GENERALIZED ECONOMETRIC MODELS

WITH SELECTIVITY

by

Lung-Fei Lee

Discussion Paper No. 81 - 140, February 1981

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

November 1980

GENERALIZED ECONOMETRIC MODELS WITH SELECTIVITY

By

Lung-Fei Lee(*)

1. Introduction

During the recent years, there is substantial interest in the econometric models with qualitative and censored dependent variables. The important contributions on these topics by Amemiya [1973], McFadden [1973] and Heckman [1974] among others stimulate the recent development and empirical applications. The specifications of the econometric models with censored dependent variables are based on the normal distributions, see, for example, Amemiya [1973,1974], Heckman [1976,1979] and Lee et al. [1980]. For the development of qualitative response models, normal distributions play an important but limited role. The qualitative response models have attractive theoretical properties, see Hausman and Wise [1978], but are computationally complicated and almost intractible for polychotomous responses with many categories. For the complexity and the limitations in the empirical implementation of the multinomial probit models, one can consult the reports in Lerman and Manski [1980] and Daganzo [1979]. The conditional logit models of McFadden [1973] based on extreme value distributions are apparently much easier to be implemented and are the widely used models for multiple responses.

The normality assumption in the censored regression models may not be appropriate theoretically for some cases. For example, when the dependent variables take only positive values, Poirier [1978] recommends the Box-Cox transformation to transform the dependent variables and then specify normality for the transformed variable. Olsen [1979] and Goldberger [1980] investigate the misspecification of the normal distribution assumption on various estimation methods. Models with nonnormal marginal distributions are of interest and need to be developed.

In Lee [1979] and Duncan [1980], among others, econometric models with both continuous and discrete variables are formulated. These models unified the censored regression models and discrete choice models. These models formulated with normal distributions are restricted for computational tractibility to binary choice. Many economics problems, such as immigration and occupational choice, do involve multiple choice and censored dependent variables. Computationally tractible and theoretical sound economic models need to be developed for the multiple choice problems with mixed discrete and continuous dependent variables. Some recent attempts to model these cases can be found in Dubin and McFadden [1980].

In this article, we will suggest some approaches to formulate models with given marginal distributions and models with multiple discrete choice and censored dependent variables. Our generalized models have tractible likelihood functions and can be computationally implemented. Two-stage estimation methods similar to the two-stage methods in Amemiya [1974], Heckman [1976], and Lee [1976] can also be derived for some of our generalized models. This article is organized as follows. In section 2, we suggest some methods to formulate selectivity models with specified marginal distributions. In section 3, we consider the two-stage estimation procedure and provide the proper asymptotic covariance matrix for the estimators. In section 4, multiple-choice models with mixed discrete and censored continuous dependent variables are specified and analyzed. Finally, we make our conclusions.

2. Abnormal Selectivity Models

Consider the simple two equations censored regression model

$$y_1 = x\beta + u \tag{2.1}$$

$$y^* = z\gamma - \varepsilon \tag{2.2}$$

where x and z are exogenous variables, $E(u|x,z) = 0$, $E(\varepsilon|x,z) = 0$ and $\text{var}\,(\varepsilon|x,z) = 1$. The disturbances u and ε conditional on x and z have absolutely continuous distribution functions $G(u)$ and $F(\varepsilon)$, which are specified except for some unknown finite dimensional vectors of parameters which are suppressed in our notations for simplicity. The dependent variable y^* is unobservable but has a dichotomous observable realization I which is related to y^* as follows:

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

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

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

$$y = x\beta + u \quad \text{if and only if} \quad z\gamma \geqslant \varepsilon.$$

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 distributions 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. For any bivariate distribution $H(u,\varepsilon)$ with marginal distributions $G(u)$ and $F(\varepsilon)$, it is known that, as given by Fréchet,

$$H_{-1}(u,\varepsilon) \leqslant H(u,\varepsilon) \leqslant H_1(u,\varepsilon)$$

where $H_1(u,\varepsilon) = \min\{G(u),F(\varepsilon)\}$ and $H_{-1}(u,\varepsilon) = \max\{G(u) + F(\varepsilon)-1,0\}$. The boundary distributions $H_1(u,\varepsilon)$ and $H_{-1}(u,\varepsilon)$ are two distribution functions with the marginal distribution functions $F(\varepsilon)$ and $G(u)$. u and ε are perfect positive dependent when they give H_1; perfect negative dependence when they give H_{-1}. The proof of the above inequalities are straightforward and is omitted here. Interested readers can find the proofs in Mardia [1970], page 31. It is true that there may exist many families of bivariate distributions which have the specified marginal distributions and contain the boundary distributions as special cases. However, the following bivariate distribution of (u,ε) is of particular interest.

Let $\Phi(\cdot)$ be the standard normal distribution function and $B(\cdot,\cdot;\rho)$ 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 $F(\varepsilon)$ of u and ε, respectively, each of them can be transformed into a standard normal random variable $N(0,1)$. Let

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

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 $F(\varepsilon)$ and $G(u)$ can be specified as

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

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. Let $f(\varepsilon)$ and $g(u)$ be the corresponding density functions of ε and u respectively. The joint density function corresponding to the distrubtion H is

$$h(\varepsilon,u;\rho) = (1 - \rho^2)^{-1/2} f(\varepsilon)g(u)\ \exp\{-\rho(2(1 - \rho^2))^{-1}[\rho(J_1^2(\varepsilon) + J_2^2(u)) - 2J_1(\varepsilon)J_2(u)]\} \tag{2.6}$$

This class contains the Fréchet's boundary distributions. Specifically, $H(\varepsilon,u;1) = H_1(\varepsilon,u)$ and $H(\varepsilon,u;-1) = H_{-1}(\varepsilon,u)$. 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.

With this specification, one can easily derive the likelihood function for the censored regression model in (2.1) and (2.2). Let

$$\psi(y,1|x,z) = \int_{-\infty}^{z\gamma} h(\varepsilon,y-x\beta)d\varepsilon \tag{2.7}$$

be the mixed continuous-discrete density function of y and I = 1. It follows

$$\psi(y,1|x,z) = \frac{\partial}{\partial u} H(\varepsilon,u;\rho)\Big|_{\substack{u=y-x\beta \\ \varepsilon=z\gamma}}$$

$$= \frac{\partial B[J_1(\varepsilon),\ J_2(u);\rho]}{\partial J_2(u)} \cdot \frac{g(u)}{\phi(J_2(u))}\Big|_{\substack{u=y-x\beta \\ \varepsilon=z\gamma}}$$

where $\phi(\cdot)$ is the standard normal density function. Since

$$\frac{\partial B(t,s;\rho)}{\partial s} \frac{1}{\phi(s)} = \Phi\left((t - \rho s)\Big/\sqrt{1 - \rho^2}\right)$$

we have

$$\psi(y,1|x,z) = \Phi\left[(J_1(z\gamma) - \rho J_2\ (y - x\beta))\Big/\sqrt{1 - \rho^2}\right] \cdot g(y - x\beta) \tag{2.8}$$

Let (y_i,x_i,z_i,I_i), $i = 1, \ldots, N$ be the given random samples. The log likelihood function based on this specification is

$$\begin{aligned}\ln L(\beta,\gamma,\theta_1,\theta_2,\rho) \\ = \sum_{i=1}^{N} \Big\{ I_i \ln g(y_i - x_i\beta) + I_i \ln \Phi\Big[(J_1(z_i\gamma) \\ - \rho J_2(y_i - x_i\beta))\Big/\sqrt{1 - \rho^2}\ \Big] \\ + (1 - I_i) \ln (1 - F(z_i\gamma))\Big\}\end{aligned} \tag{2.9}$$

where θ_1 and θ_2 are the unknown parameters in $F(\varepsilon)$ and $G(u)$, respectively. Maximum likelihood method can be applied to this likelihood function.

Another class of bivariate distribution with specified marginal distributions which may be of interest is known as the contingency type distributions. Given marginal distributions $F(\varepsilon)$ and $G(u)$, a contingency-type distribution with margins F and G is

$$H_c(\varepsilon,u;\lambda) = \left\{S - [S^2 - 4\lambda(\lambda - 1)F(\varepsilon)G(u)]^{1/2}\right\}\Big/(2(\lambda - 1)),\ \lambda > 0 \tag{2.10}$$

where $S = 1 + (F(\varepsilon) + G(u))(\lambda - 1)$. As λ tends to 1, H_c converges to $F(\varepsilon)F(u)$; λ tends to 0, H_c tends to max $(0, F(\varepsilon) + G(u) - 1)$ and λ tends to ∞, H_c tends to min $(F(\varepsilon), G(u))$. Thus this class also contains the Fréchet boundary distributions and the distribution corresponding to independent random variables. For the details, one can see Mardia [1970]. The joint density function of (ε,u) corresponding to H_c is

$$h_c(\varepsilon,u;\lambda) = \frac{\partial^2 H_c}{\partial F \partial G} f(\varepsilon)g(u)$$

$$= \lambda f(\varepsilon)g(u)[(\lambda - 1)(F(\varepsilon) + G(u) - 2F(\varepsilon)G(u))$$

$$+ 1]/[S^2 - 4\lambda(\lambda - 1)F(\varepsilon)G(u)]^{3/2} \qquad (2.11)$$

Let $K_c(\varepsilon|u)$ be the conditional distribution function of ε given u. As

$$K_c(\varepsilon|u) = \int_{-\infty}^{\varepsilon} h_c(t,u;\lambda)dt/g(u)$$

$$= \frac{\partial H_c(\varepsilon,u;\lambda)}{\partial u} /g(u)$$

$$= \frac{\partial H_c(\varepsilon,u;\lambda)}{\partial G}$$

it follows that

$$K_c(\varepsilon|u) = \frac{1}{2} + \frac{1}{2} [(\lambda + 1)F(\varepsilon) - (\lambda - 1)G(u) - 1]/[S^2$$

$$- 4\lambda(\lambda - 1)F(\varepsilon)G(u)]^{1/2} \qquad (2.12)$$

Under this bivariate distribution (2.10), the mixed probability density function ψ_c of y and I = 1 for the censored regression model is

$$\psi_c(y,1|x,z) = K_c(z\gamma|y - x\beta)\ g(y - x\beta) \tag{2.13}$$

It follows that the corresponding log likelihood function is

$$\ln L_c(\beta,\gamma,\theta_1,\theta_2,\lambda)$$

$$= \sum_{i=1}^{N}\Big\{I_i \ln g(y_i - x_i\beta) + I_i \ln K_c(z_i\gamma|y_i - x_i\beta) + (1 - I_i) \ln (1 - F(z_i\gamma))\Big\} \tag{2.14}$$

From the above derivations, one can see that both approaches give tractible analytical likelihood functions and are attractive approaches for the modeling of selectivity models. As will be shown in in the subsequent sections, simple consistent estimation methods can be derived for some important models based on the first approach. Furthermore, when both the marginal distributions are normal, the bivariate distribution derived from the first approach is bivariate normal but the bivariate distribution from the second approach is not. Based on these reasons, the first approach may be preferred over the second approach.

3. Two-Stage Estimation Method

Whether simple consistent estimation methods rather than the maximum likelihood method can be derived are apparently depended on the specific marginal distributions and the bivariate distribution. In this section, we will consider the possibility of generalizing the two-stage estimation methods in Amemiya [1974], Heckman [1976], and Lee [1976] to some of our models.

Consider the models where the marginal distribution of u is normally distributed $N(0,\sigma^2)$ and the marginal distribution of ε can be arbitrary. The bivariate distribution of (u,ε) with the given margins is specified by the translation method as in (2.5). As $g(u) = \phi(J_2(u))dJ_2(u)/du$ and $g(u)$ is a normal density function of $N(0,\sigma^2)$, $J_2(u) = u/\sigma$ and $g(u) = \phi(u/\sigma)/\sigma$. The corresponding mixed density of y, I = 1 in (2.8) becomes

$$\psi(y,1|x,z) = \Phi\left((J_1(z\gamma) - \rho(y - x\beta)/\sigma) \Big/ \sqrt{1-\rho^2}\right) \cdot \phi(u/\sigma)/\sigma \tag{3.1}$$

It follows from the above equation that, as shown in Lee [1976],

$$E(uI|x,z) = -(\sigma\rho)\phi(J_1(z\gamma)) \tag{3.2}$$

and

$$E(u^2I|x,z) = \sigma^2\Phi(J_1(z\gamma)) - (\sigma\rho)^2 J_1(z\gamma) \cdot \phi(J_1(z\gamma)) \tag{3.3}$$

Hence, conditional on I = 1, the censored regression equation can be rewritten as

$$y = x\beta - (\sigma\rho)\phi(J_1(z\gamma))/F(z\gamma) + \eta \tag{3.4}$$

where $E(\eta|I=1,x,z) = 0$ and

$$\text{var}\ (\eta|I=1,x,z) = \sigma^2 - (\sigma\rho)^2[J_1(z\gamma) + \phi(J_1(z\gamma))/F(z\gamma)]\phi(J_1(z\gamma))/F(z\gamma) \tag{3.5}$$

since $F(z\gamma) = \Phi(J_1(z\gamma))$. The above expressions are the familiar expressions in the cited literature with the term $z\gamma$ replaced by $J_1(z\gamma)$ throughout the truncated mean and variance. Thus the two-stage estimation method suggested in Heckman [1976] and Lee [1976] can be extended to our generalized abnormal selectivity models. In the first step, we estimate the parameter γ from the log likelihood function for any specified probability model where, without loss of generality, the only unknown parameter vector is assumed to be γ,

$$\ell n\ L_1(\gamma) = \sum_{i=1}^{N} \left\{ I_i\ \ell n\ F(z_i\gamma) + (1 - I_i)\ \ell n\ (1 - F(z_i\gamma)) \right\}$$

Let $\hat{\gamma}$ denote the derived maximum likelihood estimate of γ. The second stage estimation is to estimate the equation

$$y_i = x_i\beta - \sigma\rho\ \phi(J_1(z_i\hat{\gamma}))/F(z_i\hat{\gamma}) + \eta_i \tag{3.6}$$

with the observed samples on y_1 by the ordinary least square. If the choice equation is a probit equation, this two-stage-method is exactly the same one 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.

As pointed out in Heckman [1979], Lee et al. [1980], the OLS step does not provide the correct asymptotic covariance matrix since the disturbances η_i are heteroscedastic and autocorrelated. It is also true for our generalized models, the conventional OLS variance matrix with heteroscedastic errors corrected will underestimate the correct asymptotic covariances of the two-stage estimates. The detail arguments follow almost exactly those in Lee et al. [1980] with minor modifications.

For the sake of completeness, we write down explicitly the correct asymptotic variance matrix for our generalized models. For a given sample with size N, suppose there are N_1 ($<N$) observed non-censored subsamples of y_1. Without loss of generality, the observations are rearranged such that the first N_1 observations are non-censored. Let

$$\tilde{X}_1' = \begin{bmatrix} x_1' & x_2' & \cdots & x_{N_1}' \\ -\phi(J_1(z_1\gamma))/F(z_1\gamma) & -\phi(J_1(z_2\gamma))/F(z_2\gamma) & & -\phi(J_1(z_{N_1}\gamma))/F(z_{N_1}\gamma) \end{bmatrix}$$

and $D_1' = [d_1', \ldots, d_{N_1}']$ where $d_i = -\sigma\rho[J_1(z_i\gamma) + \phi(J_1(z_i\gamma))/F(z_i\gamma)]$ $(f(z_i\gamma)/F(z_i\gamma))z_i$. Furthermore, let Λ be an NxN diagonal matrix;

$$\Lambda = \text{Diag } [f^2(z_i\gamma)/(F(z_i\gamma)(1 - F(z_i\gamma)))]$$

and V_1 be an $N_1 x N_1$ diagonal matrix constructed from the first N_1 observations,

$$V_1 = \text{Diag}\,[\sigma^2 - (\sigma\rho)^2(J_1(z_i\gamma)$$

$$+ \phi(J_1(z_i\gamma))/F(z_i\gamma))\phi(J_1(z_i\gamma))/F(z_i\gamma)]$$

The correct asymptotic covariance of our two-stage estimates is

$$\text{var}\begin{pmatrix}\hat{\beta}\\ \sigma\rho\end{pmatrix} = (\tilde{X}_1'\tilde{X}_1)^{-1}\ \tilde{X}_1'(V_1 + D_1(Z'\Lambda Z)^{-1}\ D_1')\tilde{X}_1(\tilde{X}_1'\tilde{X}_1)^{-1} \qquad (3.7)$$

where $Z' = [z_1', z_2', \ldots, z_N']$. This expression becomes the same asymptotic covariance of the probit-OLS two-stage estimation when the choice equation is a probit equation. Slight differences occur as there is a nonlinear transformation $J_1(z\gamma)$ involved in the generalized model.

Our approach based on the translation method does generalize the traditional selectivity model. The similarity of the two-stage estimation methods can be easily understood from the following viewpoint when the choice is binary. From the model specification in (2.1) and (2.2), $I = 1$ if and only if $Z\gamma \geqslant \varepsilon$. Given any absolutely continuous distribution function $F(\varepsilon)$, the transformation $J_1 = \Phi_0^{-1}F$ is a strictly increasing function. Therefore, we have $I = 1$ if and only if $J_1(z\gamma) \geqslant J_1(\varepsilon)$. Define $\varepsilon^* = J_1(\varepsilon)$. 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 $F(\varepsilon)$ of ε and the bivariate distribution in (2.5) is statistically equivalent to the model with

$$y_1 = x\beta + u \tag{3.8}$$

$$y^{**} = J_1(z\gamma) - \varepsilon^* \tag{3.9}$$

where u and ε^* are bivariate normally distributed $N(0,0,\sigma^2,1,\rho)$. Thus, the results derived can be regarded as standard results of the normal selectivity model in (3.8) and (3.9).

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 [1968] and Hildebrand [1956]. Errors of approximation for those methods are less than 3×10^{-4}.

4. Multiple-Choice Selectivity Models

The approaches introduced in section 2 provide frameworks for modeling multiple-choice problems with mixed continuous and discrete dependent variables. Consider the following multiple-choice model with M categories and M regression equations:

$$\begin{aligned} y_s &= x_s\beta_s + u_s \\ & \qquad\qquad\qquad s = 1, \ldots, M \\ y_s^* &= z_s\gamma_s + \eta_s \end{aligned} \tag{4.1}$$

where all the variables x_s, z_s are exogenous, $E(u_s | x_1, \ldots, x_M, \ldots, z_M) = 0$ and $E(\eta_s | x_1, \ldots, x_M, z_1, \ldots z_M) = 0$. All the distributions u_s are assumed to have 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,

$$I = s \quad \text{if and only if} \quad z_s\gamma_s > \varepsilon_s \tag{4.2}$$

where

$$\varepsilon_s \equiv \max_{\substack{j=1,\ldots,M \\ j\neq s}} y_j^* - \eta_s \tag{4.3}$$

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 $F_s(\varepsilon)$. Consider the bivariate distribution of (u_s, ε_s) specified by the translation method as in (2.5),

$$H_s(\varepsilon, u; \rho_s) = B[J_{1s}(\varepsilon), J_{2s}(u); \rho_s] \tag{4.4}$$

where $J_{1s} = \Phi_0^{-1}F_s$ and $J_{2s} = \Phi_0^{-1}G_s$. With these specified bivariate distributions for all s = 1, . . . , M, the likelihood function for this multiple-choice model can be easilty derived. The mixed density function of y_s and I = s is

$$\psi_s(y_s, I=s \mid x_i, \ldots, x_M, z_1, \ldots, z_M)$$

$$= \Phi((J_{2s}(z_s\gamma_s) - \rho_s J_{1s}(y_s - x_s\beta_s))/(1 - \rho_s^2)^{1/2}) \cdot g_s(y_s - x_s\beta_s)$$

where $g_s(\cdot)$ is the density function of $G_s(\cdot)$. Define dummy variables D_s, s=1, . . . , M such that

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

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

$$\ell n\, L = \sum_{i=1}^{N} \sum_{s=1}^{M} \Big\{ D_{si}\ \ell n\ g_s(y_{si} - x_{si}\beta_s) + D_{si}\ \ell n\ \Phi\Big[(J_{2s}(z_{si}\gamma_s) - \rho_s J_{1s}(y_{si} - x_{si}\beta_s))/(1-\rho_s^2)^{1/2}\Big]\Big\} \tag{4.5}$$

The likelihood function depends on the density functions g_s, s=1, . . . , M and the transformations J_{1s}, J_{2s}, s=1, . . . , M. Whether this likelihood function is computationally simple or not, depends on the density functions and the transformations.

In the econometrics literature, one of the well-known and most widely used multiple-choice model is the conditional multinomial logit model of McFadden [1973]. In this model, y_s^* are stochastic utility functions and $\gamma_1 = \gamma_2 = \ldots = \gamma_M$ in (4.1), i.e.,

$$y_s^* = z_s\gamma + \eta_s \qquad s=1, \ldots, M \tag{4.6}$$

where $z_M \equiv 0$ is used as a normalization rule. The stochastic parts of the utility functions, η_s, s=1, . . . , M are assumed to be independent and identically Gumbel distributed, i.e., the distribution function of η_s is

$$Q(\eta_s) = \exp(-\exp(-\eta_s))$$

As shown in Domencich and McFadden [1975],

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

is also Gumbel distributed with parameter

$$\left(-\ln \sum_{\substack{j=1 \\ j \neq s}}^{M} \exp(z_j \gamma) \right)$$

and

$$\begin{aligned} F_s(\varepsilon) &\equiv \text{Prob}\,[\varepsilon_s < \varepsilon] \\ &= \text{Prob}\left[\left(\max_{\substack{j=1,\ldots,M \\ j \neq s}} y_j^* - \eta_s \right) < \varepsilon \right] \\ &= \frac{\exp(\varepsilon)}{\exp(\varepsilon) + \sum_{\substack{j=1 \\ j \neq s}}^{M} \exp(z_j \gamma)} \end{aligned} \tag{4.7}$$

Hence the distribution function of ε_s in (4.3) is the function F_s in (4.7) and $J_{1s}(\varepsilon) = \Phi^{-1}(F_s(\varepsilon))$. For this multiple-choice model, the likelihood function in (4.5) will be computational tractible as long as the marginal density functions g_s are not extremely complicated. Maximum likelihood method can then be applied.

When the marginal distributions of u_s are normal distributed $N(0,\sigma_s^2)$, two-stage method can be used to estimate the equations

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

$$(4.8)$$

Thus, if the multiple-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 conditional multinomial logit model will be estimated by the method of maximul likelihood. The likelihood function for the multinomial logit model is

$$\ell n\ L_1 = \sum_{i=1}^{N} \sum_{s=1}^{M} D_{si}(z_{si}\gamma - \ell n \sum_{j=1}^{M} \exp\ (z_{ji}\gamma)) \qquad (4.9)$$

Let $\hat{\gamma}$ be the multinomial logit maximum likelihood estimate of γ. The second stage estimation is to estimate the following equations for the observed noncensored samples by OLS,

$$y_{si} = x_{si}\beta_s - \sigma_s\rho_s\phi(J_{1s}(z_{si}\hat{\gamma}))/F_s(z_{si}\hat{\gamma}) + \eta_{si}$$

$$i=1,\dots,N_s \qquad (4.10)$$

The correct asymptotic covariance matrix of this multinomial logit-OLS two-state estimates has similar expression as in (3.7) with appropriate modifications. Let

$$\bar{z}_i = \sum_{s=1}^{M} F_s(z_i\gamma)z_{si}$$

As shown in McFadden [1973], the asymptotic covariance matrix of $\hat{\gamma}$ is

$$\operatorname{var}(\hat{\gamma}) = \left[\sum_{i=1}^{N} \sum_{s=1}^{M} (z_{si} - \bar{z}_i)' F_s(z_{si}\gamma)(z_{si} - \bar{z}_i)\right]^{-1}$$

Corresponding to the observations in the s^{th} regime, let $\tilde{X}_s$, D_s and V_s be the matrices constructed as in the equation (3.7). The correct asymptotic covariance matrix of the two-stage estimates $\hat{\beta}_s$, $\widehat{\sigma_s\rho_s}$ of β_s and $\sigma_s\rho_s$ is

$$\operatorname{var}\begin{pmatrix} \hat{\beta}_s \\ \widehat{\sigma_s\rho_s} \end{pmatrix} = (\tilde{X}_s'\tilde{X}_s)^{-1} \tilde{X}_s' (V_s + D_s (\operatorname{var}(\hat{\gamma}))^{-1} D_s') \tilde{X}_s(\tilde{X}_s'\tilde{X}_s)^{-1} \tag{4.11}$$

The translation approach can be easily extended to the cases where there are multiple equations or outcome functions in each regime with the joint multinomial distribution. Suppose there are L_s equations in regime s,

$$y_{s1} = x_{si}\beta_{si} + u_{si}$$

$$\vdots$$

$$y_{sL_s} = x_{sL_s}\beta_{sL_s} + u_{sL_s}$$

$$y_s^* = z_s\gamma_s + \eta_s \tag{4.12}$$

Suppose the joint distribution of $(u_{s1}, \ldots, u_{sL_s})$ is $N(0, \Sigma_s)$. A joint distribution of $(u_{s1}, \ldots, u_{sL_s}, \varepsilon_s)$ where ε_s is defined in (4.3) can be constructed as

$$H_s(u_1, \ldots, u_{L_s}, \varepsilon) = B\left[\frac{u_1}{\sigma_{s1}}, \frac{u_2}{\sigma_{s2}}, \ldots, \frac{u_{L_s}}{\sigma_{sL_s}}, J_{1s}(\varepsilon); \rho_{ijs}, \rho_{i\varepsilon}, i,j=1, \ldots, M\right] \quad (4.13)$$

where $\sigma^2_{s\ell}$ is the variance of $u_{s\ell}$, ρ_{ijs} is the correlation coefficient of u_{si} and u_{sj} and B is the standardized multivariate normal distribution function. It follows that the bivariate distribution of (u_ℓ, s) is $H_s(u_\ell, \varepsilon) = B[(u_\ell/\sigma_\ell), J_{1s}(\varepsilon); \rho]$. Furthermore, it implies that

$$H_s(u_1, \ldots, u_{L_s}, \varepsilon) = B\left[\frac{u_1}{\sigma_{s1}}, \frac{u_2}{\sigma_{s2}}, \ldots, \frac{u_{L_s}}{\sigma_{sL_s}}; \rho_{ijs}, i,j=1, \ldots, M\right]$$

and

$$H_s(\infty, \ldots, \infty, \varepsilon) = F_s(\varepsilon)$$

are the specified marginal distributions. Under this specification, the likelihood function can be derived and the two-stage method can be applied to estimate each equation in each regime as described in the previous paragraph.

6. Conclusions

In this article, we have introduced some generalized censored regression models where the disturbances are not necessarily normal. We have suggested some approaches to construct computationally and theoretically tractible models with discrete choice and selectivity. Likelihood functions are derived for the models. Simple consistent estimation procedures are also derived for some of the models. These consistent methods generalize the familiar two-stage methods in the limited dependent variables literature with minor modifications and are applicable to any binary or multiple choice models. Models with both continuous and finite discrete variables are also introduced. These models generalize the two regimes switching regession models with censored dependent variables to models with any finite number of regimes. If the multiple choice equations are specified as the McFadden's conditional multinominal logit model, the generalized switching regression model can be estimated by multinomial logit-OLS two-stage method. Correct asymptotic covariance matrix is derived.

University of Minnesota

*This author is an associate professor, Department of Economics, University of Minnesota, and visiting associate professor of 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 for his valuable comments.

References

Amemiya, T. [1973], "Regression Analysis When the Dependent Variable Is Truncated Normal," Econometrica 41, pp. 997-1016.

Amemiya, T. [1974], "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal," Econometrica 42, pp. 999-1012.

Bock, R. D. and L. V. Jones [1968], The Measurement and Prediction of Judgment and Choice, Holden Day, San Francisco.

Daganzo, C. [1979], Multinomial Probit, Academic Press, New York.

Domencich, T. A. and D. McFadden [1975], Urban Travel Demand, North-Holland Pub. Co., Amsterdam.

Dubin, J. A. and D. McFadden [1980], "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," Manuscript, Department of Economics, M.I.T.

Duncan, G. M. [1980], "Specification and a Simple Estimator for the Mixed, Continuous/Discrete Dependent Variable in Classical Production Theory," Econometrica 48.

Goldberger, A. S. [1980], "Abnormal Selection Bias," Discussion paper no. 8006, Social Systems Research Institute, University of Wisconsin, Madison.

Hausman, J. and D. Wise [1978], "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogenous References," Econometrica 46, pp. 403-426.

Heckman, J. J. [1974], "Shadow Prices, Market Wages and Labor Supply," Econometrica 42, pp. 679-694.

Heckman, J. J. [1976], "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, pp. 475-492.

Heckman, J. [1979], "Sample Selection Bias as a Specification Error," Econometrica 47, pp. 153-161.

Hildebrand, F. B. [1956], Introduction to Numerical Analysis, McGraw-Hill, New York.

Lee, L. F. [1976], Two-Stage Estimations of Limited Dependent Variable Models, Ph.D. Thesis, University of Rochester, Rochester, N.Y.

Lee, L. F. [1979], "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables," Econometrica 47, pp. 977-996.

Lee, L. F., G. S. Maddala and R. P. Trost [1980], "Asymptotic Covariance Matrices of Two-State Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity," Econometrica 48, pp. 491-503.

Lerman, S. and S. Manski [1980], "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in Manski and McFadden, Editors, Structural Analysis of Discrete Data, M.I.T. Press.

Mardia, K. V. [1970], Families of Bivariate Distributions, Hafner Pub. Co., Conn.

McFadden, D. [1973], "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, Frontiers in Econometrics, Academic Press, New York.

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

Poirier, D. J. [1978], "The Use of the Box-Cox Transformation in Limited Dependent Variable Models," Journal of the American Statistical Association 73, pp. 284-287.