

Limited Dependent Variables – Part I: Lecture XXXIII

Charles B. Moss ¹

¹University of Florida

April 10, 2012

- 1 Readings
- 2 Limited Dependent Variables in Economics
 - Large Purchases
 - General Types of Discrete Choice
 - Other Examples
- 3 Introduction to Estimation
 - Household Production Model
 - Estimation
 - Estimates

Readings

- * Cameron, A. C. and P. K. Trivedi. 2010. *Microeconometrics Using Stata Revised Edition*. Stata Press. (Chapter 14).
- * Greene, W.H. 2012. *Econometric Analysis*. Seventh Edition (Chapter 17 [Sections 17.1 - 17.3 pp. 681 - 714])
- * Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press. (Chapters 1 and 2 pp. 1-56)

Limited Dependent Variables in Economics

- Historically, economics in general and agricultural economics in particular was fairly content with continuous variables:
 - Demand systems estimated the quantity of goods consumed by consumers.
 - For an individual

$$x_{it} = \alpha_{0i} + \alpha_{1i} Y_{it} + \alpha_{2i} p_{1t} + \alpha_{3i} p_{2t} + \epsilon_{it} \quad (1)$$

for individual i at time t .

- In some instances, we aggregate the individual level of demand across individuals in a given market

$$x_t = \sum_{i=1}^N x_{it} = \alpha_0 + \alpha_1 f \left(\sum_{i=1}^N Y_{it} \right) + \alpha_2 p_{1t} + \alpha_3 p_{2t} + \tilde{\epsilon}_t \quad (2)$$

- Similar individual and aggregate models are formulated for supply functions.
- While recognizing that $x_{it} \geq 0$, the boundary condition was not a serious concern.

Large Purchases

- While the continuous (possibly bounded from below) concept "worked" for something like eggs, pounds of beef, etc. it really misses the mark for large purchases like automobiles, refrigerators, etc.
- In the case of large purchases the integer nature of demand becomes more apparent – The individual will purchase $x_{it} = 0, 1, 2, 3, \dots m$ goods in any given year (where m is relatively small).
- As a starting point, consider the question of refrigerator demand (this example is dedicated to John Kenneth Galbraith's Montgomery Martin from *A Tenured Professor*)

$$x_{it} = \alpha_0 + \alpha_1 Y_{it} + \alpha_2 p_{1t} + \alpha_3 p_{2t} + \epsilon_{it} \Rightarrow x_{it} = \{0, 1\} \quad (3)$$

where the individual either purchases a new refrigerator or does not.

Large Purchases – Continued

- Estimating the regression model from Equation 3 is not satisfactory:
 - x_{it} actually estimates the probability that the individual will purchase a refrigerator in a given year.
 - The probability formulation implies heteroscedasticity leading to the standard discriminant analysis.
 - The formulation is not consistent with maximum likelihood – The error structure does not recognize the limited dependent variable approach.
- In general, we reformulate this problem into a different Maximum Likelihood function

$$\left. \begin{array}{l} \text{Prob}[x_{it} = 1] = F(z, \beta) \\ \text{Prob}[x_{it} = 0] = 1 - F(z, \beta) \end{array} \right\} \Rightarrow \prod_{i=1}^N [F(z, \beta)]^{x_{it}} [1 - F(z, \beta)]^{1-x_{it}} \quad (4)$$

Large Purchases – Continued

- Linking Equation 4 with the market demand curve in Equation 2, we can reformulate the market demand for these large-purchase commodities

$$x_t(Y_{it}, p_{1t}, p_{2t}) = \sum_{i=1}^N F(Y_{it}, p_{1t}, p_{2t}, \alpha) \quad (5)$$

General Types of Discrete Choice

- **Binary Choice:** The individual faces a pair of choices and makes the choice between the one that provides the greatest utility (buy a refrigerator or not).
- **Multinomial Choice:** The individual chooses among more than one two choices, making the choice that provides the greatest utility – Choice of where to recreate (Green, Moss, and Spreen).
- **Ordered Choice:** The individual reveals the strength of his preference with respect to a single outcome (the order matters).
- **Event Counts:** The observed outcome is a count of the number of occurrences.

Other Examples

- **Labor Force Participation:** There are various levels and complexities of this problem.
 - In the simplest example, we hypothesize a wage model based on age and education (measure for the time being as a set of k dummy variables)

$$y_{it} = \beta_0 + \beta_1 a_i + \beta_2 a_i^2 + \sum_{j=1}^k \tilde{\beta}_j e_k + \epsilon_{it} \quad (6)$$

First, we could speculate that the individual would be unemployed so that the income distribution truncates at zero.

Other Examples – Continued

• Labor Force Participation

- Here, we anticipate two random variables (y_{it} which is the level of income and z_{it} which is the probability that the individual is working)

$$y_{it}z_{it} = \beta_0 + \beta_1 a_i + \beta_2 a_i^2 + \sum_{j=1}^k \tilde{\beta}_j e_k + \epsilon_{it} \quad (7)$$

$$z_{it} = F(a_i, e_k, \alpha)$$

which is formulated along the lines of a Tobit model (left-truncated random variable).

- Alternatively, we could formulate the same model more along the lines of a selection-bias problem

$$y_{it} = \beta_0 + \beta_1 a_i + \beta_2 a_i^2 + \beta_3 z_{it} + \sum_{j=1}^k \tilde{\beta}_j e_k + \epsilon_{it} \quad (8)$$

$$z_{it} = F(a_i, e_k, \alpha)$$

Other Examples

- **Random Utility Models:** Random utility models are popular in the valuation of environmental amenities or recreation models.
 - In these models, we assume that there are a set of choices $U_i, i = a, b$ of which the individual selects one alternative. If the individual accepts U_a then we assume that $U_a > U_b$.
 - The basic formulation is then

$$\begin{aligned}U_a &= w'\beta_a + z'_a\gamma_a + \epsilon_a \\U_b &= w'\beta_b + z'_b\gamma_b + \epsilon_b\end{aligned}\tag{9}$$

Other Examples – Continued

- **Effect of Policy**

- From my work in Rwanda - The USAID implemented a program to develop a marketing channel for high valued coffee. At the end of the policy period, the effectiveness of the policy was evaluated. A primary question is whether the policy reduced the rate of poverty.
- A similar piece asked whether government spending on "safe sex" was more effective than other programs in reducing the spread of AIDS in Africa.

Introduction to Estimation

- Starting with the logarithm of Equation 4

$$\ln(L) = \sum_{i=1}^N [x_{it} \ln(F[z_{it}, \beta]) + (1 - x_{it}) \ln(1 - F[z_{it}, \beta])] \quad (10)$$

- Two standard selections are the logit and probit functions.
 - The logit function can be written as

$$F[z_{it}, \beta] = \frac{\exp(-z'_{it}\beta)}{1 + \exp(-z'_{it}\beta)} \quad (11)$$

- Where the probit is the integral of the standard normal distribution function.

Household Production Model

- We extend Beckers model as

$$\begin{aligned} & \max_{x, \tilde{L}} U(z) \\ & \text{s.t. } \left\{ z, x, \tilde{L}, K_H \right\} \in T \\ & w'x \leq I + w_L(K_H) [L - \tilde{L}] \end{aligned} \quad (12)$$

where

- z is a vector of consumption good produced from a vector of purchased inputs x and the quantity of labor devoted to home production $\tilde{L}$,
- in this formulation K_H is a vector of quality characteristics associated with household labor (such as the level of education of each member of the household),
- T denotes a production technology of feasible combinations of inputs (including labor), outputs, and labor characteristics Fare (1988),

Household Production Model – Continued

- Continued
 - w is the vector of input prices,
 - $\tilde{I}$ is the level household income from sources other than off-farm income,
 - $w_L(K_H)$ is the wage rate associated with the quality characteristics of household labor, and
 - L is the total quantity of labor available to the farm household.

Household Production Model – Continued

- The income to the household can then expressed as

$$\begin{aligned} I &= \tilde{I} + p'_1 y_1 + p'_2 (y_2 - x_2) - w'_3 x_3 + w_L (K_H) [L - \tilde{L}] \\ G(y_1, y_2, x_3, L^*, K_H,) &= 0 \end{aligned} \quad (13)$$

where

- p_1 and p_2 are vectors of output prices which are conformable with output vectors y_1 and y_2 respectively,
 - y_2 is the vector of outputs that are also consumed by the household,
- w_3 is the vector of input prices used to produce agricultural output,
- L^* is the labor used on farm, and
- $G(.)$ is the production function for farm outputs.

Household Production Model – Continued

- If none of the outputs are directly used in consumption, this simply becomes the profit function

$$\begin{aligned}
 \pi(p, w, L^*, K_H) &= \max_{x,y} y'_1 p_1 + y'_2 p_2 - w'_3 x_3 \\
 \text{s.t. } G(y_1, y_2, x_3, L^*, K_H) &= 0 \\
 L &\leq L^* \\
 \Rightarrow I &= \tilde{I} + \pi(p, w, L^*, K_H).
 \end{aligned} \tag{14}$$

- The integrated household model then becomes

$$\begin{aligned}
 &\max_{z,x,y,\tilde{L},L^*} U(z) \\
 \text{s.t. } &F(z, x_1, x_2, \tilde{L}, K_H) = 0 \\
 &w'_1 x_1 \leq \tilde{I} + p'_1 y_1 + p'_2 (y_2 - x_2) - w'_3 x_3 + w_L(K_H) [L - \tilde{L} - L^*] \\
 &G(y_1, y_2, x_3, L^*, K_H) = 0
 \end{aligned} \tag{15}$$

Household Production Model – Continued

- What do we get out of the complications introduced by the production processes?
 - First, following from the basic results of Becker (1965) the demand for the goods used to produce consumption items are a function of the wage rate, the vector of labor quality attributes, and other factors determining the profitability of agriculture

$$x_i = x_i \left(p_1, p_2, w_1, w_2, w_3, w_L (K_H), K_H, \tilde{I} \right) \quad i \in \{ x_1, x_2 \} . \quad (16)$$

The production decisions are important in two ways. First, the household now must consider the opportunity cost of labor in off-farm employment.

- The quasi-fixed input of labor becomes variable given a sufficiently lucrative off-farm employment opportunity.

Household Production Model – Continued

- Continued
 - This opens the door for considering a variety of factors that could affect this wage rate such as quality attributes of the household labor such as education, age, experience, etc., as well as factors such as distance to an urban area.
 - Second, the price of the second output will be either the market price p_2 or the marginal value x_2 for home consumption.
- We can write a general form output supply function as

$$y_i = y_i \left(p_1, p_2, w_1, w_2, w_3, w_L(K_H, K_L), K_H, \tilde{l} \right) \quad i \in \{y_1, y_2\} \quad (17)$$

where the wage function in Equation 17 has been expanded to consider the effect of location on the wage rate.

Household Production Model – Continued

- At this point, I need to develop three models:
 - (1) The choice of producing coffee $D_1 = 1$ versus not. I think that this choice is a function of the opportunity cost of labor which is a function of (a) the distance to the labor market and (b) the human capital characteristics.
 - (2) The choice of participating in the value-chain $D_2 = 1$ versus not. I think that this is a function of (a) be selected into coffee production and (b) the distance to the coffee washing station.
 - (3) the probability of being in poverty $D_3 = 1$ which we really want to be the probability of being selected into treatment $D_2 = 1$ controlling for other factors.

Estimation

- We consider the basic logit model presented in Maddala 1983 where y_i^* is a dependent variable of the regression

$$y_i^* = x_i\beta + u_i \quad (18)$$

where x_i is a row vector of independent variables, β is a parameter vector to be estimated, and u_i is an error term.

- This unobservable regression variable is then observed through a dichotomous variable y_i where

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

Estimation

- Following the logit formulation, we estimate the dichotomous variable by maximizing

$$L = \prod_{y_i=1} F(x_i\beta) \prod_{y_i=0} (1 - F(x_i\beta)) \quad (20)$$

$$F(x_i\beta) = \frac{\exp(x_i\beta)}{1 + \exp(x_i\beta)}.$$

- The variance matrix for the β coefficients will be estimated based on the inverse of the Hessian of the likelihood function in Equation 20.
- The marginal probability of each variable is then defined by the derivative of the logit function at the mean and the standard errors for these probability are estimated by bootstrapping the marginal probabilities around the maximum likelihood estimates.

Parameter Estimates for the Logit Model

Coefficient	Parameter	Coefficient	Parameter
Constant	0.989927 (0.310810)	Distance to Secondary School	0.000866 (0.001350)
Treatment	-0.285416 (0.184129)	Distance to Hospital	-0.000628 (0.000836)
Household Size	0.099412 (0.043321)	Distance to Health Center	0.000948 (0.002203)
Sex of Head of Household	-0.007931 (0.006454)	Head of Household Reads	-0.334023 (0.240055)
Distance to Food Market	0.000773 (0.002445)	Primary School	-0.379588 (0.237835)
Distance to Common Trans	0.000571 (0.001167)	Professional School	-1.041620 (0.330006)
Distance to Primary School	0.006354 (0.003957)	Secondary School	-0.000003 (0.000001)

Readings – Next Time

- * Cameron, A. C. and P. K. Trivedi. 2010. *Microeconometrics Using Stata Revised Edition*. Stata Press. (Chapter 15).
- * Greene, G., C.B. Moss, and T.H. Spreen. 1997. Demand for Recreational Fishing in Tampa Bay, Florida: A Random Utility Approach. *Marine Resource Economics* 12(4), 293-305.
- * Greene, W.H. 2012. *Econometric Analysis*. Seventh Edition (Chapter 17 [Section 17.5 pp. 738 – 752])
- * Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics* Cambridge University Press. (Chapter 3 pp. 59 - 78)