YANGON INSTITUTE OF ECONOMICS

DEPARTMENT OF STATISTICS

LINEAR PROBABILITY, LOGIT AND PROBIT MODELS IN QUALITATIVE DATA ANALYSIS

THIDA THAN

M. Econ (Statistics)

(Roll No. 1)

MARCH 2010

CONTENTS

CHAPTER 1

INTRODUCTION

There are several methods for measuring the relationship among economic variables. The simplest methods are correlation analysis and regression analysis. Regression analysis was first developed by Sir Francis Galton who was a well known British anthropologist and meteorologist in the latter part of the 19^{th} century. It is a statistics methodology that utilizes the relation between two or more variables so that one variable can be predicted from the other, or others. This methodology is widely used in businesses, social and behavioral sciences, biological sciences, and many other disciplines.

Many regression models in which the regressand, the dependent variable, or the response variable, say Y, is quantitative, whereas the explanatory variables are either quantitative (or dummy), or a mixture thereof. In much research work, the researchers often face situations where the dependent variable of interest is a qualitative in nature. The dependent variable of interest or regressand,Y, may be two or three or multiple possible qualitative outcomes. The models in which the dependent variable or regressand, Y, is qualitative variable are called qualitative response models. These models are valuable in the analysis of survey data. The simplest possible qualitative response regression model is the binary model in which the regressand, has only two possible qualitative outcomes, and therefore can be represented by a binary indicator variable taking on values 0 and 1. So the regressand can be said that a binary or dichotomous variable and the models developed for such situations are called binary response models.

Both theoretical and empirical considerations suggest that when the response variable is binary, the shape of the response function will frequently be curvilinear. The shape of this response function is a titled S or as a reverse titled S, and they are approximately linear except at the ends. These response functions are often referred to as sigmoidal.

In a model where Y is quantitative, the objective is to estimate its predicted, or mean value given the values of the regressors, that is, $E(Y_i \mid X_{1i}, X_{2i}, X_{3i},\ldots\ldots,X_{ki})$, where the X's are regressors, may be quantitative or qualitative or both. In models where Y is qualitative, the objective is to find the probability of something happening.

Hence, qualitative response regression models are often known as a type of probability models. Qualitative response models have been extensively used in biometric applications for a much longer time than they have used in economic applications.

Among the qualitative response models, linear probability, logit and probit (also known as normit) models are studied in this paper. The objectives of this paper are to study;

(1) how to develop the qualitative response models;
(2) how to estimate the qualitative response models;
(3) how to evaluate the qualitative response models;

Firstly, the natures of qualitative response models are introduced in Chapter I. The specification and estimation procedure of the qualitative response models are discussed in Chapter II. Then, in Chapter III, diagnostic statistics for qualitative response models are discussed and, the applications of the models are studied in Chapter IV. Finally, the important characteristics of the models and findings are summarized in Chapter V.

CHAPTER II

MODEL SPECIFICATION AND ESTIMATION

In this Chapter some of the qualitative response models are considered for a binary response variable. Among the binary response models, linear probability, logit, and probit (normit) models are discussed in the following sub-sections.

2.1 Linear Probability Model (LPM)

2.1.1 Functional Form

The functional form of a linear probability model can be expressed as

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (2.1.1)$$

where $Y_i = 1$ if the event occurs and

$= 0$ if the event does not occur

β_1 and β_1 are regression coefficients. u_i is a random error term. X_i is the predictor variable.

It can be extended to more than one predictor variable.

That is,

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots\ldots\ldots\ldots + \beta_k X_{ik} + u_i \qquad (2.1.2)$$

$$Y = X\beta + u \qquad (2.1.3)$$

Assume that the model contains a constant term, that is, $X_{i1} = 1$ for all individuals. The regression coefficient is interpreted in terms of the probability of being in the interest category on Y. Hence, β_2 represents the change in he probability for each unit increase in X_i, net of the other covariates, and so on.

2.1.2 Examination of the Assumption of u_i

Assuming $E(u_i) = 0$, the conditional expectation of Y_i given X_i is obtained as:

$$E(Y_i|X_i) = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots\ldots\ldots\ldots + \beta_k X_{ik} = \sum \beta_k X_i = x_i'\beta \qquad (2.1.4)$$

If π_i is the probability that $Y_i = 1$ (that is, the event occurs), and $(1 - \pi_i)$ is the probability that $Y_i = 0$ (that is, the event does not occur), then the variable Y follows Bernoulli probability distribution. The expectation of Y is obtained as

$$E(Y_i) = 1.\ \pi_i + 0\ (1-\pi_i) \quad = \pi_i \quad = Pr\ (Y_{i=1}) \qquad (2.1.5)$$

Comparing Equation (2.1.4) with Equation (2.1.5), the conditional expectation of the model (2.1.2) can be interpreted as the conditional probability of Y. That is,

$$E\ (Y_i|X_i) = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots\ldots\ldots\ldots + \beta_k X_{ik}$$
$$= \pi_i$$
$$= Pr\ (Y_i=1)$$

Since the probability π_I must lie between 0 and 1 , this is a restriction.

That is, $0 \leq E\ (Y_i|X_i) \leq 1$.

Then the disturbances (u_i) also take only two values; that is , they follow the Bernoulli Distribution.

Y_i	u_i	Pr (Y_i)
1	1- $x_i'\ \beta$	π_i
2	- $x_i'\ \beta$	1- π_i
		1

Obviously,u_i cannot be assumed to the normally distributed; the follow the Bernoulli distribution. The OLS point estimators still remain unbiased. Besides, as the sample size increases indefinitely, statistical theory shows that the OLS estimators tend to be normally distribute generally .As a result, in large samples the statistical inference of the LPM will follow the usual OLS procedure under the normality assumption.

Even if $E(u_i) = 0$ and Cov (u_i, u_j) = 0 for $i \neq j$ (i.e., no serial correlation), it can no longer be maintained that in the LPM the disturbances are homoscedastic.

As statistical theory shows that for a Bernoulli distribution the theoretical mean and variance are, respectively, π_I and (1- π_i), where π_I is the probability of success (i.e., something happening) showing that the variance is a function of the mean. Hence the error variance is heteroscedastic. The variance of the error term is

$$Var\ (u_i) = \pi_I\ (1-\pi_i).$$

That is, the variance of the error term in the LPM is heteroscedastic. Since $\pi_i = E(Y_i \mid X_i) = \sum \beta_k X_{ik}$ the variance of u_i ultimately depends on the values of X and hence is not homoscedastic.

2.1.3 Estimation

For a model with heteroscedastic error disturbances it can be assumed that each error term u_i is normally distributed with variance σ_i^2, where the variance $Var(u_i) = E(u_i^2) = \sigma_i^2$ is not constant over observations. When heteroscedasticity is present, ordinary least squares estimation places more weight on the observations with large error variances than on those with small error variances. In the presence of heteroscedasticity, the OLS estimators, although unbiased, are not efficient; that is, they do not have minimum variance. If the heteroscedasticity is present, the appropriate estimation technique is the weighted least-squares estimation procedure, which can be derived from the maximum likelihood function.

Consider the simple linear probability model

$$Y_i = \beta_1 + \beta_2 X_i + u_i \text{ ; where } V(u_i) = \sigma_i^2. \qquad (2.1.1)$$

By minimizing the expression where the original variables are written in deviation form, the appropriate estimation can be obtained as

$$\hat{\beta} = \frac{\sum x_i y_i/\sigma_i^2}{\sum x_i^2/\sigma_i^2}$$

$$= \frac{\sum (x_i/\sigma_i)(y_i/\sigma_i)}{\sum (x_i/\sigma_i)^2}$$

$$= \frac{\sum x_i^* y_i^*}{\sum (x_i^*)^2} \qquad \text{where } x_i^* = \frac{x_i}{\sigma_i},\ y_i^* = \frac{y_i}{\sigma_i}$$

To use weighted least-squares, the variables in the original regression model of Equation (2.1.1) are redefined as;

$$Y_i^* = \frac{y_i}{\sigma_i},\ X_i^* = \frac{x_i}{\sigma_i},\ u_i^* = \frac{u_i}{\sigma_i}$$

where Var (u_i^*) = Var $(\frac{u_i}{\sigma_i})$ $= \frac{1}{\sigma_i^2}$ Var (u_i)

$= \frac{\sigma_i^2}{\sigma_i^2}$

$=1$

Now, the new error term is homoscedastic.

Since there are many situations in which the relative magnitude of the error variances is not known, it is important to consider special cases in which sufficient sample information is available to make reasonable guesses of the true error variances.

One possibility is the existence of existence of a relationship between the error variances and the values of explanatory variable in the regression model. Specifically, assume that

Var $(u_i) = CX_i^2$

where C is a nonzero constant and X_i is an observation of the independent variable in the linear probability model.

If the variances are unknown, the variables in the above equation can be transformed as;

$Y_i^* = \frac{y_i}{x_i}$, $X_i^* = \frac{1}{x_i}$, $u_i^* = \frac{u_i}{x_i}$

Where Var (u_i^*) = Var $(\frac{u_i}{x_i})$

$= \frac{1}{\sigma_i^2}$ Var (u_i)

$= \frac{1}{x_i^2}$ Var (u_i)

$= \frac{1}{x_i^2} CX_i^2$

$= C$

Now, error term u_i^* is homoscedastic.

The LPM is plagued by problems, such as

(1) non – normality of u_i

(2) heteroscedasticity of u_i

(3) possibility of $\widehat{Y_i}$ lying outside the 0-1 range, and

(4) the generally lower R^2 values.

But these problems are surmountable.

As mentioned above, WLS can be used to resolve the heteroscedasticity problem or increase the sample size to minimize the non-normality problem. By resorting to restricted least-squares or mathematical programming techniques the estimated probabilities can be made to lie in the 0-1 interval.

But even then the fundamental problem with the LPM is that it is not logically a very attractive model because it assumes that $\pi_i = E(Y = 1 | X)$ increases linearly with X, that is the marginal or increment effect of X remains constant throughout.

Therefore, what we need is a (probability) model that has these two features;

(1) as X_i increases , $\pi_i = E(Y = 1 | X_i)$ increases but never steps outside the 0-1 interval , and

(2) the relationship between π_i and X_i is nonlinear, that is "one which approaches zero at slower rates as X_i gets small and approaches one at slower and slower rates as X_i gets very large.

2.2 Logit Model

Both theoretical and empirical considerations suggest that when the response variable is binary, the shape of the response function will frequently be curvilinear. The response functions are shaped either as a title S or a reverse titled S and that they are approximately linear except at the ends. These response functions are often referred to as sigmoid. They have asymptotes at 0 and 1 and thus automatically meet the constraints on E (Y).

The commonly used non-linear probability models are logit and probit models. The two distributions most often employed are the standard normal distribution and the standard logistic distribution. The standard normal distribution employed can be called as probit and the standard logistic distribution, as logit.

2.2.1 Functional Form

The simple logit model is expressed as

$$\pi_i = \frac{\exp(\sum \beta_k X_{ik})}{1+\exp(\sum \beta_k X_{ik})}$$

$$\pi_i = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} \quad (2.2.1)$$

Letting $Z_i = \sum \beta_k X_{ik}$

$$\pi_i = \frac{e^{Z_i}}{1+e^{Z_i}}$$

$$= \frac{1}{1+e^{-Z}} \quad (2.2.2)$$

2.2.2 Features

The features of the logit model are as follows;

(1) Logistic regression effects can be expressed in terms of percent changes in the odds. Odds ratios are useful in estimating changes in the probability of event occurrence with changes in predictors once a baseline probability has been calculated.

$$\pi_i = \frac{e^{Z_i}}{1+e^{Z_i}}$$

$$1 - \pi_i = 1 - \frac{e^{Z_i}}{1+e^{Z_i}}$$

$$= \frac{1+e^{Z_i} - e^{Z_i}}{1+e^{Z_i}}$$

$$= \frac{1}{1+e^{Z_i}} \quad (2.2.3)$$

The ratio of Equation (2.2.2) to (2.2.3)

$$\frac{\pi_i}{1+\pi_i} = \left(\frac{e^{Z_i}}{1+e^{Z_i}}\right) / \left(\frac{1}{1+e^{Z_i}}\right) \quad (2.2.4)$$

$$= e^{Z_i}$$

$\frac{\pi_i}{1+\pi_i}$ can be called the odds ratio.

Take the natural log of Equation (2.2.4)

$$L_i = \ln\left(\frac{\pi_i}{1+\pi_i}\right)$$
$$= Z_i$$
$$= \sum \beta_k X_{ik} \qquad (2.2.5)$$

The logit L goes from $-\alpha$ to $+\alpha$ as π goes from 0to1. That is, although the probabilities (of necessity) lie between 0 and 1, the logits are not so bounded.

(2) Although L is linear in X, the probabilities themselves are not. This property is in contrast with the LPM model where the probabilities increase linearly with X.

(3) If L, the logit, is positive, it means that when the value of the regressor (s) increases, the odds that the regressand equals 1 (meaning some event of interest happens) increases . It L is negative,the odds that the regressand equals 1 decreases as the value of X increases. To put it differently, the logit becomes negative and increasingly large in magnitude as the odds ratio decreases from 1 to 0 and becomes increasingly large and positive as the odds ratio increases from 1 to infinity.

(4) More formally, the interpretation of the logit model given in Equation (2.2.4) is as follows; β_2, the slope, measures the change in L for a unit change in X. The intercept β_1 is the value of the log odds in favor of occurring an event if the other event does not occur (or) is zero.

(5) If we actually want to estimate not the odds in favor of event but the probability of event itself, this can be done directly from Equation (2.2.2) once the estimates of β_1 and β_2 are available.

(6) Whereas the LPM assumes that π_i is linearly related to X_i, the logit model assumes that the log of the odds ratio is linearly related to X_i.

2.2.3 Estimation

A logistic response function is either monotonic increasing or monotonic decreasing, depending on the sign of the slope coefficients. It can be linearized easily. Logistic response functions, like the other response functions which have been considered are used for describing the nature of the relationship between the mean response and one (or more) predictor variable (s). They are also used for

making predictions. The weighted least squares and maximum likelihood estimation procedures can be used to estimate the parameters of the logistic response function.

For estimation purposes, consider Equation (2.2.5), that is

$$L_i = \ln\left(\frac{\pi_i}{1+\pi_i}\right)$$
$$= \sum \beta_k X_{ik} \qquad (2.2.6)$$

In estimating the above equation, Logit , L_i depends on the two types of data which are categorized by

(1) data at the individual, or micro level, and

(2) grouped or replicated data

Individual data

Let π_i = 1 if the event occurs

π_i = 0 if the event does not occur.

If these values put directly into the logit L_i, it is obtained as

L_i = $\ln\left(\frac{1}{0}\right)$ if an event occurs

L_i = $\ln\left(\frac{0}{1}\right)$ if an event does not occur.

Obviously, these expressions are meaningless. Therefore, if the data are situated at the micro, or individual level, the model cannot be estimated by the standard OLS routine. In this situation, maximum likelihood method can be used to estimate the parameters. This method is well suited to deal with the problems associated with the responses Y_i being binary. Instead of using the normal distribution for the binary random variable Y, Bernoulli distribution will be used to develop the joint probability function of the sample observations.

Since each Y_i observation is an ordinary Bernoulli random variable, where;

$P(Y_i = 1) = \pi_i$

$P(Y_i = 0) = 1 - \pi_i$

It's probability distribution is represented as follows;

$$f_i(Y_i) = \pi_i^{y_i}(1-\pi_i)^{1-Y_i} \; ; \quad Y_i = 0, 1, ; i = 1\ldots\ldots,n \qquad (2.2.7)$$

Here, $f_i(1) = \pi_i$ and

$f_i(0) = (1- \pi_i)$

Hence, $f_i(Y_i)$ simply represents the probability that $Y_i = 1$ or 0

Since the Y_i observations are independent, their joint probability function is;

$$g(Y_1,\ldots\ldots,Y_n) = \prod_{i=1}^{n} f_i(Y_i)$$
$$= \prod_{i=1}^{n} \pi_i^{y_i} (1- \pi_i)^{1-Y_i} \quad (2.2.8)$$

Again, it will be easier to find the maximum likelihood estimates by working with the logarithm of joint probability function:

$$\text{Log}_e g(Y_i,\ldots\ldots,Y_n) = \log_e \prod_{i=1}^{n} \pi_i^{Y_i} (1- \pi_i)^{1-Y_i}$$

$$= \log_e \prod_{i=1}^{n} \left(\frac{\pi_i}{1+\pi_i}\right)^{Y_i} (1- \pi_i)$$

$$= \sum_{i=1}^{n} \left(Y_i \log_e \left(\frac{\pi_i}{1-\pi_i}\right) + \right) \sum_{i=1}^{n} \log_e (1- \pi_i) \quad (2.2.9)$$

Since $E(Y_i) = \pi_i$ for a binary variable, it follows from Equation (2.2.1), and according to Equation (2.2.5), the above Equation (2.2.9) can be expressed as follows:

$$\text{Log}_e L(\beta) = \sum_{i=1}^{n} Y_i \left(\sum \beta_k X_{ik}\right) - \sum_{i=1}^{n} \log_e [1+\exp(\sum \beta_k X_{ik})] \quad (2.2.10)$$

where $L(\beta)$ replaces $g(Y_1,\ldots\ldots,Y_n)$ to show explicitly that function can be viewed as the likelihood function of the parameters to be estimated, given the sample observation.

Equation (2.2.10) can be expressed more clearly as follows;

$$\text{Log}(L(\beta)) = \sum_{i=1}^{n} Y_i \log(\pi_i) + \sum_{i=1}^{n} (1 - Y_i) \log (1 - \pi_i)$$
$$= \sum_{i=1}^{n} Y_i \log(F(x_i'\beta)) + \sum_{i=1}^{n} (1 - Y_i) \log (1 - (F(x_i'\beta))$$
$$= \sum_{i;y_i=0}^{n} \log (1-F(x_i'\beta)) + \sum_{i;y_i=1}^{n} \log (1 - (F(x_i'\beta)) \quad (2.2.11)$$

The maximum likelihood estimates of β in the logistic regression model are those values of β that maximize the log-likelihood function in Equation (2.2.10). No

closed-form solution exists for the values of β in Equation (2.2.10) that maximize the log likelihood function. There are many widely used numerical search procedures; one of these employs iteratively reweighted least squares.

Once the maximum likelihood estimates are found, these values are substituted into the response function in Equation (2.2.1) to obtain the fitted response function.

The fitted logit model is as follows;

$$\widehat{\pi_i} = \frac{\exp(\sum b_k X_{ik})}{1+\exp(\sum b_k X_{ik})} \qquad (2.2.12)$$

If the logit transformation is utilized in Equation (2.2.5), the fitted response function in Equation (2.2.11) can be expressed as follows;

$$\widehat{L_i} = \sum b_k X_{ik} \qquad (2.2.13)$$

where,

$$\widehat{L_i} = \text{In}\left(\frac{\widehat{\pi_i}}{(1-\widehat{\pi_i})}\right) \qquad (2.2.14)$$

Once the fitted logit model has been obtained, the usual next steps are to examine the appropriateness of the fitted response function and , if the fit is good, to make a variety of inferences and predictions.

Grouped or replicated data

let N_i = total number of observations

n_i = no. of possibility among the interest category ($n_i \leq N_i$)

Therefore, π_i can be estimated as

$$\widehat{\pi_i} = \frac{ni}{Ni}$$

that is, the relative frequency can be used as an estimate of the true π_i corresponding to each X_i. If N_i is fairly large, $\widehat{\pi_i}$ will be a reasonably good estimate of π_i

Using the estimated $\widehat{\pi_i}$, the estimated logit can be obtained as

$$\widehat{L_i} = \text{In}\frac{\widehat{\pi_i}}{1-\widehat{\pi_i}} = \widehat{\beta_1} + \widehat{\beta_2}X_{i2} + \widehat{\beta_3}X_{i3} + \ldots + \widehat{\beta_k}X_{ik}$$

which will be a fairly good estimate of the true logit L_i if the no. of observations Ni at each X_i is reasonably large.

If N_i is fairly large and if each observation in a given $X_i$is distributed independently as a binomial variable, then

$$u_i \sim N\left[o, \frac{1}{N_i\pi_{i(}1-\pi_1)}\right]$$

that is, u_i follows the normal distribution with zero mean and variance equal to $1/[N_i\pi_i (1-\pi_i)]$. Therefore, as in the case of LPM the disturbance term in the logit model is hetroscedastic. Thus, instead of OLS the weighted lest squares (WLS) should be used .

For empirical purposes, replace the unknown π_i by $\widehat{\pi_i}$ and use

$$\hat{\sigma}^2 = \frac{1}{N_i\widehat{\pi_i}(1-\widehat{\pi_i})} \text{ as estimator of } \sigma^2$$

To resolve the problem of heteroscedasticity, Equation (2.2.6) can be transformed as]

$$\sqrt{W_i}L_i = \beta_1\sqrt{W_i} + \beta_2\sqrt{W_i}X_{1i} + \beta_3\sqrt{W_i}X_{2i} + \ldots\ldots\ldots + \beta_k\sqrt{W_i}X_{ki} + \sqrt{W_i}u_i \quad (2.2.15)$$

$$L_i^* = \beta_1\sqrt{W_i} + \beta_2X_{2i}^* + \beta_3X_{3i}^* + \ldots\ldots\ldots + \beta_kX_{ki}^* + V_i \quad (2.2.16)$$

where the weights $W_i = N_i\widehat{\pi_i}(1-\widehat{\pi_i})$;

L_i^* = transformed or weighted L_i; X_i^* = transformed or weighted X_i; and

v_i = transformed error term.

Now, the transformed error term v_i is homoscedastic. Estimate Equation (2.2.14) by OLS recall that WLS on the transformed data.

2.3 Probit Model

The model that emerges from the normal cumulative distribution function (CDF) is popularly known as the probit model, although sometimes it is also known as the normit model.

2.3.1 Functional Form

To motivate the probity model, assume that the decision of an event will occur or not depends on an unobservable utility index I_i, that is determined by one or more explanatory variables, in such a way that the larger the value of the index I_i, the greater the probability of occurrence of an event.

The index I_i can be expressed as

$$I_i = \sum \beta_k X_{ik} \quad (2.3.1)$$

Let Y_i = 1 if the event occurs and

= 0 if the event does not occur.

Now it is reasonable to assume that there is a critical or threshold level of the index, call it I_i^* such that if I_i exceeds I_i^*, the event will occur, otherwise it will not. The threshold, I_i^*, like I_i, is not observable, but it is assumed to be normally distributed with the same mean and variance it is possible not only to estimate the parameters of the index given in Equation (2.3.1) but also to get some information about the unobservable index itself.

Under the assumption of normality, the probability that I_i^* is less than or equal to I_i can be computed from the standard normal cumulative distribution function. That is,

$$\pi_i = P(Y = 1 \mid X) = P(I_i^* \leq I_i) = P(Z_i \leq \sum \beta_k X_{ik}) = F(\sum \beta_k X_{ik})$$

$$= F(x_i^* \beta) \qquad (2.3.2)$$

where $P(Y = 1 \mid X)$ means the probability that an event occurs given the value (s) of the X, or explanatory variable(s), i.e $Z \sim (0, \sigma^2)$.

F is the standard normal cumulative distribution function. The functional form of the probity model in two- variable case is.

$$F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-a}^{I_i} e^{-z^2/2} dz$$

$$= = \frac{1}{\sqrt{2\pi}} \int_{-a}^{\sum \beta_k X_{ik}} e^{-z^2/2} \qquad (2.3.3)$$

where

$I_i = \sum \beta_k X_{ik}$

= unobservable utility index (latent variable)

To obtain information on I_i, the utility index, as well as on β take the inverse of Equation (2.2.3) to obtain:

$$I_i = F^{-1}(I_i)$$

$$= F^{-1(}\pi_i)$$

$$= \sum \beta_k X_{ik}$$

Where F^{-1} is the inverse of the normal cumulative distribution function.

2.3.2 Estimation

Once the estimated I_i was obtained, estimating β are relatively straightforward. Since the normal equivalent deviate (n.e.d) or I_i will be negative whenever $\pi_i < 0.5$, in practice the number 5 is added to the n.e.d and the result is called a probit. Probit model is also constructed by assuming that a particular density underlies the data. Hence, this model is typical estimated using maximum likelihood rather than least squares.

Data for the probit model may also be two types. They are

(a) grouped data and

(b) ungrouped or individual data

As in the case of the logit model, a nonlinear estimating procedure based on the method of maximum likelihood can be used to estimate the probit model.

2.4 Comparison of the Models

In the LPM, the slope coefficients measure directly the change in the probability of an event occurring as the result of a unit change in the value of a regressor, with the effect of all other variables held constant. In the logit model the slope coefficient of a variable gives the change in the log of the odds associated with a unit change in that variable, again holding all other variables constant. But as noted previously, for the logit model the rate of change in the probability of an event happening is given by $\beta_j \pi_i(1- \pi_i)$, where $\beta_j$is (the partial regression) coefficient of the j^{th} regrerssor. But in evaluating π_i, all the variables included in the analysis are involved.

In the probit model, the rate of change in the probability is somewhat complicated and is given by $\beta_j f(Z_i)$ where $f(Z_i)$ is the density function of the standard normal variable and $\sum \beta_k X_{ik}$, that is, the regression model used in the analysis.

Thus, in both logit and probit models all the regressors are involved in computing the changes in probability, whereas in the LPM only the j^{th} regressor is involved. This difference may be one reason for the early popularity of the LPM model. One advantage of the LPM over logit or probit is that estimates of coefficients are available under complete or quasi complete separation.

The linear probability model has disadvantage. It places implicit restrictions on the parameters β, as $P(Y_i = 1) = E(Y_i) = x_i'\beta$ requires that $0 \leq x_i'\beta \leq 1$ for all i = 1,……..,n. Further, the error terms u_i are not normally distributed. This is because the variable y_i can take only the values zero and one, so that u_i is a random variable with discrete distribution given by

$u_i = 1 - x_i'\beta$ with probability $x_i'\beta$

$u_i = -x_i'\beta$ with probability $1-x_i'\beta$.

The distribution of u_i depends on x_i and has variance equal to $Var(u_i) = x_i'\beta(1-x_i'\beta)$, so that the error terms are heteroskedastic with variances that depends on β. The assumption that $E(u_i) = 0$ implies that OLS is an unbiased estimator of β (provided that the regressors are exogenous), but clearly it is not efficient and the conventional OLS formulas for the standard errors do not apply. Further, if the OLS estimates b are used to compute the estimated probabilities $\hat{P}[y_i=1] = x_i'b$, then this may give valuessmaller than zero or larger than one, in which case they are not real 'probabilities'. This may occur because OLS neglects the implicit restrictions $0 \leq x_i'\beta \leq 1$.

In most applications logit and probit models are quite similar, the main difference being that the logistic distribution has slightly fatter tails. That is to say, the conditional probability $\pi_i$approaches zero or one at a slower rate in logit than in probit. Therefore, there is no compelling reason to choose one over the other. In practice many researchers choose the logit model because of its comparative mathematical simplicity.

Though the models are similar, one has to be careful in interpreting the coefficients estimated by the two models. The reason is that, although the standard logistic (the basis of logit) and the standard normal distributions (the basis of probit) both have a mean value of zero and their variances are different;1 for the standard normal and $\pi^2/3$for the logistic distribution , where $\pi \approx 22/7$. Therefore, if the probit coefficient is multiplied by about 1.81(which is approximately = $\pi/\sqrt{3}$, the logit coefficient will be got approximately.

Incidentally, Amemiya (1981) has also shown that the coefficients of LPM and logit models are related as follows:

$$\beta_{LPM} = 0.25\ \beta_{Logit} \qquad \text{except for intercept}$$

and

$$\beta_{LPM} = 0.25\ \beta_{Logit} + 0.5 \qquad \text{for intercept}$$

Amemiya also suggested multiplying a logit estimate by 0.625 to get a better estimate of the corresponding porbit estimate. Conversely, multiplying a probit coefficient by 1.6 (=1/0.625) gives the corresponding logit coefficient.

CHAPTER III

DIAGNOSTIC STATISTICS FOR QUALITATVE RESPONSE MODELS

Some diagnostic statistics for qualitative response models namely, t-test (Z-test), the predictive quality (classification table and hit rate), and analysis of the residuals (in particular an LM test for heteroscedasticity), the likelihood ratio test and goodness-of-fit (R^2) will be presented in this Chapter.

3.1 Z statistic

The significance of individual explanatory variables can be tested by the usual t-test. The sample size should be sufficiently large to rely on the asymptotic expressions for the standard errors, and the t-test statistic then follows approximately the standard normal distribution. Since the method of maximum likelihood is generally a large sample method, the estimated standard errors are asymptotic. As a result, instead of using the t statistic to evaluate the statistical significance of a coefficient, (standard normal) Z statistic has to be used.

3.2 Likelihood Ratio (LR) Statistic

To test the null hypothesis that all the slope coefficients are simultaneously equal to zero, the equivalent of the F test in the linear regression model is the likelihood ratio (LR) statistic. Under the null hypothesis, H_0: $\beta_2 = \beta_3 = \ldots = \beta_k = 0$; the LR statistic follows the X^2 distribution with degree of freedom equal to the number of explanatory variables. That is,

$$2\ \mathrm{Ln}(L_1 - L_0) \sim X^2_{(k-1)}$$

where L_0 is the likelihood function when all parameters except the intercept, are set to zero and L_1 is likelihood function of the model of interest. Sometimes this measures similar to the R^2 of linear regression models. Joint parameter restrictions can be tested by the likelihood ratio test.

3.3 R^2 Statistic

A goodness-of-fit measure is a summary statistic indicating the accuracy with which the model approximates the observed data, like the R^2 measure in the linear

regression model. In linear regression model, R^2 is the most commonly used measure for assessing the discriminatory power of the model. R^2 possesses three properties. First, it is standardized to fall in the range (0, 1), equaling 0 when the model affords no predicted efficacy over the marginal mean and equaling 1 when the model perfectly accounts for, or discriminates among the responses. Second, it is non decreasing in X, meaning that it cannot decrease as regressors are added to the model. Third, it can be interpreted as the proportion of variation in the response accounted for by the regression.

In the case in which the dependent variable is qualitative, accuracy can be judged either in terms of the fit between the calculated probabilities and observed response frequencies or in terms of the model's ability to forecast observed responses. Contrary to the linear regression model, there is not single measure for the goodness-of-fit in qualitative response models and a variety of measures exists in nonlinear models.

Often, goodness-of-fit measures are implicitly or explicitly based on comparison with a model that contains only a constant as explanatory variable. A first goodness-of-fit measure defined by Amemiya (1981) is known as Pseudo-R^2 which is formulated by

$$\text{pseudo-}R^2 = 1 - \frac{1}{1+2(\log L_1 - \log L_0)/N}$$

where N denotes the number of observations.

An alternative measure suggested by McFadden (1974) is

$$\text{McFadden } R^2 = 1 - {}^{\text{Log}L_1}/_{\text{Log}L_0}$$

which is sometimes referred to as the likelihood ratio index. Like R^2, R^2_{MCF} also ranges between 0 and 1.

Another comparatively simple measure of goodness of fit is the count R^2, which is defined as:

$$\text{Count } R^2 = \frac{no.of\ correct\ predictions}{Total\ no.of\ observations}$$

Since the regressand in the model takes a value of 1 or zero, the number of correct predictions can be counted. If the predicted probability is greater than 0.5, it is classified as 1, but if it is less than 0.5, it is classified as 0.

3.4 Predictive Quality

Alternative specifications of the model may be compared by evaluating whether the model gives a good classification of the data into the two categories $y_i = 1$ and $y_i=0$. The estimated model gives predicted probabilities $\hat{\pi}_i$ for the choice $y_i = 1$, and this can be transformed into predicted choices by predicting that $\hat{y}_i = 1$ if $\hat{\pi}_i \geq c$ and $\hat{y}_I = 0$ if, $\hat{\pi}_i < c$. The choice of c can sometimes be based on the costs of misclassification. In practice on often takes $c = {}^1/_2$, or, if the fraction $\hat{\pi}_I$ of successes differs much from 50 per cent, one takes $c = \hat{\pi}_i$. This leads to a 2x2 classification table of the predicted responses $\hat{y}_i$ against the actually observed responses y_i. The hit rate is defined as the fraction of correct predictions in the sample. Formally, let w_i be the random variable indicating a correct prediction – that is, $w_i= 1$ if $Y_i = \hat{y}_I$ and $w_i =0$ if $Y_i \neq \hat{y}_I$, then the hit rate is defined by $h=\frac{1}{n}\sum_{i=1}^{n} w_i$.

In the population the fraction of successes is. If the prediction 1 with probability π_i and 0 with probability $(1-\pi_i)$ were randomly made, then a correct prediction is with probability $q= \pi^2 + (1- \pi)^2$. Using the properties of the binomial distribution for the number of correct random predictions, it follows that the 'random' hit rate h_r has expected value $E(h_r) = E(w) = q$ and variance $Var(h_r) = Var(w)/n = q(1-q)/n$. The predictive quality of the model can be evaluated by comparing hit rate h with the random hit rate h_r. Under the null hypothesis that the predictions of the model are no better than pure random predictions, the hit rate h is approximately normally distributed with mean q and variance q(1-q)/n. Therefore, reject the null hypothesis of random predictions in favor of the (one-sided) alternative of better- than random predictions if

$$z = \frac{h-q}{\sqrt{q(1-q)/n}} = \frac{nh-nq}{\sqrt{nq(1-q)}}$$

is large enough (larger than 1.64 at 5 per cent significance level). In practice, $q= \pi^2 + (1- \pi)^2$ is unknown and estimated by $\hat{\pi}^2 + (1- \hat{\pi})^2$, where is the faction of successes in the sample. In the above expression for the z-test, nh is the total number of correct predictions in the sample and *nq* is the expected number of correct random predictions.

3.5 Analysis of Residuals

3.5.1 Standardized Residuals and Consequences of Heteroskedasticity

The residuals u_i of a binary response model are defined as the differences between the observed outcomes y_i and the fitted probabilities $\hat{\pi}_i$. As the variance of y_i (for given values of x_i) is $\pi_i(1-\pi_i)$, the standardized residuals are defined by

$$u_i^* = \frac{y_i - \widehat{\pi_i}}{\sqrt{\widehat{\pi_i}(1-\widehat{\pi_i})}} \qquad (3.5.1)$$

A histogram of the standardized residuals may be used, to detect outliers. Further, scatter diagrams of these residuals against explanatory variables are useful to investigate the possible presence of heteroskedasticity. Heteroskedasticity can be due to different kinds of misspectfication of the model. It may be, for instance, that relevant explanatory variable is missing or that the function F is misspecified. In contrast with the linear regression model, where OLS remains consistent under heteroskedasticity, maximum likelihood estimators of binary response models become inconsistent under this kind of misspecification. For instance, if data generating process is a probit model but one estimates a logit model, then the estimated parameters and marginal effects are inconsistent and the calculated standard errors are not correct. However, as the differences between the probit function and the logit function are not so large, the outcomes may still be reasonably reliable.

3.5.2 Likelihood Ratio Test for Heteroskedasticity

A formal test for heteroskedasticity can be based on the index model $y_i^* = x_i^*\beta + u_i$. Until now it has been assumed that the error terms u_i all follow the same distribution (described by F). As an alternative can be considered the model where all μ_i/σ_i follow the same distribution F where

$$\sigma_i = u^{z_i'\gamma}$$

with z_i a vector of observed variables. The constant term should not be included in this vector because the scale parameter of a binary response model should be fixed,

independent of the data. Assume that the density function *f* (the derivative of F) is symmetric – that is, *f*(t) = *f*(-t). It then follows that

$P[y_i = 1] = [y_i^* \geq 0]$

$= P[u_i \geq -x_i'\beta]$

$= P[(u_i/\sigma) \geq -x_i'\beta/\sigma]$

$= P[(u_i/\sigma) \leq x_i'\beta/\sigma]$

$= F(x_i'\beta/\sigma)$, so that

$$P[y_i = 1] = F(x_i'\beta/u^{z_i'\gamma}) \qquad (3.5.2)$$

The null hypothesis of homoskedasticity corresponds to the parameter restriction $H_o : \gamma = 0$. This hypothesis can be tested by the LR-test. The unrestricted likelihood function is obtain from the log-likelihood by replacing the term

$\pi_i = F(x_i'\beta)$ by $\pi_i = F(x_i'\beta/u^{z_i'\gamma})$.

3.5.3 Lagrange Multiplier Test for Heteroskedasticity

Alternative is to use the LM-test, so that only the model under the null hypothesis (with $\gamma = 0$) needs to be estimated. By working out the formulas for the gradient and the Hessian of the unrestricted likelihood, it can be shown that the LM-test can be performed as if Equation (3.5.2) were a non-linear regression model.

First estimate the model without heteroskedasticity – that is, under the null hypothesis that $\gamma = 0$. This amounts to estimating the model $P(y_i = 1) = F(x_i'\beta)$ by ML. The residuals of this model are denoted by

$u_i = y_i - \widehat{\pi_i}$

$= y_i - F(x_i'\beta)$

As a second up step, regress the residuals u_i on the gradient of the non-linear model $P(y_i = 1) = F(x_i'\beta/u^{z_i'\gamma})$, taking into account that the residuals are heteroskedastic. This amounts to applying (feasible) weighted least squares- that is, OLS after division for the i^{th} observation by the (estimated) standard deviation. The variance of the 'error term' $y_i - \pi_i$ is $Var(y_i - \pi_i) = Var(y_i) = \pi_i(1-\pi_i)$. π_i is replaced by $\widehat{\pi_i}$ obtained in the first step, so that the weight of the i^{th} observation in WLS is given

by $1/\sqrt{\hat{\pi}_i(1-\hat{\pi}_i}$. Further, the gradient of the function F ($x_i'\beta/u^{z_i'\gamma}$) in Equation (3.5.2) , when evaluated at γ=0, is given by

$\frac{\partial F\,(x_i'\beta/u^{z_i'\gamma})}{\partial F} = f\,(x_i'\beta)$ X, $\frac{\partial F\,(x_i'\beta/u^{z_i'\gamma})}{\partial F} = -\,f\,(x_i'\beta)\,x_i'\beta z.$

Therefore, the required auxiliary regression in this second step can be written in terms of the standardized residuals as

$$u_i^* = \frac{y_i-\hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}} = \frac{f\,(x_i'b)}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}\,x_i'\delta_1 + \frac{f\,(x_i'b)x_i'b}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}\,z_i'\delta_1 + n_i. \qquad (3.5.3)$$

Under the null hypothesis of homoskedasticity, there holds that LM = nR^2_{nc}, where nR^2_{nc} denotes the non-centered R^2-that is , the explained sum of squares of Equation (3.5.3) is divided by the non-centered total sum of squares $\sum_{i=1}^{n}(u_i^*)^2$. As the regression in Equation (3.5.3) does not contain a constant term on the right-hand side, one should take here the non-centered R^2 defined by $R^2_{nc} = \sum(u_i^*)^2/\sum(u_i^*)^2$, where $\hat{u}^*_i$ denotes the fitted values of the regression in Equation (3.5.3). Reject the null hypothesis for large values of the LM-test, and under the null hypothesis of homoskedasticity (γ = 0) it is asymptotically distributed as X^2 (g), where g is the number of variables in Z_i-that is , the number of parameters in γ.

CHAPTER IV

APPLICATION OF LINEAR PROBABILITY LOGIT AND PROBIT MODELS

4.1 Introduction

In this chapter, the application of linear probability, logit and probit models are demonstrated by survey data. The survey data used in this chapter are provided by Ma Moe Sandar Oo who collected the data for her Master of public Administration Thesis. The data were responses of the mother of 300 children under 3 years of age in Thingungyun Township. The weights of the children were assessed from the standard weight chart using by Township Health Center. There are four different colours (red, yellow, green, white) to present the condition of child's weight on this chart. Red colour represents the child's weight, which reflects the severe malnutrition. Yellow colour stands for moderate malnutrition of child's condition and green colour signifies as good condition. White colour zone shows another form of malnutrition which is known as over-eight child. In general, malnutrition can be defined as underweight in developing countries, which is a serious public health problem that has been linked to a substantial increase in the risk of morbidity and mortality. The term malnutrition refers to both over-nutrition and under-nutrition. Malnutrition is a general term for a medical condition caused by an improper or inadequate diet and nutrition. In This study, if child's weight colour is green, the child can be determined by nutrition, and if child's weight colour is yellow (or) red, the child can be determined by malnutrition. The white colour case is very rare in Myanmar. So, white colour case is omitted from this study.

Out of these collected information, mother's age, mother's education level and child's weight colour variable are used to develop the models. Mother's education levels are divided into 4 categories such as primary, middle, high, and graduate. Child's weight colour is divided into 3 categories such as green, yellow, and red. To estimate the models, mother's age and mother's education level are used as independent variables and child's weight colour is used as dependent variable.

4.2 Models for Child's Weight Colour

In construction the models, the variables are noted as:

Y_i	= 1	if child's weight colour is green
	= 0	otherwise
$MAGE_i$	=	mother's age
$MEDU_i$	= 1	if mother's education is primary school level
	= 0	otherwise
$MEDU_2$	= 1	if mother's education is middle school level
	= 0	otherwise
$MEDU_3$	= 1	if mother's education is high school level
	= 0	otherwise

The Linear Probability Model (LPM)

$$Y_i = \beta_1 + \beta_2 MAGE_i + \beta_3 MEDU_1 + \beta_4 MEDU_2 + \beta_5 MEDU_3 + u_i$$

where u_i is disturbance term and the unknown parameters β_1, β_2, β_3, β_4 and β_5 in the LPM are estimated by using the weighted least squares method using Statistical Package for Social Science (SPSS). It is assumed that the variance of u_i is proportional to the variable $MAGE_i$.

The Logit Model

The ligit model here can be written as;

$$L_i = \ln \frac{\pi_i}{1-\pi_i} = \beta_1 + \beta_2 MAGE_i + \beta_3 MEDU_1 + \beta_4 MEDU_2 + \beta_5 MEDU_3 + u_i$$

where π_i = the probability that child's weights colour is green

1- π_i = the probability that child's weight colour is not green

The Probit Model

Assume that I_i = unobservable untility index (latent variable)

I_i^* = critical or threshold level of the index

If I_i exceeds I_i^*, the child's weight colour will be green , otherwise it will not.

$$I_i = \beta_1 + \beta_2 MAGE_i + \beta_3 MEDU_1 + \beta_4 MEDU_2 + \beta_5 MEDU_3$$

$$I_i = F^{-1}(I_i)$$

$$= F^{-1}(\pi_i)$$

$$= \beta_1 + \beta_2 MAGE_i + \beta_3 MEDU_1 + \beta_4 MEDU_2 + \beta_5 MEDU_3$$

where F^{-1} is the inverse of the normal cumulative distribution function (CDF).

$$\pi_i = P_r(Y = 1/X)$$

$$= P_r(I_i^* \leq I_i)$$

$$= F(\beta_1 + \beta_2 MAGE_i + \beta_3 MEDU_1 + \beta_4 MEDU_2 + \beta_5 MEDU_3)$$

π_i represents the probability that child's weight colour is green , it is measured by the area of the standard normal curve from = $\alpha$to I_i.

The unknown parameters in the logit and probit models are estimated by using method of Maximum Likelihood and Enter Regression Method through computer software of (SPSS).

4.3 Results

The estimated models and their results are described in this section. The estimated standard error (se) and computed p-values are shown in parentheses.

Linear Probability Model

$\hat{Y}_i$ = 1.079 - 0.009 $MAGE_i$ – 0.052 $MEDU_i$ – 0.009 $MEDU_2$ + 0.139 $MEDU_3$

se	(0.044)	(0.003)	(0.094)	(0.065)	(0.083)
P. values	(0.000)	(0.001)	(0.580)	(0.131)	(0.096)

R^2 = 0.146, $\overline{R}^2$ = 0.134, count R^2 = 0.76, Pseudo R^2 = 0.15,

McFadden R^2 = 0.157, F = 12.598

According to the p-values it van be said that the variable $MEDU_i$ and $MEDU_2$ are insignificant and the variables $MAGE_i$ and $MEDU_3$ are significant at % level, and 10% level, respectively. The insignificant variables $MEDU_1$ and $MEDU_2$ are dropped from the model and estimate the model for child's weight colour with the variable $MAGE_i$ and $MEDU_3$.

The re-estimated model is as follows:

$\widehat{Y}_i =$ 1.082 – 0.012 $MAGE_i$ + 0.217 $MEDU_3$

se (0.044) (0.022) (0.063)

p.values (0.000) (0.000) (0.000)

R^2 = 0.105, $\overline{R}^2$ = 0.102, count R^2 = 0.76, Pseudo R^2 = 0.141,

McFadden R^2 = 0.15, F = 34.87

The results imply that the variable $MAGE_i$ and $MEDU_3$ are important factors in explaining the changes of probability of child's nutrition . It can be said that if the mother's age increases by 1-year and being mother's education in high school level remained unchanged, the probability of child's nutrition will decrees by about 1.2% IF the mother's education is in high school level and being mother's age remained unchanged, the probability of child's nutrition will increase by 21.7%.

Logit Model

$\widehat{L}_i$ = In $\frac{\hat{\pi}}{1-\hat{\pi}}$

= 11.146 - 0.248 $MAGE_i$ – 2.657 $MEDU_i$ – 3.029 $MEDU_2$ + 0.268 $MEDU_3$

se	(1.768)	(0.044)	(1.134)	(1.052)	(1.271)
p.values	(0.000)	(0.000)	(0.019)	(0.004)	(0.833)

count R^2 = 0.79 , pseudo R^2 = 0.102, McFaddenR^2 = 0.211, X^2 = 64.241

According to the p.values it can be said that each variable, except $MEDU_3$ is significant at 1% level and X^2 = 64.241 indicates that the whole model is highly significant . The insignificant variable $MEDU_3$ is excluded from the model and estimate the model for child's weight colour with the variables $MAGE_i$, $MEDU_i$, and $MEDU_2$. The re-estimated model is as follows;

$\widehat{L}_i$ = ln $\frac{\hat{\pi}}{1-\hat{\pi}}$

= 10.986 - 0.248 $MAGE_i$ – 2.488 $MEDU_i$ – 2.86 $MEDU_2$

se	(1.577)	(0.044)	(0.765)	(0.693)
p.values	(0.000)	(0.000)	(0.001)	(0.000)

count R^2 = 0.62 , pseudo R^2 = 0.187, McFaddenR^2 = 0.211, X^2 = 69.195

From the re-estimated ligit model, $MAGE_i$, $MEDU_1$ and $MEDU_2$ are found to be important factors in explaining the changes of the log of odds for child's nutrition. It can be found that being other factors remained unchanged, with an increase of 1-year of mother's age, there is an expectation of decrease in the log of odds for child's nutrition about 0.25. Moreover, if the mother's education is in primary school level, it is expected to have a decrease of 2.488 and if the mother's education is in middle school level, it is expected to have a decrease of 2.86, in the log of odds for child's nutrition, respectively.

Probit Model

$\hat{I}_i$= -2.658 - 0.002 $MAGE_i$ + 0.103 $MEDU_i$ + 0.017 $MEDU_2$ + 0.050 $MEDU_3$

se	(0.164)	(0.005)	(0.098)	(0.079)	(0.056)
p.values	(0.000)	(0.763)	(0.291)	(0.826)	(0.446)

count R^2 = 0.76 , X^2 = 101.624

According to the p. values it can be said that the all variables are insignificant at 1% and 10% level.

In summarizing the results and findings of estimated models, the diagnostic statistics such as p-values, computed F-values and computed X^2 values indicate that the LPM and logit model are found to be significant models.

From the estimated LPM and logit models the variable mother's age and mother's education are important factors in explaining the changes of child's nutrition.

For the estimated models, the count R^2 value is high, whereas the McFadden R^2 vale and pseudo R^2 are low. Although these R^2 values are not directly comparable, they can give some idea about the orders of magnitude. Besides, one should not overplay the importance of goodness of fit in models where the regressand is dichotomous. The estimated R^2 may seem rather low, but in view of the large sample size, this R^2 is still significant on the basis of the F test.

CHAPTER V

CONCLUSION

In this paper, qualitative response models: linear probability, logit, and probit models in which the dependent variable involves only two qualitative choices are studied together with their specification and estimation procedure. These models are valuable in the analysis of survey data. The important characteristics of this study are as follows:

1. Qualitative response regression models refer to models in which the response, or regressand, variable is not quantitative or an interval scale.
2. The simplest possible qualitative response regression model is the binary model in which the regressand is of the yes/no or presence / absence type.
3. The simplest possible binary regression model is the linear probability model (LPM) in which the binary response variable is regressed on the relevant explanatory variables by using the standard OLS methodology. Simplicity may not be a virtue here, fore the LPM suffers from several estimating problems. Even if some of the estimation problems can be overcame, the fundamental weakness of the LPM is that it assumes that the probability of something happening increases linearly with the level of the regressor. This very restrictive assumption can be avoided by using the logit and probit models.
4. In the logit model the dependent variable is the log of the odds ratio, which is a linear function of the regressors. The probability function that underlies the logit model is the logistic distribution. If the data are available in grouped form, OLS can be used to estimate the parameters of the logit model, provided the heteroscedastic nature of the error term is taken into account explicitly. If the data are available at the individual, or micro level, nonlinear-in-the-parameter estimating procedures, like as method of maximum likelihood can be used.
5. If the normal distribution is chosen as the appropriate probability distribution, then the probit model can be used. This model is mathematically a bit difficult as it involves integrals.
6. The estimated model can be interpreted in terms of the signs and significance of the estimated coefficient. The model can be evaluated in different ways, by

using diagnostic tests (t or Z-test, LR-test) and by measuring the model quality (goodness of fit R^2).

As an application, these models are developed and estimated by using SPSS computer software with the survey data of the mother of 300-children in Thingungyun Township.

The findings are as follows:

(1) According to the computed F value and X^2 value, the LPM and logit models are significant but probit is not.

(2) IN the estimated LPM, it can be concluded that the variables mother's age and mother's education are found to be important factor in explaining the child's nutrition. From the estimated model, being other factors remained unchanged, an increase in the mother's age of 1-year will decrease the probability of child's nutrition by about 1.2%.

(3) In the estimated logit model, it can be said that the mother's age and mother's education are found to be important factors in explaining the child's nutrition From the estimated model, being other factors remained unchanged, an increase in the mother's age of 1-year will decrease the odds for child's nutrition by about 22%. If the mother's education is in primary school level, it is expected to have an decrease about 92%, if the mother's education is in middle school level, it is expected to have a decrease about 94% in the odds for child's nutrition, respectively.

REFERENCES

1. ALFRRED DEMARIS, "*Regression with Social Data: Modeling continuous and Limited Response Variables"*, Published by John Willey & Sons Inc, Hoboken, New Jersey, U.S.A.

2. Christiaan H, P. de Boer, P.H.Franses, T.Kloek, and H.K. van Dijk (2004) *"Econometric Methods with Applications in Business and Economics"*, First Edition, Oxford University Press.

3. G.K. David and Mitchel K. (2002) *" Logistic Regression"*.A Self-Learning Text Second Edition, Springer-Verlag, New York, Inc.

4. Gujarati, D.N. and Sangeetha (2008) *" Basic Econometrics*", Fourth Edition, McGraw-Hill Publishing Company Ltd.

5. MarnoVerbeek (2008),*"A guide To Modern Econometrics"*, Third Edition, John Wiley & sons, Ltd.

6. Moe Sandar Oo (July 2009), "*A study on the Influential Factors of Underweight children (age under 3) in Thingungyun Township"*, MPA Thesis.

7. Neter J., Michael H.K, Christopher J.N, and William Wasserman, (1996), *" Applied Linear Statistical Models"*, 4th Edition, McGraw-Hill.

8. Takeshi Amemiya (1918) "*Qualitative Response Models"* A Survey, JASA. Vol. XIX.

9. William. H. Green (2000), *"Econometric analysis"* 4th edition.
   Copyright 2000 by Prentice- Hall,Inc, upper Saddle river, New Jersey 07458 Printed in the United States of America.