International Journal of Academic Accounting, Finance & Management Research(IJAAFMR)
ISSN: 2643-976X
Vol. 7 Issue 5, May - 2023, Pages: 11-16

A Discussion of Limited Dependent Variable

Robert Odek[1] and Dr. Gordon Opuodho[2]

1 (PhD Accounting Student)
roba.odek@gmail.com
2 (Lecturer)
Department Of Economics, Accounting & Finance
Jomo Kenyatta University of Agriculture And Technology

***Abstract**: In every research work, there are variables. Variables aid in clarifying in detail the research objective and what the study intends to achieve. Limited dependent variable is referred to as being non-linear and usually qualitative in addition to being restrictive. Research shows that there is a challenge in understanding and drawing accurate inferences with the use of limited dependent variable models among researchers that opt to use non-linear models. Thus the Limited Dependent Variable models are discussed herein. Further, this article discusses the general interpretation of the modern sophisticated models utilized in the interpretation of discrete data analysis. The article is also presented in a manner that aids easy comprehension by business students who are taking econometrics as a unit but are otherwise not having very extensive knowledge of econometrics.*

Keywords: Limited dependent variable, logit regression, probit model

1.0 INTRODUCTION

It has been evidenced for a longer period of time that the most utilized model in data analysis is the ordinary least square. The use of continuous data has advised the wide use of ordinary least squares. A continuous type of data is that which can take any value and it only changes over time. So that it can easily develop an observable trend. For instance, the weight of a baby is a continuous type of data and can be presented in a graphical nature over a period of time. However, recent studies have evidence that there is a need to interpret discrete data (Wiersema & Bowen, 2009). That is, as a result of dynamism in various organizations, not only is continuous data interpreted, but also discrete data. Discrete data is that which only takes particular values as compared to continuous data. For instance, the bottom line of a corporation for the first quarter may be Ksh. 151,670.51.

Discrete data are not best interpreted with the aid of straight graphical representations. As a result, there has been the introduction of mitigatory models to help in the interpretation of discrete data. Thus the Limited Dependent Variable models are discussed herein. Further, there is a need to briefly discuss the general interpretation of the modern sophisticated models utilized in the interpretation of discrete data analysis. This article is also presented in a manner to aid easy comprehension by business students who are taking econometrics as a unit but are otherwise not having very extensive knowledge of econometrics. Recent studies have shown that one of the factors undermining adequate conclusions among researchers applying non-linear models is the inability to properly interpret results from the non-linear models utilized.

1.1 Research Variables

In every research work, there are variables. Variables aid in clarifying in detail the research objective and what the study intends to achieve. A variable is any circumstance that can fluctuate in magnitude or value. In most cases, research works consist of two variables; independent and dependent variables. An Independent variable is a variable that can be manipulated so that any transformation on it leads to a direct effect on the dependent variable. On the other hand, the dependent variable is that which remains controlled and is only affected by changes in the independent variable. Therefore if the changes in the independent variable directly influence the dependent variable, then a causal relationship or rather a functional relationship is established. It is centered on the causal or functional relationship that researchers can develop theories and deductive inferences contemplating future behaviours on the same subject under investigation. It is also important to note that variables can be both quantitative (containing values) and qualitative (containing descriptive statements). Example of independent and dependent variable is illustrated in Figure 1.

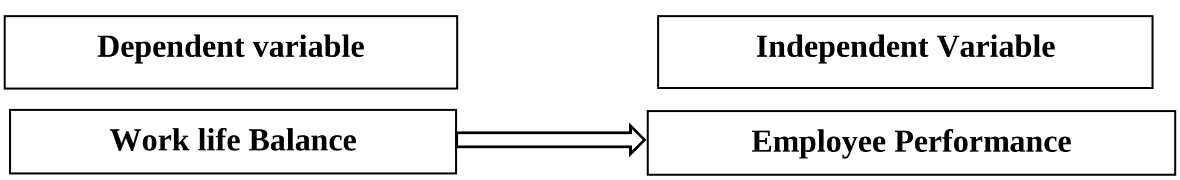

Figure 1: Independent and dependent variables

1.2 Limited Dependent Variable (LDV)

Due to the increasing change in the organizational environment, researchers have resorted to analyzing what really causes the changes and not just what the change is. As a result various variable particularly the dependent variable has been frequently modeled into discrete values which requires more advanced analysis thus the limited dependent variable. Limited dependent variable is referred to as being non-linear and usually qualitative. That is, the equation cannot be written as an addition of various sub-variables (Wiersema & Bowen, 2009). A limited dependent variable is therefore referred to as a rejoinder variable whose value is circumscribed to a small number so that the restraint requires analysis in a precise tactic (Bowen, 2020). An example of a dependent variable is being employed and not being employed.

(Amemiya, 1985) who has widely contributed to the field of econometrics acknowledges that the first person to come up with the rule of generalization of the non-linear models otherwise the limited dependent variable was (Tobin, 1958)in his study on measuring the relationship of dependent variables.

1.2.1 Limited Dependent Variable Verses ordinary least square regression

(Wiersema & Bowen, 2009) notes that there is a difference between the ordinary least square method and the limited dependent variable. Besides, to adequately interpret results using the limited dependent variable, there is a need to comprehend the difference between the two models. (Phoebus, 1986) suggests the following main differences;

i. Linear regression forecasts actual choice vis a vis Limited Dependent Variable which estimates the probability of choices.
ii. The stochastic error term conditions applicable in ordinary least square are not congruent with the conditions under limited independent variable. In other words, Limited dependent variable are estimated on the method of maximum likelihood compared to OLS which lacks minimizing error.

2.0 Literature Review

In a study conducted by (Hoetker, 2007) , between 2000-2005, he presents that the use of probit and logit models is currently being embraced as compared to the use in the 1980s. (Hoetker, 2007) , reviewed 157 articles that applied logit or probit models. He concluded that this signified an increase in the use of non-linear regression methods in data analysis.

(Wiersema & Bowen, 2004) noted that the use of limited dependent variables have been on the increase. However, they distinguished that to a greater extent, just minimal attention has been placed in comprehension to the adequate application of Limited dependent variables in data interpretation, particularly strategy research. This, therefore, signifies that there is a challenge in understanding and drawing accurate inferences with the use of limited dependent variable models among researchers that opt to use the same technique.

(Taplin, 2016) in a study on limited dependent variables penned that precision is the key aspect for the application of non linear models. Thus, the author comments that most of the researchers that have utilized limited dependent variables have faced problems in result interpretation while applying Limited Dependent Variable models. This concern is equally affirmed by (Wiersema & Bowen, 2009).

(Taplin, 2016) analyzed research articles that utilized non-linear models between the year 2003- 2007 and 2009-2011. The author points out that between the years 2003-2007, there was an increase in those studies which had utilized the interpretation of regression co-efficient which transformed from 29% to 41 %. That is between the years 2003-2007, a total of 66 papers discussed non-linear models and out of this, 19 papers interpreted regression coefficients. On the other hand, between 2009-2011, a total of 103 articles had been published and out of which 42 articles interpreted regression coefficients. However, (Taplin, 2016) presents a concern that as much as there is an increase in papers that have utilized non-ordinary linear models, the majority of these articles are yet to present any interpretation of regression co-efficient other than noting the sign of co-efficient and the statistical significance. Based on the foregoing, it can then be concluded that the interpretation of study results applying non-linear models is still a challenge. The majority of the researchers are still incapable of utilizing these models to the conclusion, therefore, undermining their research capability.

Other suggestions by (Taplin, 2016) are that some other factors hindering proper conclusions and interpretation of results by the researchers using non-linear models have been a poor statement of the hypotheses to be tested coupled with a lack of dependent variables with more than two categories.

3.0 Limited Dependent Variable Interpretation Models

Other researchers (Michael, 1997) support the idea that non-linear models are useful in the analysis of discrete data which only utilizes particular figures. Hence, there are various kinds of discrete consequences;

a) Binary outcomes which only take two value-1 or 0, YƐ{0,1}. For instance, 0=is a drug addict, 1= not a drug addict. Binary kind of discrete data is analyzed by the use of probit LDV.

b) Ordinal data which utilizes three or even more variables in its rankings. For instance, the impact of leave entitlement on employee performance so that the response is as indicated in parenthesis (Are you entitled to leave days as a hotel employee; 1=strongly agree, 2=Agree, 3=disagree, 4=not sure). **This is analyzed using ordered probit regression.**

c). Nominal results utilize three or more values, though lack intrinsic order. For instance, various estates where people leave in Kisumu Kenya (Nyawita=1, Polyview=2, Milimani=3) YƐ{0,1,2,3,4}. This is analyzed using **multi-nomial logistic regression.**

d). Count which considers how many times a particular event occurred. For instance, the number of days a customer visited the supermarket in a month. This is analyzed using **Poisson regression.**

e). Censoring; this is where the value of the dependent variable is zero and is known to the researcher. On the other hand, the value for the independent variable is known as congruent to the dependent variable value YƐ{Y*:Y*≥0}.

f). Truncation; here there are particular observations for the values for the dependent variable that cannot be completely made so that the dependent variables are totally not identifiable by the researcher.

3.1 Logistic Model

(Nelder & Wedderburn, 1972) posits that the logistic model and linear regression are all generalized models. For instance, in the contingency table, with a multiplicative model, the logarithm of the expected probability is assumed linear on classifying factors defining the table. Linear model is given as below;

$$Y=\beta 0+\beta 1X1+\beta 2X2+\beta 3X3+\mu \text{----------------} (1)$$

Where; Y=Dependent variable

$\beta 0$=Constant of the model

$\beta 1$=Coefficients of the regression equation

$X1$= independent variable 1

$X2$= independent variable2

$X3$ = independent variable2

ɥ=Tolerable error

However, (Nelder & Wedderburn, 1972) noted that linear regression assumes only one error component. For analysis of designed surveys and experiments, extensions of multiple errors have been developed. This then leads us to the logistic regression. So that using the same linear model, we transform it to the logistic regression model by introducing another error term. This will appear as below;

$$l=\beta 0+\beta 1X1+\beta 2X2+\beta 3X3+\mu \text{-----} (2)$$

Here, the equation has taken the form of a linear regression equation with the dependent variable *l* measured on a different scale to the dependent variable.

Logit model creates a link between the linear predictor *l* to the probability of outcome p with the logit function being;

$$l=\ln\left(\frac{\quad}{1-p}\right) \text{-------------} (3)$$

Putt ıer equation 2 and 3 will give what is referred to as logit model or logistic regression with binary independent variables as depicted below;

$$l=\ln\left(\frac{p}{1-p}\right) = \beta 0+\beta 1X1+\beta 2X2 \text{-------------} (4)$$

Where;

$\ln\left(\frac{p}{1-p}\right)$ *is the logodds*

$\frac{}{1-p}$ *is the odds*

P $1-p$ *e probability*

Interpreting the results in terms of log odds is similar to ordinary regression. In other words, the logit function heads to infinity as p approaches 1 and heads to negative infinity as p approaches 0.

3.1.2 Concept of odds ratio and log odds

The probability of an event occurring is **(p)** and the probability of the event not occurring is **(1-p). Odds ratio** is therefore probability of p not occurring denoted as **p/(1-p). Log of odds** is thus denoted as **log p/(1-p).**

Probability ranges from 0 to 1. Odd ranges from - ꝏ to ꝏ. Using odds and log odds is a simpler and accurate interpretation of logit models compared to modeling a variable with a restricted range.

3.1.3 Practical interpretation using logit model

Table 1: Probabilities, odds and log odds

Probability	**Odds**	**Logodds**
0.01	0.01	-4.60
00.3	0.33	-0.85
0.7	2.22	0.85

The result in table 1 shows that probability of 0.7 corresponds to the odds of 2.22 and log odds of 0.85. What follows is an example from the literature which portrays how results can be interpreted using the log odds.

Table 2: Extract of logit outcomes computed from those in (Wiersema & Bowen, 2009)

Dismissal	**P>│Z│**	**95% confidence level**
X	0.000	-2.673053
Z	0.004	1.854669
Constant	0.000	-1.450213

The data in table 2 were generated from stata software and it sought to find out the chances of dismissing the chief executive officer in a survey of 199 firms. The variable was modeled such that 0 implied CEO succession and 1 implied CEO dismissal. The results showed among 76 firms that the CEO was dismissed. Applying the logit model in this binary circumstance will be depicted as follows;

$$l=\ln\left(\frac{p}{1-p}\right) = \beta 0+\beta 1X1+\beta 2X2 \text{ ---logit function/model}$$

p/(1-p) is the probability of CEO not being dismissed. On the hand p/(1+p) is the probability of CEO being dismissed. Hence;

P(dismissal=1│X,Z; β0, βXβZ)

$$\frac{p/\beta 0+\beta X^{X}+\beta Z^{Z}}{1+p/\beta 0+\beta X^{X}+\beta Z^{Z}}$$

Variable X and Z are both significant with p<0.004 as per table 2.

3.2 Probit model

The probit model also referred to as probit regression is utilized majorly in econometrics where there are binary choices. The word probit originates from two terms, probability, and unit so this model estimates the probability a value will fall into one of the two possible binary (i.e. unit) outcomes. Since binary choices imply choices that are two, it means the model measures the chances of falling into either choice, for instance, successful or not successful.

3.2.1 When to use probit models

i. In determining an election winner (1=win, 2=Lose)
ii. In determining acceptance/admission in a graduate school (1=Admitted, 2=Denied)
iii. In determining whether an article would be published in a particular journal (1=Accepted, 2=Rejected)

3.1.2 Dummy interpretation of probit regression

Consider a circumstance where the outcome Y is binary. This implies it can only possess two results, 1, or 0. This will come with X which is the regressor's vectors influencing the Y variable. That is the number of times Y will occur will be influenced by Y. The model will then take the function;

$\mathbf{P(Y=1 \mid X)=\Phi(X^T \beta)}$

Where P; is the probability

Φ is the cumulative distribution of the standard normal distribution

β is estimated by maximum likelihood

This can still be simply put just as was the case of logistic regression as;

$\mathbf{\mathit{l}= \Phi^{-1}(p)}$

Here the cumulative distribution can be used to create the link in transforming the linear predictor *l* to probability between 0 and 1

Maximum likelihood estimation

Consider data set $\{y_i, x_i\}^n_{i=1}$ containing independent variables.

For single observation;

$P(y_i=1 \mid x_i)= \Phi(x_i' \beta)$

$P(y_i=0 \mid x_i)= \mathbf{1}\text{-}\Phi(x_i' \beta)$

Likelihood of single observation is thus;

$L(\beta_i, y_i\ x_i,)= \Phi(x_i' \beta)^{\ y_i} [1\text{-}\Phi(x_i' \beta)]^{(1- y_i)}$

So that if $y_i=1$, then $L(\beta_i, y_i\ x_i,)= = \Phi(x_i' \beta)$

If $y_i=0$, then $L(\beta_i, y_i\ x_i,)= 1\text{-}\Phi(x_i' \beta)$

If the observation is identical, then we can have joint likelihood given as follows;

$L(\beta, y, x)=\Sigma[(y_i \ln \Phi(x_i' \beta) + (1- y_i) \ln(1- \Phi(x_i' \beta)]$

The estimator β which maximizes the function is consistent

4.0 Conclusions

It has been evidenced from this article that Limited dependent variables are constrained variables that only take particular values. Besides, Limited dependent variables are utilized in designed particular studies or surveys. Since the variables are limited as the name suggests, it has also been noticed that they may not be best interpreted using linear regressions since linear regression only have one possibility of error. Instead, more sophisticated methods of interpretations are utilized thus; logistic regression and probit model. Logistic regression is simple to use. It is actually a generalized linear regression transformed by the introduction of multiple errors so that there is no one error component like that in the linear regression model. Using modeled data in statistical software like Stata, it is easy to therefore draw inferences utilizing logit and probit models for the limited dependent variable under study. It has also been evidenced that limited dependent variables are more prone in strategic research compared to other types of studies which in most cases utilizes continuous data which may be modeled linearly. Literature review evidenced that the majority of the empirical studies on limited dependent variables have always not interpreted the results correctly, therefore, undermining these empirical studies' capability to advance knowledge. This article has overcome this by being presented in a more precise and easy to comprehend form. I recommend that more practical integration of results from the field using limited dependent variable models be embraced by lecturers, especially among business accounting and finance students who are not necessarily having the origin of in-depth statistics and mathematics.

References

Amemiya, T. (1985). Limited Dependent Variable. *Econometric Workshop.* California: University of Stanford. Retrieved March 12, 2023

Hoetker, G. (2007, February 5). The Use Of Logit And Probit Models In Strategic Management Research: Critical Issues. *Strategic Management Journal, 28*, 331-343. doi:10.1002/smj.582

Michael, R. (1997). Regression Models for Discrete and Limited Dependent Variables. Buffalo, New York: Research Institute on Addictions. Retrieved from http://division.aomonline.org/rm/1997_forum_regression_models.html

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society., 135*(3), 370-384. Retrieved from http://www.jstor.org/stable/2344614

Phoebus, J. (1986). Limited Dependent Variables. In C. University, *Hand Book of Econometrics.* Elsevier Science publishers BV.

Taplin, H. R. (2016). Research Involving Limited dependent variables: issues in the literature and recommendations for improvement. *Qual Quant, 50*, 2121-2140. doi:10.1007/s11135-015-0254-5

Tobin, J. (1958). Estimation of relationships for Limited Depndent Variables. *Econometrica, 26*, 24-36.

Wiersema, F., & Bowen, H. (2009). The Use Of Limited Dependent Variable Techniques In Strategy Research: Issues And Methods. *Strategic Management Journal, 30*, 679-692. doi:10.1002/smj.758

Wiersema, M., & Bowen, H. (2004). The Use Of Limited Dependent Variable Techniques In Strategy Research: Assessment And Critique.