Econometrica, Vol. 51, No. 2 (March, 1983)

NOTES AND COMMENTS

GENERALIZED ECONOMETRIC MODELS WITH SELECTIVITY

BY LUNG-FEI LEE[1]

1. INTRODUCTION

IN LEE [**12**] AND DUNCAN [**5**], among others, econometric models with both continuous and discrete variables are formulated. These models unify the censored regression models and discrete choice models. These models, formulated with normal distributions, are restricted for computational tractability to binary choice. Many economics problems, such as immigration and occupational choice, do involve multiple choice and censored dependent variables. The multinomial probit models have attractive theoretical properties (see Hausman and Wise [7], but are computationally complicated and almost intractable for polychotomous responses with many categories. The conditional logit models of McFadden [**15**] based on extreme value distributions are apparently much easier to be implemented and are the widely used models for multiple responses. The realism of the normality assumption has been questioned by Olsen [**17**] and Goldberger [**6**] and they have shown that existing estimators for Tobit models are not robust to departures from normality.

In this article, we will suggest an approach to formulating models with given marginal distributions and models with discrete choice and censored dependent variables. Our approach will be useful when the investigators have a priori theoretical reasons on the use of specific marginal distributions. Our generalized models have tractable likelihood functions and can be computationally implemented. Two-stage estimation methods similar to the two-stage methods in Amemiya [**1**], Heckman [**9**], and Lee [**12**] can also be derived for some of our generalized models.

2. BINARY CHOICE SELECTIVITY MODELS

Consider the simple two equations censored regression model:

(2.1) $$y_1 = x\beta + \sigma u, \qquad \sigma > 0,$$

(2.2) $$y^* = z\gamma - \epsilon,$$

where x and z are exogenous variables, $E(u \mid x, z) = 0$, $E(\epsilon \mid x, z) = 0$, and $\operatorname{var}(\epsilon \mid x, z) = 1$. The disturbances u and ϵ conditional on x and z have absolutely continuous distribution functions $G(u)$ and $F(\epsilon)$, which are completely specified.[2] The dependent variable y^* is unobservable but has a dichotomous observable realization I which is related to y^* as follows:

$$I = 1 \quad \text{if and only if} \quad y^* \geq 0,$$

$$I = 0 \quad \text{if and only if} \quad y^* < 0.$$

[1] This article was completed during my visit to the Center for Econometrics and Decision Sciences, University of Florida, Gainesville, Florida. I appreciate having financial support from the National Science Foundation under Grant SES-8006481 to the University of Minnesota. I thank G. S. Maddala and two anonymous referees for their valuable comments.

[2] When u is normally distributed, it will be standardized to have unit variance. For this case, σ is the standard deviation for the nonstandardized distribution.

The dependent variable y_1 conditional on x and z has well-defined marginal distribution but y_1 is not observed unless $y^* \geq 0$. The observed samples y's of y_1 are thus censored and

$$y = x\beta + \sigma u \qquad \text{if and only if} \quad z\gamma \geq \epsilon.$$

For the model in equations (2.1) and (2.2), the distributions u and ϵ are allowed to be correlated. Since only the marginal disturbances of u and ϵ are specified but not the joint bivariate distribution of u and ϵ, the formulation of the complete model is to suggest some interesting proper bivariate distributions which have the specified marginal distributions. Any joint bivariate distribution that will be of interest should allow unrestricted correlation between the disturbances u and ϵ.[3]

Let $\Phi(\cdot)$ be the standard normal distribution function and let $B(\cdot, \cdot; \rho)$ be the bivariate normal distribution $N(0, 0, 1, 1, \rho)$ with zero means, unit variances, and correlation coefficient ρ. With the completely specified marginal distributions $G(u)$ and $F(\epsilon)$ of u and ϵ, respectively, each of them can be transformed into a standard normal random variable $N(0, 1)$. Let

$$\epsilon_* = J_1(\epsilon) = \Phi^{-1}(F(\epsilon)) \tag{2.3}$$

and

$$u_* = J_2(u) \equiv \Phi^{-1}(G(u)). \tag{2.4}$$

Both the transformed random variables u_* and ϵ_* are standard normal variables with zero means and unit variances. A bivariate distribution having the marginal distributions $F(\epsilon)$ and $G(u)$ can be specified as

$$H(\epsilon, u; \rho) = B[J_1(\epsilon), J_2(u); \rho]. \tag{2.5}$$

Thus this bivariate distribution of (ϵ, u) is derived by assuming that the transformed variables u_* and ϵ_* are jointly normally distributed with zero means, unit variances, and correlation coefficient ρ. When $\rho = 0$, it corresponds to statistical independence of u and ϵ. When the marginal distributions of u and ϵ are normally distributed, the above bivariate distribution will be a bivariate normal distribution. Let (y_i, x_i, z_i, I_i), $i = 1, \ldots, N$, be the random samples. The log likelihood function based on this specification is

$$\begin{aligned} \ln L(\beta, \gamma, \rho) = \sum_{i=1}^{N} \Big\{ & I_i \ln g((y_i - x_i\beta)/\sigma) \\ & + I_i \ln \Phi\Big(J_1(z_i\gamma) - \rho J_2((y_i - x_i\beta)/\sigma) \Big/ \sqrt{1-\rho^2}\Big) \\ & - I_i \ln \sigma + (1 - I_i)\ln(1 - F(z_i\gamma)) \Big\} \end{aligned} \tag{2.6}$$

where g is the density function of u.

This approach is of a special interest when the marginal distribution of u is normally distributed $N(0, 1)$ and the marginal distribution of ϵ can be arbitrary. From the model

[3] It is known that for any bivariate distribution $H(u, \epsilon)$ with marginal distributions $G(u)$ and $F(\epsilon)$, $H_{-1}(u, \epsilon) \leq H(u, \epsilon) \leq H_1(u, \epsilon)$ where $H_1(u, \epsilon) = \min\{G(u), F(\epsilon)\}$ and $H_{-1}(u, \epsilon) = \max\{G(u) + F(\epsilon) - 1, 0\}$. The disturbances u and ϵ are perfect positive dependent when they give H_1; they are perfect negative dependent when they give H_{-1}. Only those bivariate distributions which can attain the boundary distributions will have unrestricted range of correlation.

specification in (2.1) and (2.2), $I = 1$ if and only if $z\gamma \geq \epsilon$. Given any absolutely continuous distribution function $F(\epsilon)$, the transformation $J_1 = \Phi_0^{-1}F$ is a strictly increasing function. Therefore, we have $I = 1$ if and only if $J_1(z\gamma) \geq J_1(\epsilon)$. Define $\epsilon_* = J_1(\epsilon)$. It follows that ϵ_* is a standard normal random variable. The censored regression model with given normal marginal distribution $G(u)$ of u, arbitrary marginal distribution $F(\epsilon)$ of ϵ, and the bivariate distribution in (2.5) is statistically equivalent to the model with

$$y_1 = x\beta + \sigma u, \tag{2.7}$$

$$y^{**} = J_1(z\gamma) - \epsilon_*, \tag{2.8}$$

where u and ϵ_* are bivariate normally distributed $N(0, 0, 1, 1, \rho)$. Hence, conditional on $I = 1$, the censored regression equation is

$$y = x\beta - (\sigma\rho)\phi(J_1(z\gamma))/F(z\gamma) + \eta \tag{2.9}$$

where $E(\eta \mid I = 1, x, z) = 0$, the function ϕ is the standard normal density function, and

$$\operatorname{var}(\eta \mid I = 1, x, z) = \sigma^2 - (\sigma\rho)^2[J_1(z\gamma) + \phi(J_1(z\gamma))/F(z\gamma)]\phi(J_1(z\gamma))/F(z\gamma) \tag{2.10}$$

since $F(z\gamma) = \Phi(J_1(z\gamma))$. Thus the two-stage estimation method suggested in Heckman [**9**, **10**] and Lee [**12**] can be extended to our generalized selectivity models. If the choice equation is a probit equation, this two-stage method is exactly the same one as in the literature. When the choice equation is a logit equation, our method becomes a logit-OLS two-stage method. Our two-stage method is thus quite flexible and can be applied to any binary choice models. The correct asymptotic covariances of the two-stage estimates can be constructed as in Lee et al. [**13**] with minor modifications.

The transformation J_1 involves the inverse of the standard normal distribution function Φ. Computationally simple and accurate methods involving the use of approximate function can be found in Appendix II, C, in Bock and Jones [**2**] and Hildebrand [**11**]. Errors of approximation for those methods are less than 3×10^{-4}.

The above approach has demonstrated the use of the bivariate normal distribution. In fact, we can start from any convenient and flexible bivariate distribution and by this method construct a bivariate distribution with given margins F and G. With an appropriately chosen bivariate distribution, this may lead to models which can be estimated by simple consistent methods even the marginal distribution $G(u)$ of u is not normally distributed. The following case provides an example.

Suppose the marginal distribution $G(u)$ of u is known to be (central) Student t distributed with ν degress of freedom. The disturbance ϵ can be transformed to $\epsilon_* = J_1(\epsilon) \equiv G^{-1}(F(\epsilon))$ which will be Student t distributed with ν degrees of freedom. The variables u and ϵ_* will be assumed to have a bivariate Student t distribution with the density function h,

$$h(u, \epsilon_*) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \frac{\nu}{\nu - 2} \times \left[1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(u^2 - 2\rho u\epsilon_* + \epsilon_*^2\right)\right]^{-(1/2)(\nu+2)} \tag{2.11}$$

where $\nu > 2$. This distribution was introduced by K. Pearson [**18**]. The conditional mean and variance of u on ϵ_* are, respectively,

$$E(u \mid \epsilon_*) = \rho\epsilon_* \tag{2.12}$$

and

$$\text{(2.13)} \qquad \operatorname{var}(u \mid \epsilon_*) = (1-\rho^2)(\nu - 2 + \epsilon_*^2)/(\nu - 1)$$

(see Mardia [**14**, p. 92]). As shown in Raiffa and Schlaifer [**19**], the first moment of the truncated (central) Student t distribution with ν degrees of freedom is

$$\text{(2.14)} \qquad E(\epsilon_* \mid \epsilon_* \leq w) = -\frac{\nu + w^2}{\nu - 1} \frac{g(w)}{G(w)}$$

where g is the (central) Student t density function with ν degrees of freedom. It follows that, conditional on $I = 1$, the censored regression equation for the observed sample y is

$$\text{(2.15)} \qquad y = x\beta - \rho\sigma\left(\frac{\nu + J_1^2(z\gamma)}{\nu - 1}\right)\frac{g(J_1(z\gamma))}{G(J_1(z\gamma))} + \eta$$

where $E(\eta \mid I = 1, x, z) = 0$. This equation can be estimated by the two stage method.

3. POLYCHOTOMOUS CHOICE SELECTIVITY MODELS

The approaches introduced in Section 2 provide frameworks for modeling polychotomous choice problems with mixed continuous and discrete dependent variables.[4] Consider the following polychotomous choice model with M categories and M regression equations:

$$\text{(3.1)} \qquad \begin{aligned} y_s &= x_s\beta_s + \sigma_s u_s, \\ y_s^* &= z_s\gamma_s + \eta_s, \qquad (s = 1, \ldots, M), \end{aligned}$$

where all the variables x_s, z_s are exogenous, $E(u_s \mid x_1, \ldots, x_M, z_1, \ldots, z_M) = 0$ and $E(\eta_s \mid x_1, \ldots, x_M, z_1, \ldots, z_M) = 0$. All the distributions u_s are assumed to have completely specified absolutely continuous marginal distributions and the joint distribution of $(\eta_1, \ldots, \eta_M)$ has also been specified. The dependent variable or outcome y_s is observed if and only if the category s is being chosen. Category s is chosen if and only if

$$y_s^* > \max_{\substack{j=1,\ldots,M \\ j \neq s}} y_j^* .$$

Let I be a polychotomous variable with values 1 to M and denote $I = s$ if category s is chosen. Equivalently,

$$\text{(3.2)} \qquad I = s \quad \text{if and only if} \quad z_s\gamma_s > \epsilon_s$$

where

$$\text{(3.3)} \qquad \epsilon_s \equiv \max_{\substack{j=1,\ldots,M \\ j \neq s}} y_j^* - \eta_s .$$

[4] Approaches that are different from ours have been introduced in Dubin and McFadden [**4**] and Hay [**8**].

For each pair (u_s, ϵ_s), suppose the specified marginal distribution of u_s is $G_s(u)$ and the implied marginal distribution of ϵ_s is $F_s(\epsilon)$. A bivariate distribution of (u_s, ϵ_s) can be specified by the translation method, for example, the normality transformation in (2.5). Let $g_s(\cdot)$ be the density function of $G_s(\cdot)$. Define dummy variables D_s, $s = 1, \ldots, M$, such that

$$D_s = 1 \qquad \text{if and only if} \quad I = s.$$

The log likelihood function for this polychotomous choice model with random samples of size N is

$$\text{(3.4)} \qquad \ln L = \sum_{i=1}^{N} \sum_{s=1}^{M} \Big\{ D_{si} \ln g_s((y_{si} - x_{si}\beta_s)/\sigma_s) - D_{si} \ln \sigma_s + D_{si} \ln \Phi\Big((J_{1s}(z_{si}\gamma_s) - \rho_s J_{2s}(y_{si} - x_{si}\beta_s))/\big(1 - \rho_s^2\big)^{1/2} \Big) \Big\}$$

where $J_{1s} = \Phi_0^{-1} F_s$ and $J_{2s} = \Phi_0^{-1} G_s$.

In the econometrics literature, one of the well-known and most widely used polychotomous choice models is the conditional multinomial logit model of McFadden [**15**]. In this model, y_s^* are stochastic utility functions and $\gamma_1 = \gamma_2 = \cdots = \gamma_M$ in (3.1), i.e.,

$$\text{(3.5)} \qquad y_s^* = z_s \gamma + \eta_s \qquad (s = 1, \ldots, M)$$

where $z_M = 0$ is used as a normalization rule. The stochastic parts of the utility functions, η_s, $s = 1, \ldots, M$, are assumed to be independent and identically Gumbel distributed. As shown in Domencich and McFadden [**3**],

$$\text{(3.6)} \qquad \begin{aligned} F_s(\epsilon) &\equiv \text{Prob}[\epsilon_s < \epsilon] \\ &= \text{Prob}\left[\left(\max_{\substack{j=1,\ldots,M \\ j \neq s}} y_j^* - \eta_s \right) < \epsilon \right] \\ &= \frac{\exp(\epsilon)}{\exp(\epsilon) + \sum_{j=1, j\neq s}^{m} \exp(z_j \gamma)}. \end{aligned}$$

When the marginal distributions of u_s are normal distributed $N(0, 1)$, two stage method can be used to estimate the equations

$$\text{(3.7)} \qquad y_s = x_s \beta_s - \sigma_s \rho_s \phi(J_{1s}(z_s\gamma))/F_s(z_s\gamma) + \eta_s \qquad (s = 1, \ldots, M).$$

Thus, if the polychotomous choice model is multinomial logit model and the marginal distributions of the potential outcome functions y_s are normal, we have a multinomial logit-OLS two-stage estimation method. The correct asymptotic covariance matrix of this multinomial logit-OLS two-stage estimates has similar expression as in the binary choice case with appropriate modifications.[5]

University of Minnesota

Manuscript received January, 1981; revision received February, 1982.

[5] It should be pointed out that our approach is ready to be generalized to more complicated polychotomous choice models, such as those models in McFadden [**16**].

REFERENCES

[1] AMEMIYA, T.: "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal," *Econometrica*, 42(1974), 999–1012.

[2] BOCK, R. D., AND L. V. JONES: *The Measurement and Prediction of Judgement and Choice*. San Francisco: Holden Day, 1968.

[3] DOMENCICH, T. A., AND D. MCFADDEN: *Urban Travel Demand*. Amsterdam: North-Holland Pub. Co., 1975.

[4] DUBIN, J. A., AND D. MCFADDEN: "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," Manuscript, Department of Economics, Massachusetts Institute of Technology, 1980.

[5] DUNCAN, G. M.: "Specification and a Simple Estimator for the Mixed, Continuous/Discrete Dependent Variable in Classical Production Theory," *Econometrica*, 48(1980), 839–852.

[6] GOLDBERGER, A. S.: "Abnormal Selection Bias," Discussion Paper No. 8006, Social Systems Research Institute, University of Wisconsin, Madison, 1980.

[7] HAUSMAN, J., AND D. WISE: "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46(1978), 403–426.

[8] HAY, J.: "Selectivity Bias in a Simultaneous Logit-OLS Model," Manuscript, University of Southern California, 1980.

[9] HECKMAN, J. J.: "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5(1976), 475–492.

[10] ———: "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1979), 153–161.

[11] HILDEBRAND, F. B.: *Introduction to Numerical Analysis*. New York: McGraw Hill, 1956.

[12] LEE, L. F.: "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables," *Econometrica*, 47(1979), 977–996.

[13] LEE, L. F., G. S. MADDALA, AND R. P. TROST: "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity," *Econometrica*, 48(1980), 491–503.

[14] MARDIA, K. V.: *Families of Bivariate Distributions*. Darien, Conn.: Hafner Publishing Co., 1970.

[15] MCFADDEN, D.: "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press, 1973.

[16] ———: "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data*, ed. by C. Manski and D. McFadden. Cambridge: The MIT Press, 1980.

[17] OLSEN, R. J.: "Tests for the Presence of Selectivity Bias and Their Relation to Specifications of Functional Form and Error Distribution," Institution for Social and Policy Studies Working Paper 812, Yale University, 1979.

[18] PEARSON, K.: "On Non-Skew Frequency Surfaces," *Biometrika*, 15(1923), 231–244.

[19] RAIFFA, H., AND R. SCHLAIFER: *Applied Statistical Decision Theory*. Boston: Harvard U. Press, 1961.