy and the square length of the current solution vectors. The mentioned cost functions should be minimized synchronously by optimally choosing the minimum length vector solution. Application of Bellman to each of the cost function and usage of the functions' semi definite quadratic form subsequently leads to recurrence relation for DP algorithm.

2. Methodology

The two cost functions are introduced in following way:

$f_k(b)$ = the smallest square of the length of the vector

$$A_k x^k - b$$

$g_k(b)$ = the smallest square of the length of the vector x^k

where x^k is the subject to the restriction

$$\left| A_k x^k - b \right| = \min \qquad (1)$$

for k = 1, 2, …, n. The k[th] column of the matrix A is denoted by a_k and the matrix A_k consist of the first k columns of A, and x^k is a k-dimensional vector. The vector x^n is the optimal vector.

Using the principle of optimality leads to the recurrence relation k = 1, 2, …, n for two cases. Firstly it depends on if the columns A are linearly dependent. Secondly, if the columns A are independent. Consider that a_k is linearly dependent on $a_1, a_2, \ldots, a_{k-1}$. Then

$$f_k(b) = f_{k-1}(b) \qquad (2)$$

As, a linear combination of $a_1, a_2, \ldots, a_{k-1}, a_k$ cannot be brought closer to b than the linear combination of $a_1, a_2, \ldots, a_{k-1}$ is,

$$g_k(b) = \min_{x_k} [x_k^2] + g_{k-1}(b - x_k a_k) \quad (3)$$

In the similar way, if the vector a_k is independent then the following equations are used as mentioned below,

$$f_k(b) = \min_{x_k} (b - x_k a_k) \quad (4)$$

$$f_k(b) = (x_k^*)^2 + g_{k-1}(b - x_k^* a_k) \quad (5)$$

where x_k is the k^{th} component of the vector x^k. Here the minimization is to be performed over k^{th} component x^k and the minimum value is denoted by an asterisk, x_k^*. There are initial conditions for those functions and they are used to derive the recurrence relations. Observe the argument (b - $a_k x_k$); it represents the new target vector at stage k - 1. For the case, when k = 1, the smallest square of the length of the vector is as

$$f_1(b) = \min_{x_1} (a_1 x_1 - b)^T (a_1 x_1 - b) \quad a \neq 0 \quad (6)$$

Thus,

$$f_1(b) = \min_{x_1} (a_1^T a_1 x_1^2 - 2a_1^T b x_1 + b^T b) \quad (7)$$

Then, differentiation with respect to x_1 yields the optimality condition as

$$a_1^T a_1 x_1 - a_1^T b = 0 \quad (8)$$

Hence,

$$x_1^* = \frac{a_1^T b}{a_1^T a_1} = a_1^+ b \quad (9)$$

where $a_1^+ = \frac{a_1^T b}{a_1^T a_1}$, assuming that $a_1 \neq 0$

Since the columns of A are assumed to be linearly independent, $a_1 \neq 0$ and minimum of $f_1(b)$ is written as

$$f_1(b) = b^T (a_1^+)^T a_1^T a_1 a_1^+ b - 2b^T a_1 a_1^+ b + b^T b \quad (10)$$

The residual vector is

$$f_1(b) = b^T [I - a_1 a_1^+] b \quad (11)$$

$$= b^T Q_1 b \quad (12)$$

where $Q_1 = [I - a_1 a_1^+]$ (13)

Suppose, Ax = b (14)

In vector form,

$$X = (a_1^T a_1)^{-1} a_1^T b \quad (15)$$

Consequently by definition, it is written as

$$g_k(b) = x_1^2 + x_2^2 + \ldots + x_k^2 = b^T a_k [(a_k^T a_k)^{-1}]^T (a_k^T a_k)^{-1} a_k^T b$$

$$= b^T a_k (a_k^T a_k)^{-1} (a_k^T a_k)^{-1} a_k^T b \quad (16)$$

Now, let us denote

$$R_k = a_k (a_k^T a_k)^{-1} (a_k^T a_k)^{-1} a_k^T \quad (17)$$

Therefore,

$$g_k(b) = b^T (a_1^+)^T a_1^T b \quad (18a)$$

Hence, we can write as

$$g_1(b) = b^T R_1 b \quad (18b)$$

Using the recurrence relation (3), we observed as

$$g_k(b) = \min_{x_k} [x_k^2 + g_{k-1}(b - a_k x_k)]$$

$$= \min_{x_k} [x_k^2 + b^T R_{k-1} b - 2b^T R_{k-1} a_k x_k + a_k^T R_{k-1} - a_k x_k^2)]$$

$$= \min_{x_k} [(1 + a_k^T R_{k-1} a_k) x_k^2 + b^T R_{k-1} b - 2b^T R_{k-1} a_k x_k] \quad (19)$$

Given that the first order condition for the minimizing the value of x_k is

$$\frac{\partial\{.\}}{\partial x_k} = 2(1 + a_k^T R_{k-1} a_k) x_k - 2b^T R_{k-1} a_k = 0 \quad (20)$$

It is observed that,

$$x_k^{OPT} = \frac{b^T R_{k-1} a_k}{1 + a_k^T R_{k-1} a_k} \quad (21)$$

or, $x_k^{OPT} = \frac{a_k^T R_{k-1} b}{1 + a_k^T R_{k-1} a_k}$ (22)

Substituting (21) or (22) in (19), we can write as

$$g_k(b) = (1 + a_k^T R_{k-1} a_k)[x_k^{OPT}]^2 - 2a_k^T R_{k-1} b x_k^{OPT} + b^T R_{k-1} b$$

$$= (1 + a_k^T R_{k-1} a_k)[\frac{b^T R_{k-1} a_k a_k^T R_{k-1} b}{1 + a_k^T R_{k-1} a_k}]^2 - 2b^T [\frac{R_{k-1} a_k a_k^T R_{k-1}}{1 + a_k^T R_{k-1} a_k}] b + b^T R_{k-1} b$$

$$= b^T R_{k-1} b - b^T [\frac{R_{k-1} a_k a_k^T R_{k-1}}{R_{k-1} a_k a_k^T R_{k-1}}] b$$

$$= b^T [\frac{R_{k-1} - R_{k-1} a_k a_k^T R_{k-1}}{1 + a_k^T R_{k-1} a_k}] b$$

$$= b^T R_K b \quad (23)$$

where, $R_k = \frac{R_{k-1} - R_{k-1} a_k a_k^T R_{k-1}}{1 + a_k^T R_{k-1} a_k}$ (24)

The denominator is noticed as

$1 + a_k^T R_{k-1} a_k \neq 0$, always holds (25)

Let, $\beta_k = R_{k-1} a_k$ (26)

(24) is equivalent to

$$R_k = \frac{R_{k-1} - \beta_k \beta_k^T}{1 + a_k^T \beta_k} \quad (27)$$

Substituting (27) into (21) or (22),

$$x_k^{OPT} = \frac{b_k^T \beta_k}{1 + a_k^T \beta_k} \quad (28)$$

$$x_k^{OPT} = \frac{\beta_k b_k^T}{1 + a_k^T \beta_k} \quad (29)$$

Secondly, if a_k is linearly independent of the vector $a_1, a_2, \ldots, a_{k-1}$ then the algorithm uses the general form of (12) to yield the scalar x^k that minimizes the current solution vector x^k as,

$$x_1^{OPT} = \alpha_k^T b_k \Big/ \alpha_{k1}^T \alpha_k \quad (30)$$

where,

$$\alpha_k = Q_{k-1} a_{k-1} \quad (31)$$

$$b_k = b - (a_1 x_1 + a_2 x_2 + \ldots + a_{k-1} x_k) \quad (32)$$

3. Measuring the Precession of Equation and Examining the Significance of Independent Variables

Lawson and Hanson (1974) discussed that the coefficient of determination R^2 is a scale free summary of the degree to which the variables $a_1, a_2, \ldots, a_{k-1}, a_k$ predict the dependent variable b.

Let us define TSS as total sum of the squared deviation of the vector b about its mean. Considering Ax = b, TSS can be split into two parts. The first part is the sum of squares, i.e., variance explained by regression equation. Another part is the sum of square which cannot be explained by regression equation.

Thus, TSS = RSS + SSE

As, R^2 is the measure of how much of the total variation is accounted for using the sum of square regression, we have,

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

The residual vector α_{n+1} is the measure of the unexplained error, where $b^T b$ gives total sum of squares.

Hence, $SSE = \alpha_{n+1}^T \alpha_{n+1}$ and

$$R^2 = [b^T b - f_k(b)] \Big/ b^T b = 1 - \alpha_{n+1}^T \alpha_{n+1} \Big/ b^T b$$

Now, considering that the residuals are normal allowing us to compute the *F*- statistics from RSS and SSE and compare it to a theoretical distribution providing a criterion to accept or reject the model. Comparing the values computed in Table 1 to the theoretical value from the F-tables for any chosen level of significance gives us a decision rule to reject or accept the hypothesis that the model fits the data.

Table 1: ANOVA for the F-test

Source	df	SS	MSS	F-ratio
Regression	n - 1	$b^T b - \alpha_{n+1}^T \alpha_{n+1}$	$\frac{b^T b - \alpha_{n+1}^T \alpha_{n+1}}{n-1}$	$\frac{b^T b - \alpha_{n+1}^T \alpha_{n+1}}{n-1} \Big/ \frac{m-n}{\alpha_{n+1}^T \alpha_{n+1}}$
Error	m - n	$\alpha_{n+1}^T \alpha_{n+1}$	$\frac{\alpha_{n+1}^T \alpha_{n+1}}{m-n}$	
Total	m - 1			

After examination of the degree of fit between the model as a whole and the data we inspect whether each individual predictor variable x_k contributes significantly to the regression equation. When the coefficient of the variable examined is statistically close to zero, then it can be concluded that the variable does not contribute significantly to the regression equation. The counterpart of the formal statistical test for this work is called the t-test.

Now, the standard deviation of the coefficient for each predictor variable is required to estimate for the construction of the test. Johnson and Kalaba (2003) derived the test by noting that the t-distribution is closely related to the F-distribution because the αQβR algorithm does not provide the standard deviation. The exact distribution of the square of a t-distribution with m-degrees of freedom (t_m^2) is an F-distribution with degrees of freedom 1, m, namely, $F_{(1, m)}$ (Casella and Berger, 1990).

The DP algorithm provides sufficient information for computing $F_{(1, m)}$. Our interest is in computing t-value of the coefficient of the predictor a_i. Now RSS is compared for the full model with all the n predictors in it with the RSS of the model without predictor a_i. The difference of two RSS gives the reduction in sum of squares that is achieved by including the predictor a_i. The ratio of the mean squared error from the model with and without the predictor gives as the required F-value (Casella and Berger, 1990). Now

$$F_{(1,\, m-n)} = \frac{RSS_{n-1} - RSS_n}{\frac{RSS_n}{m-n}} \quad (33)$$

where $\frac{RSS_n}{m-n}$ = variance for the full model.

The required t-value for the coefficient of the predictor a_i is computed by taking the square-root of (33). Thus,

$$t_i = \sqrt{\frac{RSS_{n-1} - RSS_n}{\frac{RSS_n}{m-n}}} = \sqrt{\frac{\alpha_{n-1}^T \alpha_{n-1} - \alpha_n^T \alpha_n}{\frac{\alpha_n^T \alpha_n}{m-n}}} \quad (34)$$

where t_i = calculated t-value for the coefficient of a_i, α_{n-1} = residual vector from the model without predictor i, α_n = residual vector from the model with all the predictors.

Table 2: ANOVA Table for the t-test

Source	df	SS	MSS	t-ratio
Full model	n	RSS_n		
Model without predictor x_k	n - 1	RSS_{n-1}		
Predictor	1	RSS_{n-1} - RSS_n	RSS_{n-1} - RSS_n	$\frac{RSS_{n-1} - RSS_n}{\frac{RSS_n}{m-n}}$
Error (Full Model)	m - n	RSS_n	$\frac{RSS_n}{m-n}$	

Proceeding in the same way, the t-value for the coefficient of each variable is sequentially derived by dropping each one and retaining the others. The calculated t-value is compared against the theoretical t-value with any desired level of significance and a decision to accept or reject is assessed through contribution of each variable to the regression model. Table 2 shows how the t-value for any predictor k is computed by analyzing the change in variance of the model with and without the predictor k.

4. Conclusion

In this paper it has been discussed that, how the DP algorithm can produce the coefficients of least squares. Advantage of using this algorithm is that even in the presence of multicollinearity one can estimate the coefficients and those can be applied for future and related research works. This area is less developed till now, more and frequent research works are required to be done in this area for its development. Various standard statistical measures like the coefficient of determination R^2, t-and F-statistics and others can be computed using the DP algorithm.

References

[1] Casella, G. and Berger, R. L. (1990), "Statistical Inference", Duxbury Press, California.
[2] Johnson, J. and Kalaba, R. E. (2003), "Statistical Measures for Ordinary Least Squares Using the αQ Algorithim", Journal of Optimization theory and Applications, Vol. 117 (3), pp. 461 - 474.
[3] Kalaba, R.E., Natsuyama, H., Uneo, S. (1999), "The Estimation of Parameter in Time Dependent Transport Problems: Dynamic Programming and Associative Memories", Computers and Mathematics with Applications, Vol. 37, 41 - 45.
[4] Koutsoyiannis, A. (1984), "The Theory of Econometrics", Macmillan Publishing, NY.
[5] Lawson, C. and Hanson, R. (1974), "Solving Least Square Problems", Prentice-Hall Englewood Cliff, NJ.
[6] Saikia, B. and Singh, R. (2014), "A Comparative Study on Countermeasures for Handling Multicollinearity in Regression Analysis", Asian-African Journal of Economics and Econometrics, Vol. 14 (2), 163 - 174.