The Economic Journal of Nepal, Vol.12, No.1, Issue 45, January-March 1989 © DOE-TU

An Introduction to Probit and Logit Models.

KISHOR KUMAR GURU GHARANA*

INTRODUCTION

Economics as a theory of choice, can be applied not only to questions about how much to produce or consume but also "whether" to produce or consume a certain item. Thus, economic units often have to choose from a finite set of alternatives. The econometric models designed to deal with such situations are called models with qualitative dependent variables. Probit and Logit models are the most popular models of choice from a finite set of alternatives. Some examples of situations where such choices arise are:

i. a household deciding whether to buy or rent a suitable dwelling;

ii. a consumer choosing one from several shopping areas and a mode of transportation;

iii. members of a household deciding whether to take part-time or full-time employment, or whether or not to seek a second job;

iv. a person deciding whether or not to attend college; and,

v. a committee member deciding whether to vote 'yes' or 'no' on a particular proposal.

If the situation involves only two alternatives (as in yes-no situation), it is called binary choice situation. If, however, the situation involves choice from a finite set of (more than two) discrete alternatives, it is a case of multinomial choice.

We will study both types of situations in this article, and also derive Probit and Logit models using Random Utility Models.

BINARY CHOICE MODELS

Suppose $$Y_i = \begin{cases} 1 \text{ if } X_i'b \geq s_i \\ o \text{ if } X_i'b < s_i \end{cases}$$

where $Y_i = 1$ means the individual i chooses the object and $y_i = o$ means the individual does not choose the object. Also, s_i is called the threshold level of stimulus for individual i, and X_i' is the vector of attributes of the object for the individual i. The vector b is considered to be constant over i.

*Dr. Guru Gharana is Assistant Lecturer at Central Department of Economics, Tribhuvan University, Kirtipur, Kathmandu, Nepal.

In this case we have,

(a) $\Pi_i = Pr\ (y_i = 1) = Pr\ \{\ s_i \leq X_i'b | X_i\}$

$$= F\ (X_i'b),\ \ldots\ldots\ (1)$$

where F is the distribution function of s_i.

(b) $1 - \Pi_i = Pr\ (y_i = o) = 1-F(X_i'b)\ \ldots\ldots\ (2)$

Therefore, the likelihood for n individuals is given by

$$L(y_1,\ y_2,\ \ldots\ y_n) = \prod_{i=1}^{n} F(X_i'b)^{y_i} [1-F(X_i'b)]^{1-y_i}\ \ldots\ldots\ (3)$$

If F is assumed standard normal, then we obtain Probit model. On the other hand if F is assumed logistic, that is,

$$F(s) = \frac{1}{1+e^{-s}} = \frac{e^s}{1+e^s},$$

then we obtain Logit model. Equation (3) can be used for maximum likelihood estimation of b.

PROBIT MODEL

Suppose $F(s) \sim N(o,1)$

then $\Pi_i = N(X_i'b)\ \ldots\ldots\ (4)$, and

$N^{-1}\ (\Pi_i) = X_i'b\ \ \ \ldots\ldots\ (5)$

The variate $N^{-1}\ (\Pi_i)$, or, sometimes $N^{-1}\ (\Pi_i) + 5$ is called Probit. The estimation model will be,

$$N^{-1}\ (p_i) = X_i'b + u_i\ \ldots\ldots\ (6),$$

where p_i is the observed proportion of choice when stimuli are provided to T_1, T_2, ... In subjects in 1st, 2nd, ...nth groups respectively. Also, $N^{-1}\ (p_i)$ is observed Probit.

The problem here is that the error term is heteroscedastic (see Johnston for a proof).

Hence, we have to estimate b in the Probit model by generalised least squares method.

LOGIT MODEL

Let the probabilities Π_i be modelled by the logistic distribution.

That is,

$$\Pi_i = F(X_i'b) = \frac{e^{X'_i b}}{1+e^{X'_i b}} = \frac{1}{1+e^{-X'_i b}} \quad \ldots\ldots (7)$$

Clearly $o<\Pi_i<1$ and Π_i increases monotonically with the stimulus $X_i'b$.

We obtain from (7),

$$\ln\left(\frac{\Pi_i}{1-\Pi_i}\right) = X_i'b \quad \ldots\ldots (8)$$

Thus, the logit $\ln\left(\frac{\Pi_i}{1-\Pi_i}\right)$ is the natural log of the true odds ratio and can also be interpreted as the utility for individual i of choosing alternative one over alternative two.

But the observed and estimated relation is,

$$Y_i = \ln\left(\frac{p_i}{1-p_i}\right) = X_i'b+u_i \quad \ldots\ldots (9)$$

In this case too the variance of u_i changes with i (heteroscedasticity) and generalised least squares method is required for estimation.

RANDOM UTILITY MODELS

Logit and Probit models can also be derived from Random Utility Models. In this case let us consider multiple choice situation. Suppose there are J objects of choice, and we are studying a group of individuals with same representative tastes plus idiosyncracies reflected in the error term as in the following. For an individual, the utility function in linear form is expressed as,

$$U_j = X_j'b+u_j, \; j = 1, 2, \ldots\ldots J, \quad \ldots\ldots (10)$$

where X_j is the vector of the measurable attributes of good j and their interaction with the character of the individual. Thus, $X_j'b$ reflects the representative tastes of the group. The error term u_j reflects the effects of unobservable attributes and characteristics.[j]

The linear form of utility function is not so restrictive because any transformation of variables can be used. Secondly, the utility function can be viewed as a linear approximation to a more complicated functional form where error term also includes approximation error.

MULTINOMIAL LOGIT MODEL

We will now discuss the Logit model first, because, we want to show that the Probit model is theoretically more appropriate by first pointing out the problems with Logit model.

The distinctive characteristic of Logit model stems from the assumption made concerning the error term. The error term for each alternative is assumed to be "independently and identically distributed" extreme value variable (also called Weibull Distribution). The scale parameter is assumed to be unity (because utility is unique up to a scalar multiple), and centering parameter is assumed zero (because only the differences in utility matters).

Thus we assume $Pr(u_i \leq a) = e^{-e^{-a}}$, and ..... (11)

$$Pr(u_1 \leq a_1, u_2 \leq a_2 \; ..... \; u_j \leq a_j)$$

$$= Pr(u_1 \leq a_1) . Pr(u_2 \leq a_2) \; .... \; Pr(u_j \leq a_j) \; .... (12)$$

The individual selects the alternative that yields the highest utility level. That is alternative k is chosen if

$$Uk \geq U_j \text{ for all } j \neq k$$

In probability terms,

$$Pk = Pr(k \text{ chosen by the individual})$$

$$= Pr(Uk \geq U_1, Uk \geq U_2 \; ....., Uk \geq U_j)$$

$$= Pr(Uk \geq U_1) \; P(Uk \geq U_2) \; ..... \; Pr(Uk \geq U_j)$$

$$= Pr[u_1 \leq (X_k - X_1)'b + Uk] . Pr[U_2 \leq (X_k - X_2)'b + Uk] \; ..... \; Pr[U_j \leq (X_k - X_j)'b + Uk]$$

$$= \int_{-\infty}^{\infty} f(uk)duk \int_{-\infty}^{(X_k - X_1)b+uk} f(u_1)du \; ... \int_{-\infty}^{(X_k - X_j)'b+uk} f(u_J)du_J \; ..... (13)$$

If the probability density function f() is the extreme value (Weibull) density function, then the integral in (13) reduces to,

$$Pk = \frac{\exp(X'_k b)}{\sum_{1}^{j} \exp(x'_j b)} \; ..... (14)$$

Clearly, equation (14) is the Logit model which is just an extension of (7) above for the case of many objects of choice.

The likelihood function for N individuals and J alternatives is,

$$L = \prod_{i=1}^{N} \prod_{j=1}^{J} \left(\frac{\exp(x_{ij}'b)}{\sum_{j=1}^{j} \exp(x_{ij}'b)} \right)^{y_{ij}}, \quad \ldots\ldots (15)$$

where, $y_{ij} = 1$ if alternative j is chosen by individual i, otherwise zero. This likelihood function can be maximised with respect to b to yield maximum likelihood estimator. Note that for two objects, the Logit model for individual i can be written as,

P_{i1} = Probability of individual i choosing object 1 over 2

$$= \frac{e^{x_{i1}'b}}{e^{x_{i1}'b} + e^{x_{i2}'b}} \quad \ldots\ldots (16)$$

Now we will discuss some problems with the Logit model. As will be shown below, Logit model satisfies Independence of Irrelevant Alternatives (IIA) axiom. However, this axiom is implausible for alternative sets containing choices that are close substitutes. Another important problem with Logit specification involves the equality of cross-partial elasticities of the choice probabilities at the individual level.

If X_{k1} be attribute 1 for alternative k, then

$$\frac{\partial \log P_i}{\partial \log X_{ki}} = \frac{\partial \log P_j}{\partial \log X_{k1}} = b_1 P_k X_{k1} \quad \ldots\ldots (17)$$

However, there is little reason why the cross partial elasticities should be exactly equal. Finally, the independence between errors may be violated in practice.

MULTINOMIAL PROBIT MODEL

This model avoids the problems mentioned above. Multinomial Probit begins with the same utility maximisation assumption as the Multinomial Logit. The basic difference is in the specification of the probability distribution of the error term. The Probit model assumes that the errors U_j's for an individual are distributed as multivariate normal, and that dependence between error terms is possible. Let the multinormal density function be,

f(u,S) where S is the variance coviriance matrix.

Then we have,

$$P_k = P(U_k \geq U_1, U_k \geq U_2, \ldots, U_k \geq U_J)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{(X_k - X_1)'b + u_k} \cdots \int_{-\infty}^{(X_k - X_J)'b + u_k} f(u,S)\, du_1 \ldots du_J du_k \quad \ldots\ldots (18)$$

This choice probability, however, cannot be written in a closed analytic form without an integral as was possible for Logit Model. Thus, the likelihood function also involves these integrals. Therefore, interative maximum likelihood maximisation has to be used to estimate the parameters.

COMPARISON OF LOGIT AND PROBIT MODELS

Apart from different error term distribution, the two models can be compared as follows:

(a) Logit model is computationally simpler, more popular, and less expensive. Probit model has high computer costs, because for each observation in each interation, an integral of order J-1 must be evaluated.

(b) The choice probability can be written in a simple closed form without integrals for Logit. Moreeover, the formula for probability of choice allows a ready interpretation in terms of the relative representative utilities of alternatives. In case of Probit, the integrals cannot be avoided.

(c) The Logit probability choice formula makes it simple to ascertain the effect of introducing a new alternative to an alternative set because Logit satisfies Independence of Irrelevant Attributes (IIA). This is a strength as well as a weakness of Logit model.

(d) There are problems with IIA because it rules out a pattern of differential substitutability and complementarity between alternatives. To make these concepts clear, we will next discuss IIA axiom and and derivation of Logit from it.

(e) We have already mentioned some other problems with Logit model such as independence of errors assumption and equality of cross-partial elasticities.

INDEPENDENCE OF IRRELEVANT ALTERNATIVES

The Independence of Irrelevant Alternatives Axiom says that relatives odds of one alternative being chosen over a second should be independent of the presence or absence of unchosen third alternative.

Let s = vector of measured attributes of decision makers
B = a choice set of objects; $B \subseteq X$
X = Universe of objects of choice

then, for all $B \subseteq X$ and members $x, y \in B$ the IIA axiom says,

$$\frac{Pr[y/s, \{x,y\}]}{Pr[x/s, \{x,y\}]} = \frac{Pr(y/s,B)}{Pr(x/s,B)} \quad \ldots\ldots (19)$$

In words, the relative odds of y being chosen over x in a multiple choice situation B, where both x,y are available, equals the relative odds of a binary choice of y over x.

If new objects are added in the choice set, the choice probabilities of both x and y are reduced proportionately. Thus, it becomes easy to ascertain the effects of introducing new objects of choice.

However, when objects which are very similar to the existing ones are added, this axiom tends to break down. Suppose, for example, that a population has alternatives of travel by a car and bus, and that 2/3rd choose car. If a new colour bus (otherwise identical to the existing one) is added, then intuitively 2/3rd of the population should still choose car and the remainder (those who preferred bus) should split between the bus alternatives. But IIA axiom requires that the probabilities of choice of both car and first bus should decline proportionately. This requires that Pr(car) drops to ½ and Pr(Red bus or first bus) equals Pr(Blue bus or 2nd bus) and both equal to ¼.

Obviously, the reduction in the probability of choice of car is counter intuitive.

Now we show that Logit model satisfies IIA and, therefore, is subject to the same criticism.

According to equation (14)

$$\frac{P_k}{Pl} = \frac{\exp(X_k' b)}{\exp(X_l' b)} \quad \ldots\ldots (20)$$

Thus, the odds ratio of choice probabilities between objects k and l does not depend on the set j = 1,2 ... J, as long as the alternatives k and l are available. This is what IIA stipulates.

Another way to prove the compatibility of Logit model and IIA is to derive the Logit model starting from IIA assumption. This is what we intend to do next.

DERIVATION OF LOGIT MODEL FROM IIA

Suppose we have a choice set B containing alternatives x, y and z. From equation (19) which states the IIA principle we obtain,

$$Pr(y/s,B) = \frac{P_{yx}}{P_{xy}} \cdot Pr(x/s,B), \quad \ldots\ldots (21)$$

where Pxy = Pr[x/s,{x,y}] = probability of choosing x over y in a binary choice situation.

Similarly we define Pyx and also put Pxx = ½

Then we have,

$$1 = \sum_{y \varepsilon B} Pr(y/s,B) = \left(\sum_{y \varepsilon B} \frac{Pyx}{Pxy}\right) Pr(x/s,B) \quad \ldots\ldots (22)$$

Now multiple choice selection probabilities can be written in terms of binary odds as,

$$Pr(x/s,B) = \frac{1}{\sum_{y \varepsilon B} \frac{Pyx}{Pxy}} \quad \ldots\ldots (23)$$

Moreover, using different permutations of x,y, and z we obtain from (21),

$$\frac{Pyx}{Pxy} = \frac{Pr(y/s,B)}{P(x/s,B)} = \left(\frac{Pyz}{Pzy}\right) \left(\frac{Pxz}{Pzx}\right), \quad \ldots\ldots (24)$$

since Pr(z/s,B) cancels out.

Therefore, we have

$$\sum_{y \varepsilon B} \left(\frac{Pyx}{Pxy}\right) = \frac{\sum_{y \varepsilon B} \left(\frac{Pyz}{Pzy}\right)}{(Pxz/Pzx)} \quad \ldots\ldots (25)$$

Now consider z as a "bench-mark" member of the alternatives' set B and define V(s,x,z) as the utility of choosing x when x and z are both available.

Let $V(s,x,z) = \log\left(\frac{Pxz}{Pzx}\right)$, or ,

$$\frac{Pxz}{Pzx} = \exp[V(s,x,z)] \quad \ldots\ldots (26)$$

Then from equations (23), (25) and (26) we have,

$$Pr(x/s,B) = \frac{1}{\sum_{y \varepsilon B} \frac{Pyx}{Pxy}} = \left(\frac{Pxz}{Pzx}\right) \sum_{y \varepsilon B} \left(\frac{Pyz}{Pzy}\right) = \frac{e^{V(s,x,z)}}{\sum_{y \varepsilon B} e^{V(s,y,z)}} \quad \ldots\ldots (27)$$

Finally, if we express $V_i = X_i'b$, then we have Logit model from (27).

CONCLUSION

We saw that Logit model is preferable on the grounds of simplicity of expression and interpretation, and of less costly estimation. But the error term assumption of Probit model is much more flexible, the parameters of the utility function can vary across individuals and cross-partial elasticities of choice probability need not be equal. Moreover, the satisfaction of IIA principle may be the strength of Logit model in some situations and it may also be its weakness when alternative sets contain choices that are close substitutes (as the Red bus-Blue bus case).

With respect to the estimation technique, it can be said in general that the appropriate estimation technique depends upon the nature of the sample data that are available. If repeated observations exist on individual decision-makers, a feasible generalised least squares estimation procedure can be used. If only a few observations exist for each decision maker, maximum likehood estimation is possible for the two models, although numerical optimisation methods must be employed. A similar maximum likehood estimation procedure can be used when the number of choice alternatives facing an individual is greater than two but still relatively few.

Hausman and Wise have developed an extremely accurate lagorithm which calculates normal cumulative distribution function rapidly, thus, reducing the computer costs for Multinomial Probit model. However, this approach is still not suited for more than four alternatives due to excessive computer costs.

As regards forecasting performance, Hausman and Wise present an artificial forecasting example which shows that a Probit model that assumes independent error term is likely to generate forecasts almost identical to those of a Logit model. However, a Probit model that permits variation in tastes among individuals can lead to substantially different forecasts.

We conclude this introductory article with this remark: The Probit model is intuitively appealing as it provides a sound theoretical basis for specification in which error terms are not necessarily distributed independently across alternatives, and can easily incorporate stochastic parameters. But it will not be popular until computational costs can be greatly reduced.

SELECTED REFERENCES

Akin, J.S., Guilkey, D.R., and Sickles, R. (1979), "A Random Coefficient Probit Model With an Application to a Study of Migration," Journal of Econometrics, 11.

Amemiya, T. (1973), "Qualitative Response Models", Annals of Economic and Social Measurement, 4.

Amemiya, T. (1978a), "On a Two-step Estimator of a Multivariate Logit Model", *Journal of Econometrics*, 8.

____ (1978b), "The Estimation of a Simultaneous Equation Generalized Probit Model", *Econometrica*, 46.

____ (1981), "Qualitative Response Models: A Survey", *Journal of Economic Literature*, 19.

Ashford, J.R. and Sowden, R.R. (1970), "Multivariate Probit Analysis", *Biometrics*, 26.

Daganzo, C. (1979), Multinomial Probit, NY: Academic Press.

Fomby, T.B., Hill, R.C., and Johnson, S.R. (1984), *Advanced Econometric Methods*, Springer-Verlag.

Guilkey, D.R., and Schmidt, P. (1979), "Some Small Sample Properties of Estimators and Test Statistics in the Multivariate Logit Model", *Journal of Econometrics*, 10.

Hauser, J.R. (1977), "Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information Theoretic Approach", *Operations Research*, 26.

Hausman, J.A., and Wise, D.A. (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences", *Econometrica*, 46.

Judge, G.G., Griffiths, W.E., Hill, R.C., and Lee, T.C. (1980), *The Theory and Practice of Econometrics*, NY: Wiley.

____ and Lutkepohl, H. (1982), *Introduction to the Theory and Practice of Econometrics*, NY: Wiley.

Madlala, G.S. (1977), *Econometrics*, NY: McGraw-Hill.

____ (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behaviour", *Frontiers in Econometrics*, ed. by P. Zarembka, NY: Academic Press.

____ (1976), "Quantal Choice Analysis, A Survey", *Annals of Economic and Social Measurement*, 5.

Pindyck, R.S. and Rubinfield, D.L. (1976), *Econometrics Models and Economic Forecasts*, NY: McGraw-Hill.

Schmidt, P. and Strauss, R.P. (1975), "The Prediction of Occupation Using Multiple Logit Model", International Economic Review, 16.

Westin, R.B. (1974), "Predictions from Binary Choice Models", Journal of Econometrics, 2.

Zellner, A. and Lee, T.H. (1965), "Joint Estimation of Relationships Involving Discrete Random Variables", Econometrica, 33.