Vidyasagar University Journal of Commerce
Vol. 18, 2013/ISSN 0973-5917

EXPLAINING THE DEMAND FOR DOMESTIC ALTERNATIVE POWER BACK-UP SYSTEM – A LOGISTIC REGRESSION APPROACH

Arijit Roy*

Abstract

Power supply is a basic input required for progress of a civil society. Growing imbalance between demand and supply of power has resulted in a very steep increase in the demand for alternative domestic power backup system like battery and inverter, mini generator set etc. Specially in the urban areas. In statistical case studies where categorical results such as "successful-unsuccessful", "present-not present", "ill-not ill" are obtained as a result of data evaluation, the logistic regression is a rather suitable statistical method. In this paper, we have tried to estimate the probability of having an alternative domestic power generation system in an urban set up in West Bengal using logistic regression approach. Primary data were collected in Nadia district of West Bengal during 2009-2010. It has been observed that the level of per capita income, the duration of average daily power cut and presence of children in a family pursuing education do have a positive and significant influence on the demand for domestic alternative power back-up system.

Key words: Alternative power back-up system, Binary logistic regression.

1. Introduction

The researchers and model designers always try to convert the data they obtain from real events or experiments to functional structures by means of various models. Although to establish a mathematical model is very difficult, doing so ensures beneficial information.

Classification of the data used in models constitutes the very important part of the statistical analysis. The following are examples of such logistic regression studies. A pioneering study was made to econometrically examine the impact of group ownership on conduct of a bank in an emerging economy like India (Deb et al. 2006). The paper substantiates the findings from case studies through estimating a logit model. The results clearly show that group banks differ in their conduct from non-group banks. A hospital based case study was carried out to investigate the role of sexual risk factors in cervical cancer among rural Indian women (Biswas

* Asstt. Professor in Economics, Dept. Of Commerce, V.M. Mahavidyalaya, Purba Medinipur

et al. 2008). A case control design was used in which a total of 268 subjects, comprising 134 women with invasive cervical cancer as cases and 134 control women were studied. A multiple logistic regression model was used to analyse the data. A qualitative model was built up consisting of a logistic function that uses economic variables to determine the probability that a country will face a financial crisis the following year (Sperber 2008). The logistic model developed in this paper is analysed to identify the economic variables that have the greatest impact on a country having a financial crisis. The paper presents a case study that applies the model to India. In a study, the data for the years 2004, 2005 and 2006 from 53 companies that are active in the insurance sector in Turkey were evaluated by using logistic regression method to classify them as "successful" and "unsuccessful" (Ruzgar et al. 2007). A comparison of the mobile nursing system that was established between the years of 1977-1985 in America to restrict the health expenses to the former system was examined with multi logistic regression analysis (Elston et al. 1990). The American data obtained from extraordinary events such as wars, elections, political crisis and epidemic diseases were used to determine differences between the periods when such events occurred and other periods by means of logistic regression (King et al. 2000). Binary logistic regression was used to calculate the retirement age of people depending on age, sex, economical and social statuses (McNamara et al. 2001). Between the years of 1980-1995, the data of bankrupted American companies were examined; 237 of the bankrupted companies were handled as samples for the year 1992, and the financial and non-financial values of their final bankruptcy resolutions were examined with logistic analysis and their classifications were estimated (Bran et al. 2002). The health insurance classification of insured and uninsured low-income children in America between the years of 1995-1999 and the classification of uninsured ones according to their sex, age and economical status were made by using the logistic regression model (Campbell et al. 2002). The national health researches of the Australian households were made by using 2001 data; the rate of switching to private health insurance and the reasons for switching including economical, social and health factors were examined by means of multi logistic regression analysis (Temple 2004). In classification of car accidents in America, the logistic regression analysis was performed using the variables such as literacy rate, economical status and sex (Lee et al. 2006). In Japan, 57 big parent companies that were very important for the Japanese economy between the years 1998- 2001 were classified as "financially under stress" and "peaceful" (Sullivan 2003). Logistic modelling was used to determine whether a Treatment Centre established for purpose of treating visually disabled or blind people to help them find new jobs was beneficial for such people (Capella 2005). Between the years 1980- 2004, the disability risk and disability risk insurances were examined in America and a classification was made by using the logistic regression according to workability limits, non workability situations and the need to get health care for people who retired due to a physical disability (Chandra et al. 2005). In another study, the logistic regression was used in determination and classification of car insurance tariffs of insurance companies (Christmann et al. 2004). Traditionally, a researcher desires to find whether there is a relation between two or more variables and to express such relation with an equation (Agresti 2002). For instance, an engineer may want to know the relation between the pressure and temperature, an economist between the income level and consumption expenses, an insurer between the number of policies sold and profitability, and an educator between the absent days of students and their success ranks. An equation showing the relation between two (or more) variables not only demonstrates the functional form of the relation between variables but it also estimates any variable if the value of another is known (Menard 1995). When a dependent variable is a classified variable which depends on two situations while independent variables can be continuous, discrete or classified, binary logistic regression has a quite functional relation and it is suitable for category classification by using the structure of regression analysis.

2. Methodology & Model

With similar geographical and socio – economic features, two cities, namely, Kalyani and Krishnagar of Nadia district of West Bengal were selected. The total sample size was determined by available budget. At the planning stage this paper targeted a total of 268 samples for urban households. The questionnaires were randomly assigned for households to find required information. The data were collected in the year 2009-2010. For power generation and distribution, both Kalyani and Krishnagar are under the jurisdiction of West Bengal State Electricity Distribution Company Limited (W.B.S.E.D.C.L). The data on arrangement of alternative power system, average duration of power cut per day, the per capita income of the households, size of households, the number of old people (age>= 60 years) and the number of children pursuing education in each of the 268 families were collected. We formulate duration of power cut (DPC) as a categorical variable taking value of 0 if tolerable and 1 if not. We also formulate old people (OLD) as a categorical variable taking a value of 1 if there is any old person in a household and 0 if none. We formulate children pursuing education (CPE) as a categorical variable taking a value of 1 if there is any child pursuing education and 0 if not. The dependent variable is arrangement of alternative power system (APS) where APS is coded 1 if it exists and 0 if does not. Our aim is to identify the predictors that help to discriminate the households having arrangement for alternative power system (APS). Since APS can take only two values, 0 and 1, we use the binary logistic regression to investigate whether arrangement of alternative power system (APS) in an urban set-up can be explained on the basis of per capita household income (Y), average duration of power cut (DPC), children pursuing education (CPE) and presence of old person (OLD) in a household. In other words, the binary logistic model selects the statistically significant variables (predictors) that help to distinguish between existence and non-existence of APS in the urban set-up. Here, we are trying to estimate the probability that a given urban household will have an alternative power system (APS). At each observation, the logit probabilities are represented by:

$$\ln (p/1\text{-}p) = \text{logit}\, p = \beta_0 + \beta_1 Y + \beta_2 (DPC) + \beta_3 (CPE) + \beta_4 (OLD) + u$$

where p is the probability that a given household has arrangement for APS; (1-p) is the probability that a given household has no APS; (p/1-p) is the odds ratio; ln(p/1-p) is the logit transform of p and is written as logit p which varies from $-\infty$ to $+\infty$; β_0 is the constant; β_j is the coefficient of the j-th (j = 1,2,3,4) predictor variable.

3. Result & Discussion

In logistic regression the parameters are estimated such that the coefficients make our observed results most likely. We run the binary logistic regression in SPSS to get the following results:

Table - 1: Case Processing Summary

Unweighted Cases		*n*	Percent
Selected Cases	Included in Analysis	268	100.0
	Missing Cases	0	0.0
	Total	268	100.0
Unselected Cases		0	0.0
Total		268	100.0

Table - 2 : Dependent Variable Encoding

Original Value	Internal Value
0.00	0
1.00	1

n is the number of cases in each category e.g. included in the analysis, missing and total, while percent gives percent of cases in each category e.g. included in the analysis, missing and total. In this case study, 268 cases are included. Since we have no missing data, this also corresponds to the total number of cases.

SPSS logistic regression is run in 2 steps. The first step is called Step 0.This part of the output describes a "null model", which is a model with no predictors and just the intercept. Consequently, all of the variables that I put into the model appear in the table titled "Variables not in the equation".

Block 0: Beginning Block

Table - 3 : Classification Table [a,b]

Observed			Predicted		
			APS		Percentage Correct
			0.00	1.00	
Step 0	APS	0.00	0	64	0.0
		1.00	0	204	100.0
	Overall Percentage				76.1

a. Constant included in the model.

b. The cut value is .500.

Table - 4: Variables In The Equation

	B	S.E.	Wald	d.f.	Sig	Exp(β)
Step 0 Constant	1.159	0.287	16.367	1	0.000	3.187

In this null model, the number of observed 0's and 1's in the dependent variable is 64 and 204 respectively. In this null model, SPSS has predicted that all cases are 0 on the dependent variable APS. The percent of cases for which the dependent variable is correctly predicted is 204/268 =76.10. The coefficient for the intercept is 1.159. The standard error around the coefficient for the intercept is 0.287. The value of Wald chi-square is 16.367 which tests the null hypothesis that the constant equals 0. This hypothesis is rejected because the *p*-value is smaller than the critical *p*-value of 0.05. Hence, we conclude that the constant is not 0. The degree of freedom for the Wald chi-square test is 1 because there is only one predictor in the model, namely, the constant. The exponentiation of the β coefficient which is an odds ratio is 204/64 = 3.187.

Table - 5 : Variables Not In The Equation

			Score	d.f.	Sig
Step 0	Variables	Y	20.404	1	0.00
		DPC	26.361	1	0.00
		CPE	24.419	1	0.00
		OLD	16.944	1	0.00
	Overall Statistics		39.839	4	0.00

This is a Score test that is used to predict whether or not an independent variable would be significant in the model. Looking at the *p*-values, we can see that each of the predictors would be statistically significant. Each variable to be entered into the model, e.g., Y, DPC, CPE and

OLD , has 1 degree of freedom, which leads to the total of four shown at the bottom of the column. The overall statistics show the result of including all of the predictors into the model.

The next section contains the most interesting part of the output: the overall test of the model in the "Omnibus Tests of Model Coefficients" table and the coefficients and the odds ratios in the "Variables in the Equation" table.

Block 1: Method = Enter

Table - 6 : Omnibus Tests Of Model Coefficients

		Chi-square	d.f.	Sig
Step 1	Step	51.732	4	0.00
	Block	51.732	4	0.00
	Model	51.732	4	0.00

This is the first step with predictors in it. In this case, it is the full model that we specified in the logistic model. This shows the chi-square statistic and its significance level. In our case, the statistics for the Step, Block and Model are the same because we have not used stepwise logistic regression or blocking. The value given in the Sig. column is the probability of obtaining this chi-square statistic (51.732) if there is infact no effect of the independent variables, taken together, on the dependent variable. The *p*-value (0.00) is compared to the critical *p*-value (0.05) to determine if the overall model is statistically significant. In our case, the model is statistically significant because the *p*-value is 0.00. There is 1 d.f. for each predictor in the model. Here, we have 4 predictors: Income per head (Y) and 3 dummies for duration of power cut (DPC), children pursuing education (CPE) and presence of old people (OLD).

Table -7 : Model Summary

Step	-2Loglikelihood	Cox & Snell R Square	Nagelkerke R Square
1	21.928 [a]	0.538	0.807

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than 0.001.

-2 loglikelihood for the final model is 21.928 which can be used to compare reduced models. Cox & Snell R Square (= 0.538) and Nagelkerke R Square (= 0.807) are pseudo R-Squares.

Table - 8: Classification Table

Observed			Predicted		
			APS		Percentage Correct
			0.00	1.00	
Step 0	APS	0.00	56	8	87.5
		1.00	12	192	94.1
Overall Percentage					92.5

a The cut value is 0.500.

Table 8 shows the number of 0's and 1's that are observed in the dependent variable along with the predicted values based on the full logistic regression model. This table shows 192 cases are observed to be 1 and are correctly predicted to be 1; 56 cases are observed to be 0 and are correctly predicted to be 0; 12 cases are observed to be 1 but are predicted to be 0 while 8 cases are observed to be 0 but predicted to be 1. The overall percent of cases that are correctly predicted by this full model has increased to 92.50 from 76.10 in the null model.

Table - 9: Variables In The Equation

		β	S.E.	Wald	d.f.	Sig	Exp(β)
Step 1[a]	Y	1.021	0.506	4.073	1	0.044	2.776
	DPC	2.473	1.221	4.103	1	0.043	11.861
	CPE	2.970	1.338	4.928	1	0.026	19.488
	OLD	1.442	1.139	1.602	1	0.206	4.229
	Constant	-9.664	4.010	5.810	1	0.016	0.000

a. Variables entered on Step 1 : Y, DPC, CPE, OLD.

The β values show the values for the logistic regression equation for predicting the dependent variable from the independent variables. They are in log-odds units. The prediction equation is $\ln(p/1-p) = -9.664 + 1.021\ Y + 2.473\ DPC + 2.970\ CPE + 1.442\ OLD$

where p is the probability of having alternative power system (APS).

According to this regression, per capita income of household (Y), average daily duration of power cut (DPC) and presence of children pursuing education (CPE) are statistically significant and carry a positive sign. Although presence of old people (OLD) carries a positive sign, the predictor is not statistically significant.

These estimates tell about the relationship between the independent variables and the dependent variable, where the dependent variable is on the logit scale. These estimates tell the amount of increase or decrease in the predicted log-odds of Alternative Power System = 1 that would be predicted by a one unit increase or decrease in the predictor, holding all other predictors constant. For every one unit increase in per head income, we expect a 1.021 increase in the log-odds of Alternative Power System, holding all other independent variables constant. The coefficients for the dummies DPC and CPE are statistically significant while the coefficient for the dummy OLD is not.

4. Conclusion

In this study, the applicability of the binary logistic regression in determining the probability of finding an alternative power back-up system in an urban set-up in West Bengal is tried to be demonstrated. 268 urban households in West Bengal that have sufficient data were examined by means of the binary logistic regression to categorize whether they have alternative power back-up system or not. It is seen that the binary logistic regression equation explains the impacts of income and demographic factors on the demand for domestic alternative power back-up system. In conclusion, the logistic regression is deemed to be a good method for explaining the probability of finding an alternative power back-up system in an urban domestic set-up.

References

Deb, A.T. and K.V. B. Murthy (2006), 'Business Group Ownership of Banks: Issues and Implications', *Paper presented in the 8th Annual Conference on Money and Banking,* I.G.I.D.R, Mumbai.

Biswas, L.N., B. Manna, P.K. Maity and S. Sengupta (1997), 'Sexual Risk Factors for Cervical Cancer among Rural Indian Women: A Case-Control Study', *International Journal of Epidemiology*, 26(3) : pp 491-495.

Sperber, M. (2008), 'Predicting Financial Debt Crises : A Case Study of India', Unpublished Honours Thesis Submitted to Duke University, Durham.

Ruzgar, B. and N.S. Ruzgar (2007), 'Classification of the Insurance Sector with Logistic Regression', *International Journal of Mathematical Models and Methods in Applied Sciences,* 1(3) : pp168-174.

Elston, J.M., G.G. Koch and W.G. Weissert (1990), 'Risk of Institutionalization', 1977–1985 U.S. Department of Health and Human Services.

King, G. and L. Zeng (2000), 'Logistic regression in rare events data', *Political Analysis,* 9(2) : pp137–310.

McNamara, T.K. and J.B. Williamson (2001), 'Why Some Workers Remain in the Labor Force Beyond The Typical Retirement Age?' *Center for Retirement Research Working Papers*, no. 47, Boston College.

Bran, R.A. and A.R. Leach (2002), 'Predicting Bankruptcy Resolution', *Journal of Business Finance and Account,* 29(3) : pp497-507.

Campbell, J. and R. Fisher (2002), 'Health Insurance Estimates for States', *U.S. Census Bureau Govt. Statistics Proceedings of the American Statistical Association Annual Meeting,* New York.

Temple, J. (2004), 'Explaining the Private Health Insurance Coverage for Older Australians', *People and Place*, 12(2) : pp13–24.

Lee, T., R. Liu and Y. Yeh (2006), c*an corporate governance variables enhance the prediction power of accounting-based financial distress prediction models? Financial Distress and Bankruptcy Prediction- An introduction :*Magnus School of Business Publisher.

Sullivan, K.(2003), 'Transportation Work: Exploring Car Usage and Employment Outcome in the LSAL data', *NCSALL Occasional Papers*, Cambridge.

Capella, M.E. (2005), 'Predictors of Competitive Employment for Blind and Visually Impaired Consumers of Vocational Rehabilitation Services', *Journal of Visual Impairment and Blindness,* 99(5) : pp303-315.

Chandra, A. and A. Samwick (2005), 'Disability Risk and the value of Disability Insurance', In D. Culter and D. Wise (ed), *Health in older ages: The Causes and Consequences of Declining Disability Among the Elderly,* Univ. of Chicago Press, Chicago.

Christmann, A. and M.M. Galiano (2004), 'Insurance: An R-program to Model Insurance Data', *Technical Report :* pp5-8, Univ. of Dortmund.

Menard, S. (1995), *Applied logistic regression analysis,* pp. 37-42. USA : Sage Pub.

Agresti, A. (2002), *Categorical Data Analysis*, pp. 165-257, Florida : Wiley Pub.