"Dimitrie Cantemir" Christian University

Knowledge Horizons - Economics

Volume 6, No. 3, pp. 100–102
P-ISSN: 2069-0932, E-ISSN: 2066-1061
© 2014 Pro Universitaria
www.orizonturi.ucdc.ro

ANALYZING THE LOGISTIC REGRESSION OF AN ECONOMETRIC MODEL WITH QUALITATIVE VARIABLES

Carmen-Judith GRIGORESCU[1], **Raluca-Georgiana MOSCU**[2], **Ligia PRODAN**[3]

[1]*Faculty of Finance, Banking and Accountancy, Bucharest – UCDC, E-mail:* judithgrigorescu@yahoo.com
[2]*Faculty of Finance, Banking and Accountancy, Bucharest – UCDC, E-mail:* moscu.raluca@yahoo.com
[3]*Faculty of Finance, Banking and Accountancy, Bucharest – UCDC, E-mail:* prodanligia@yahoo.com

Abstract *It is our intention in this paper to highlight the role and importance of the regression analysis in the case of an econometric model with qualitative variables, proposing a model that may underline the cause relation between the factor effect-effort. The study is based on a questionnaire, carried out at the level of 3 groups of students (70 persons) of the four existing at our faculty, regarding the rate of their rate of success at the tests they have stood for during the session of examinations.*

Key words
Regression, model, variables, modelling
JEL Codes:
M16

1. Introduction

The qualitative variables or the dummy ones relate to talents, qualities, categories, etc., of which size is expressed by attributes or names. These variables, also called variables attributives, shall be divided into two categories: variables dichotomist (binary or alternative) and polytechnic variables (non-alternative).

The use of the qualitative variables in an econometric model can be summed up by three scenarios: the endogenous variable Y displays a binary qualitative nature; the exogenous variable X has a variable nature of qualitative alternative (binary or dichotomist) or non -alternative (polychotomous).

2. Logistic regression of an econometric model with qualitative variables

The logistic regression analysis aims at modelling the connection between a dependant variable, effect (endogenous, in the case of our model being noted R_1) and one or more independent variables, causal chain (exogenous, in models R_2 and R_3) to highlight form and the direction of the connection between the variables taken into consideration. The dependant variable can take only logic values, namely: Yes or No. Also, in the majority of the literature applications, the independent variables also take all logical values, such as Yes or No.

The models are the representations of objects or actual situations and can be defined as simplified representations of processes or phenomena. Modelling method is a useful tool for scientific knowledge that has as its object the construction of representations to have a better knowledge of the various scientific fields. The hypotheses of the described model must be a real life type (range) on which it is intended to apply the model. The assumptions should not be an exact representation of reality, but only its reasonable abstracting, i.e. it contains only those aspects of reality that are considered relevant.

In the event that the assumptions are realistic enough for the purpose of the analysis, even if they do not represent particularly and entirely the reality, we may draw conclusions which can be successfully applied in real life. Hypothetical models are designed for carrying out intellectual experiments, to determine the nature of important variables or are used as a criterion for assessing the current state of the system.

Therefore, assumptions may contain in certain situations hypothetical construction (abstracting reality) or refer to items which are not directly observable, being extremely useful to reach conclusions which are relevant in reality and can be used for explanation, interpretation, prediction and control. This means that not all assumptions need to relate or to comply with some observable elements. An assumption in a theory may not be at the same time a conclusion of another theory of higher level. This hierarchical structure of models is in fact the grounds on which Economics, as a subject, has been systematically developing.

Obtaining the conclusions based on the assumptions of a model is a deductive process in which the logical aspects of realism or the empirical truth of the assumptions or conclusions are irrelevant. Therefore, the logic test of consistency, the validity of our assumption does not guarantee the truth of empirical conclusion or its meaning.

The model approached in this paper is a Multy-Input-Single-Output (MISO). In such a context, the model variables are:

R_1 = "Have you passed the examinations this session ?", exogenous variable and it takes the value 1, affirmatively, or 0, contrarily;

R_2 = „Have you attended at least half of lectures and seminars?", first exogenous variable and it takes the value 1, affirmatively, or 0, contrarily.;

R_3 = "Have you had didactic supporting material taught during lectures or seminars?", the second exogenous variable and it takes the value 1, affirmatively, or 0, contrarily;

R_4 = „Have you acquired any didactic material for self-study?", third exogenous variable and it takes the value1, affirmatively, or 0, contrarily;

R_5 ="Have you spent time to study?", the fourth exogenous variable and it takes 1, affirmatively, or 0 contrary;

As for more independent variables, the general model is:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_k x_k ,$$

where: p is $P(y = 1/x_1, x_2, ..., x_k).$

The equivalent exponential form can immediately be obtained.

The interpretation of β_i coefficients, is obvious: the increase of the quantity logit (logarithm of OR –odds report, chances ratio) when x_i increases by a unit (the other x variables staying constant). OR represents the ratio between the successful probability and the unsuccessful probability, thus: OR = $\frac{p}{1-p}$

For a complex interpretation, we may write the model as follows:

$$P(y = 1/x_1, x_2, ..., x_k). = \frac{\exp\ (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_k x_k)}{1 + \exp\ (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_k x_k)}$$

We get: $\exp\ (\beta_0) = \frac{P\ (y = 1/x_1 = x_2 = ... = x_k = 0)}{1 - P\ (y = 1/x_1 = x_2 = ... = x_k = 0)} = \frac{P\ (y = 1/x_1 = x_2 = ... = x_k = 0)}{P\ (y = 0/x_1 = x_2 = ... = x_k = 0)}$

Which implies that OR is a base, $x_1 = x_2 = ... = x_k = 0.$

For β_i coefficients , we may have the following values :s

$$\exp\ (\beta_i) = \frac{P(y = 1/x_i = 1, x_j = 0\ pentru\ \ j \neq i)}{1 - P(y = 1/x_i = 1, x_j = 0\ pentru\ \ j \neq i)} \times \frac{1}{OR_{baza}} = \frac{OR_{x_i=1, x_j=0\ pentru\ \ j\neq i}}{OR_{baza}} .$$

Thus, regarding the multiplicative character of the logistic model, namely:

$OR_{x_1, x_2, ... x_k} = \exp\ (\beta_0) \times \exp\ (\beta_1 x_1) \times ... \times \exp\ (\beta_k x_k),$

each β_i expresses the contribution of factor x_i when explaining the probability (under the form OR) of generating the event y=1. Thus, setting up $x_i = 1$, exp(β_i) will represent the multiplicative constant factor, no matter what the other independent value may be.

The estimation of our model parameters through the econometric soft Eviews will generate the following output:

Dependent Variable: R1; Method: ML - Binary Probit (Quadratic hill climbing)

Sample: 1 70; Included observations: 70
Convergence achieved after 34 iterations

WARNING: Singular covariance - coefficients are not unique Covariance matrix computed using second derivatives

WARNING: Quasi-complete separation detected at estimated parameters (results may not be valid)

Table 1.

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-6.477132	NA	NA	NA
R2	-1.169674	NA	NA	NA
R3	-1.035098	NA	NA	NA
R4	7.637104	NA	NA	NA
R5	7.541954	NA	NA	NA

McFadden R-squared	**0.337555**	Mean dependent var	0.500000
S.D. dependent var	0.503610	S.E. of regression	0.415279
Akaike info criterion	1.061200	Sum squared resid	11.20969
Schwarz criterion	1.221807	Log likelihood	-32.14201
Hannan-Quinn criter.	1.124995	Deviance	64.28402
Restr. deviance	97.04061	Restr. log likelihood	-48.52030
LR statistic	32.75658	Avg. log likelihood	-0.459172
Prob(LR statistic)	0.000001		
Obs with Dep=0	35	Total obs	70
Obs with Dep=1	35		

3. Conclusions

“Experiment” is often used in statistics to refer to any activity for which the outcome or final condition of an activity or action cannot be specified in advance, but for which it can be identified lots containing all possible results of that activity. Before analysing the possibility of a particular result of an experiment, it is necessary to identify possible results to be obtained. This situation leads to define statuses of a space experiment. The results of several experiments are expressed inherently under the form of real numbers. Statistics R-squared measures "success" by means of which the regression equation estimated manages to explain the value for the variable dependent (in our case, the rate of graduated) in a sample. Normally, this statistic can be interpreted as the proportion of the variant’s dependent variable explained by independent variables.

Statistics is equal to 1 in the case in which the regression equation fits perfectly and zero, in the case in which it does not fit better than the dependent average variable. A major problem in terms of the use of statistics R-squared, as a measure of the "matching" model to available data, refers to the fact that the value of this statistics does never decrease while several regressors are added. Thus, in extreme cases, we can obtain a statistics equal to 1 if there are included as many independent regressors equal to the comments of the sample.

Adjusted R-squared Statistica represents an alternative, with the advantage that “penalizes” the addition of regressors which do not contribute to the explanatory power of the model. Thus, this statistic may drop while the regressors are added, and for the models where the “match” of data is not very good, it can even be negative.

The results of the proposed model (McFadden R-squared 33,755 %) shows us that there is a strong enough connection between the students’ rate of success and the degree to which the students have attended the lectures, acquired the materials needed for the preparation of exams, but also on the time allocated to study each subject. Therefore, in order to increase students’ rate of success, it is necessary to meet the exogenous variables (R_2, R_3, R_4, R_5).

References

1. Biji, E. M.,, Lilea, E., Vatui, M., Gogu, E., (2012) *Compendiu de statistica aplicata în economie*, Oscar Print Publishing House, Bucharest, 2012

2. Brooks, C., (2008), *Introductory Econometrics for Finance*, 2nd Edition, Cambridge University Press

3. Canova, F., (2007), *Methods for Applied Macroeconomic Research*, Princeton University Press Enders, W., (2004), Applied Econometric Time Series, 2nd Edition, Wiley

4. Greene, W.H., (2008), *Econometric Analysis*, 6th Edition, Prentice Hall Hamilton, J., (1994), Time Series Analysis, Princeton University Press

5. Stancu, S., (2011), *Econometrie. Teorie şi aplicaţii utilizând Eviews,* ASE, Publishing House Bucharest, 2011

6. Turdean, M.S, (2013), S*tatistică,* Pro Universitaria Publishing House, Bucharest.