ASR Vol. 17 No. 2 September 2003, pp 145-152

TESTING THE STABILITY OF REGRESSION AND THE ROLE OF GENERALIZED DUMMY VARIABLE TECHNIQUE

D. DUTTA
D. D. R. College, Chabua
Dibrugarah, Assam
and
S. C. KAKATY
Department of Statistics
Dibrugarh University

1. INTRODUCTION

In this paper the generalized dummy variable technique is dissussed to test the stability of regression. An attempt is made in this paper how dummy variable technique can be used to study the spatial effetcs arising in regression. An experiment has been conducted for a period of one month in a broiler firm collecting informations on two independent variables – initial age and initial weight of broilers and dependent variable – the average daily grain of weight under four treatments and a numerical analysis has been made here to demonstrate the generalized dummy variable approach. The investigation proceeds in the line of the works by Gujarati (1970).

2. Objective of the Study

The main objective of this paper is to test stability of coefficients in linear regression unsing generalized dummy variable technique. The method of generalized dummy variable technique is essentially a device for analysis of more than two linear regressions in more than two variables. Secondly, our aim is to investigate spatial effect over treatment arising in regression analysis. The technique to study stability of regression coefficients has found increasing in many econometric analysis and social phenomenon.

The subsequent discussion proceeds as follows : Section 3 is devoted to a brief discussion of methodology of general dummy variable technique with the process of data collection. A numerical example based on the data generated by an experiment conducted by ourselves in field had been discussed in section 4. Section 5 is devoted to discussion of proposed invstivation.

3. Methodology

In this section generalized dummy variable technique has been demonstrated by considering 'n' independent variables $X_1, X_2, \ldots\ldots, X_n$ and one dependent variable Y over k treatments incorporating k-1 dummy variables. The model is presented as follows by considering Y linearly related to $X_1, X_2, \ldots\ldots, X_n$.

$$y_{ij} = \alpha_0 + \sum_{i=1}^{k-1} \alpha_{0i} D_i + \alpha_1 X_{1j} + \sum_{i-1}^{k-1} \alpha_{1i} D_i X_{1j} + \ldots\ldots + \alpha_n X_{nj} + \sum_{i-1}^{k-1} \alpha_{ni} D_i X_{nj} + U$$

where,

D_1 = 1, if the observation lies in the second group
= 0, otherwise

D_2 = 1, if the observation lies in the third group
= 0, otherwise

D_3 = 1, if the observation lies in the fourth group
= 0, otherwise

D_{k-1} = 1, if the observation lies in the k^{th} group

α_0 = Intercept for group 1.

α_{10} = differential Intercept for group 2.

α_{0i} = differential Intercept for group i + 1.

α_1 = partial regression coefficient of Y w.r.t. X_1 for group 1.

α_1 = partial regression coefficient of Y w.r.t. X_1 for group 1.

α_{1i} = partial regression coefficient of Y w.r.t. X_1 for group i + 1.

α_{ni} = partial regression coefficient of Y w.r.t. X_n for group i + 1.

The generalized model above reveals the following :

(1) If α_{01}, α_{02}, ..., α_{ok-1}; α_{11}, α_{12}, ..., α_{1k-1}, ..., α_{n1}, α_{n2}, ..., α_{nk-1}, are statistically significant, there is a signal of shifts in regression over k treatments, otherwise there will be no shifts in regression, viz. common regression for all the treatments.

(2) If α_{01}, α_{02}, ..., α_{ok-1} are statistically significant, there is a signal of shifts in intercept, no shifts in regression coefficients.

(3) α_{11}, α_{12}, ..., α_{1k-1}, ..., α_{nl}, α_{n2}, ..., α_{nk-1} are statistically significant, there is a signal of shifts of regression through regression coefficients.

a numerical example based ontwo independent variable X_1 (initial age) and X_2 (initial weight) and one dependent variable Y (averge daily gain of weight) has been considered to demonstrate the generalized dummy variable approach. The observations on the variables have been conducted by ourself in a broiler firm (Dutta Broiler Firm, Chabua, Dibrugarh, Assam). The data are divided into four treatments and related to the average daily gain weight with intial age and initial weight. Let us postulate the model where Y is linearly related to X_1 and X_2 as follows :

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + U_j \qquad j=1, ..., N \qquad ... \qquad (2)$$

Our investigation is of four groups, we are interested in finding out whether (2) differs from group to group. For this purpose of illustration rewrite (2) as follows :

$$Y_{1j} = \beta_i + \beta_{i1} X_{1j} + \beta_{i2} X_{2j} + U_{ij} \qquad ... \qquad (3)$$

$i = 1, 2, 3$ and 4, $\quad j = 1, ..., N$

Again equation (3) consist of the following set of equations in the different four treatment as follows :

$$Y_{1j} = \beta_1 + \beta_{11} X_{1j} + \beta_{12} X_{2j} + U_{1j} \qquad j = 1, ..., n_1 \qquad ... \qquad (4)$$

$$Y_{2j} = \beta_2 + \beta_{21} X_{1j} + \beta_{22} X_{2j} + U_{2j} \qquad j = 1, ..., n_2 \qquad ... \qquad (5)$$

$$Y_{3j} = \beta_3 + \beta_{31} X_{1j} + \beta_{32} X_{2j} + U_{3j} \qquad j = 1, ..., n_1 \qquad ... \qquad (6)$$

$$Y_{4j} = \beta_4 + \beta_{41} X_{1j} + \beta_{42} X_{2j} + U_{4j} \qquad j = 1, ..., n_1 \qquad ... \qquad (7)$$

$$N = (n_1 + n_2 + n_3 + n_4)$$

The aim of the present work is to find out the stability of regressions (4) through (7),

To present how the dummy variable technique has been applied in different treatments, let us first consider the equations (4) through (7) differ from one another. For the purpose of this assumption, consider the following equation

$$Yj = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 \alpha_3 D_3 + \alpha_4 D_{1j} + \alpha_5 (D_1 X_{1j}) + \alpha_6 (D_2 X_{1j}) + \alpha_7 (D_3 X_{1j}) + \alpha_8 X_{2j} + \alpha_9 (D_1 X_{2j}) + \alpha_{10} (D_2 X_{2j}) + \alpha_{11} (D_3 X_{2j}) + U_j \quad \dots \quad (8)$$

$j = 1, \dots, (n_1 + n_2 + n_3 + n_4)$

Where,

D_1 = 1, if the observation lies in the second group
= 0, otherwise

D_2 = 1, if the observation lies in the third group
= 0, otherwise

D_3 = 1, if the observation lies in the fourth group
= 0, otherwise

α_0 = Intercept for group 1.

α_1 = differential Intercept for group 2.

α_2 = differential Intercept for group 3.

α_4 = partial regression coefficient of Y w.r.t. X_1 for group 1.

α_5 = differential partial regression coefficient of Y w.r.t. X_1 for group 2.

α_6 = different partial regression coefficient of Y w.r.t. X_1 for group 3.

α_7 = different partial regression coefficient of Y w.r.t. X_1 for group 4.

α_8 = partial regression coefficient of Y w.r.t. X_2 for group 1.

α_9 = different partial regression coefficient of Y w.r.t. X_2 for group 2.

α_{10} = different partial regression coefficient of Y w.r.t. X_2 for group 3.

α_{11} = different partial regression coefficient of Y w.r.t. X_2 for group 4.

Using above differential intercept and differential partial regression coefficient, the actual values of intercept and partial regression coefficients can be determined as follows :

Group 1 : $Y_{1j} = \alpha_0 + \alpha_4 X_{1j} + \alpha_8 X_{2j}$ ... (9)

Group 2 : $Y_{2j} = (\alpha_0 + \alpha_1) + (\alpha_4 + \alpha_5)X_{1j} + (\alpha_8 + \alpha_9)X_{2j}$ ... (10)

Group 3 : $Y_{3j} = (\alpha_0 + \alpha_2) + (\alpha_4 + \alpha_6)X_{1j} + (\alpha_8 + \alpha_{10})X_{2j}$ ... (11)

Group 4 : $Y_{4j} = (\alpha_0 + \alpha_3) + (\alpha_4 + \alpha_7)X_{1j} + (\alpha_8 + \alpha_{11})X_{2j}$ ... (12)

The method of ordinary least squares (OLS) is the proper device to estimate the equation (8), from which the equation (9) through (11) can be obtained. It is betteer to present the usual assumptions of error term U_j, namely $E(U_j) = 0$, $E(U_j X_{ij}) = 0$ and $E(U_j\ U_{j+k}) = \sigma$ for $k = 0$ and zero if $k \neq 0$.

To test the stability of regressions, in other sense, whether the sets of linear regressions are different, one has to test the statistical significance of estimated differential intercept and partial regression codffieients on the basis of "t" lest. If every differential intercept and partial regression coefficients in (8) are found statistically insignificant, then equation (9) gives the common regression for all the groups. This hypothesis can be interpreted as we can pool all the observations in various groups and run one regression, because the variable treatments or groups have no effect on the relationship of Y to X_1 and X_2. It is not our area of study whether the common regression Y i s significanttly related to X_1 and X_2 or not. We would like to get the answer through investigation, whether the intervening variable "treatment" or "group" have any indepent effect on the relationship of X_1 and X_2. In other words, our intention is to see whether the spatial effect (or treatment effect) occurs in the present relations of Y to X_1 and X_2.

4. Numerical Example

In table 1, the data have been placed in different treatments, which have been collected from daily observation of fourty broilers in a firm (Dutta Broiler Firm, Chabua, Dibrugarh, Assam) for a period of one month. the broiler have been classified in four groups including ten in each group feeding with different treatment to present the role of generalized dummy variable technique in our investigation.

TABLE 1

Initial age in days (X_1), Initial weight in grams (X_2), and Rate of gain in grams per day (Y) of 40 (Fourty) Broilers.

Treatment 1			Treatment 2			Treatment 3			Treatment 4		
Y	X_1	X_2	Y	X_1	X_2	Y	X_1	X_2	Y	X_1	X_2
37.33	10	221	44.83	19	510	43.67	29	923	49	38	1315
30.5	11	218	38.33	21	505	50.33	31	982	50.83	40	1212
36.83	14	372	39.83	23	550	45.83	33	1020	43.67	37	1100
32.67	16	400	33.33	18	453	40.83	35	1170	66.17	39	1295
30.67	12	275	35.67	20	490	45.33	32	960	52	42	1517
27.17	15	340	38.33	22	577	43.67	34	1082	73	41	1260
30.50	11	253	31.33	24	540	37.33	30	830	42.6	43	1307
28.83	13	302	29.50	18	433	38.67	34	1200	77.5	40	1585
31.5	16	415	33.17	21	517	53.17	31	1000	61.67	38	1388
30	10	210	31.67	22	480	50.83	29	835	43.33	37	1110

Source : Dutta Broiler Firm, Chabua, Dibrugarh, Assam

The regression results for the above example are as followos :-

$$Y_J = 47.736143 - 33.079389\ D_1 + 34.424522\ D_2 - 20.64998\ D_2 \quad 4.297452\ X_{ij}$$

$$(32.596080)^* \quad (47.854658)^* \quad (42.513576)^* \quad (4.087626)^*$$

$$+ 2.257697\ (D_1 X_{ij}) + 2.768827\ (D_2 X_{ij}) + 3.289866\ (D_3\ X_{ij}) \quad + 0.129312\ X_{2j}$$

$$(4.4442660^* \quad (4.775196)^* \quad (4.279355)^* \quad (.123989)$$

$$- 0.003953\ (D_1 X_{2j}) - 0.117899\ (D_2 X_{2j}) - 0.076833\ (D_3 X_{2j})$$

$$(0.148229)^* \quad (0.131088)^* \quad (0.125111)^* \quad \ldots \quad (13)$$

Multiple = 0.87352; DF = 28

*Statistically insignificant at the 5 per cent level.

The regressions for individual groups in equation (9) through (12) can be interpreted with the help of above results as follows.

$$\text{Group 1 : } Y_{ij} = 47.736143 - 4.297452\ X_{ij} + 0.129312\ X_{2j}$$

$$(4.087626)^* \quad (0.123989)^* \quad \ldots \quad (14)$$

$$\text{Group 2 : } Y_{2j} = (47.736143 - 33.079389) + (-4.297452 + 2.2576970)\ X_{ij} + (0.129312 - .003953)\ X_{2j}$$

$$= 14.656754 - 2.039755\ X_{1j}\ 0.125359\ X_{2j} \quad \ldots \quad (15)$$

$$\text{Group 3} : Y_{3j} = (47.736143 - 43.424522) + (-4.297452 + 2.768827)X_{1j} + (0.129312 - 0.117899)\,X_{2j}$$

$$= 82.160665 - 1.528625\,X_{1j} + 0.011413\,X_{2j} \quad \ldots \quad (16)$$

$$\text{Group 3} : Y_{4j} = (47.736143 - 20.64998) + (-4.297452 + 3.289866)X_{1j} + (0.129312 - 0.076833)\,X_{2j}$$

$$= 27.086145 - 1.007586\,X_{1j} + 0.052479\,X_{2j} \quad \ldots \quad (17)$$

5. Discussion

From equation (13), ir has been obseved that the differntial partial regression coefficients entering into equation (8) are statistically insignificant. It implies that, we should have only one regression, that is equation 14), which is common to all the groups. it has been found that in equation (14) Y is not signicantly related to X_1 and X_2, which is beyond of our discussion has been pointed out earlier. Our major objective is to search weather a given regression model differs from group to group.

Moreover dummy variable technique is so flexble, it is not necessary to differentiate each adn every coefficient. Again to assess relative importance of the variables entering into equation (8), we run a stepwise regression program to know the variable contribution. As in stepwise regression program the variable which contributes most to the explained sum of squares enters first, followed by the variable, whose contribution is next highest and so on.

TABLE 2

ANOVA of Numerical Example

[Equation (8), Data of table 2]

Source of variation	Sum of squares	D.F.	Mean Squares
Due to variable X_2	20349.476	1	20349.476
" " " D_3X_1	13702.329	1	13702.329
" " " D_2X_1	1824.9976	1	1824.9976
" " " D_1	1597.9826	1	1597.9826
" " " D_2	755.44614	1	755.44614
" " " D_1X_2	92.937188	1	92.937188
" " " X_1	0	1	0
" " " D_3	0	1	0
" " " D_1X_1	0	1	0
" " " D_2X_2	0	1	0
" " " D_3X_2	0	1	0
" " " Regression	383223.167	11	3483.9242
" " " Residual	1314.5939	28	46.949782
Total	39633.76	39	1016.3528

Having observed the table 3, we can assess the relative importance of the variables. With this point of view, we should retain the variables X_2, D_3X_1, D_2X_1 and D_1 considerably. The revised regression results of the model with these variables are as follows :

$$Y_j = 23.481148 - 2.014192\,D_1 - 0.212666\,D_2X_1 - 0.103694\,D_3X_1 + 0.02795\,X_2$$
$$\qquad (4.176412)^* \quad (0.30391)^* \quad (0.339859)^* \quad (0.012856)^{**} \quad (18)$$

Multiple R = 0.82374

D. F. = 35

* Insignificant at the 5 percent level

** Significant at the 5 percent level

In equation (13), we have found that none of the codfficients are statistically significant. On the other hand, in equation (18), Y is significantly related to X_2, in which the estimated coefficient of X_2 does not differ from group to group.

As evidenced by the D-W statistics, d = 1.78, calculated for equation (13) it appears that, the model free of serial correction. The dummy variable approach indicates here that, a common regression will be valid for all the treatments.

Dummy variable approach is extensively used in Econometric analysis. Many applications have been demonstrated in the literature of Gujarati (1972) Comanor and Wilson (1971), Halvorsen and Palmquist (1980) etc. incorporating dummy variable. Dummy variable has an advantage over Chow test, it can be extended over time and space and the analysis can be made by simply using a single equation to see wheather there are shifts in the parameters either over time or over space. It also reduces the problem of degrees of freedom what is usually seen in other approaches.

REFERENCES

1. Comanor, W. S. and wilson A. Thomas (1971) : "*On Advertising and Profitability*", Review of Economics and Statistics, Vol. 53, pp. 408-410/
2. Gujarati Damodar (1970) : "*Use of Dummy Variables in Testing for Equa;ity between sets of coeffieients in Linear Regression*", the american Statistician, Vol. 24, No 5, December, pp. 18-22.
3. Gujarati Damodar (1972) : "*The Behaviour of Unemployment and unfilled Vacancies : Great Britain, 1958-1971*", Economic journal, Vol. 82, March, pp. 195-201.
4. Halvorsen Tobert and palmquist Raymond (1980) : "*The Interpretation of Dummy Variables in Semilogarithmic Equations*" American Economic Review, Vol. 70, No. 3, pp. 474-575.