SOME APPROACHES ON THE CORRECTION OF SELECTIVITY BIAS

by

Lung-Fei Lee

Discussion Paper No. 81 - 141, February 1981

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

December 1980

SOME APPROACHES ON THE CORRECTION OF SELECTIVITY BIAS

by

Lung-fei Lee(*)

1. Introduction

The recent developments on the limited dependent variables and censored dependent variables in econometric models attempt to deal with the problems of systematic missing data on the dependent variables for cross sectional survey data. The most common cases are the existence of some selection processes which determine the observed samples. Conditional on the appropriate set of exogenous variables, if the dependent variable on potential outcome in a regression model is correlated with the selection processes, conventional estimation techniques will not provide consistent estimates of the parameters. The common solution in the literature is to specify the joint probability distributions of the random elements in the selection processes and the regression equation. Under the hypothesis that the distribution is the correct one, the maximum likelihood method is consistent and asymptotically efficient under very general conditions, see for example, Amemiya [1973] for the bivariate normal distribution case. Multivariate normal distribution is the most commonly specified assumption in those models. Under this distributional assumption, computationally simple limited information method which corrects directly the source of least squares bias has also been developed, see, for example, Amemiya [1974], Heckman [1976], Lee [1976] among others.

The limited information method cited above utilizes only the information on the first two incomplete moments of the distribution. As a simple example, consider a two equations model with a random sample of size N,

$$y_{1i} = x_i\beta + u_i \tag{1.1}$$

$$y_i^* = z_i\gamma + \varepsilon_i \qquad i=1,\ldots,N. \tag{1.2}$$

where x_i and z_i are exogenous variables, $E(u_i) = 0$, $E(\varepsilon_i) = 0$ and $var(\varepsilon_i) = 1$. The joint distribution of u_i and ε_i conditional on x_i and z_i is bivariate normal, $N(0, 0, \sigma_u^2, 1, \rho)$, where ρ is the correlation coefficient. The dependent variable y_i^* is unobservable but has a dichotomous observable realization I_i which is related to y_i^* as follows:

$$I_i = 1 \text{ if and only if } y_i^* > 0$$

$$I_i = 0 \text{ if and only if } y_i^* \leq 0.$$

The dependent variable y_{1i} conditional on x_i and z_i has well-defined marginal distribution but y_{1i} is not observed unless $y_i^* > 0$. Without lost of generality, let us assume the non-censored observations y_i of y_{1i} are the first N_1 observations.

$$y_i = x_i\beta + u_i \text{ if and only if } z_i\gamma > \varepsilon_i\,, \quad i=1,\ldots,N_1 \tag{1.3}$$

Since u_i and ε_i are bivariate normally distributed, the conditional expectation of u_i given ε_i is linear in ε_i and $u_i = \rho\sigma_u(\varepsilon_i - \mu_\varepsilon)/\sigma_\varepsilon + v_i$ where $\mu_\varepsilon = E(\varepsilon_i)$ and $\sigma_\varepsilon^2 = var(\varepsilon_i)$.[1/] Furthermore; ε_i and v_i are independent. Equation (1.1) can then be written as

$$y_{1i} = x_i\beta + \rho\sigma_u(\varepsilon_i - \mu_\varepsilon)/\sigma_\varepsilon + v_i \ , \quad i=1,\ldots,N \tag{1.4}$$

where $\sigma_\varepsilon = 1$ and $\mu_\varepsilon = 0$. Since $E(\varepsilon_i|z_i\gamma > \varepsilon_i) = -\phi(z_i\gamma)$ and $E(v_i|z_i\gamma > \varepsilon_i) = 0$ where $\phi(z_i\gamma)$ is the standard normal density function evaluated at $z_i\gamma$. The limited information method is to use the implied equation,

$$y_i = x_i\beta - \rho\sigma_u\phi(z_i\gamma)/\sigma_\varepsilon + \rho\sigma_u\mu_\varepsilon/\sigma_\varepsilon + \xi_i, \quad i=1,\ldots,N_1 \tag{1.5}$$

where $\xi_i = v_i + \rho\sigma_u(\varepsilon_i + \phi(z_i\gamma))/\sigma_\varepsilon$ and $E(\xi_i|z_i\gamma \geq \varepsilon_i) = 0$, after the correction of selection bias term. Olsen [1980] has pointed out that underlying the derivation of the equation (1.5), the crucial properties used are the linearity of the conditional expectation of u given ε, the normality of the disturbance ε and the independence of v with ε. Based on these properties, Olsen specifies the regression equation (1.4) as the basic model and suggests a linear probability modification to correct for the selectivity bias in this class of models.

This modification is useful as it provides an alternative way to specify selectivity models without restricting to the assumption of multivariate normal disturbances. The correction of selectivity bias in the equation (1.5) is insensitive to the distribution of v. The correction of the selectivity bias in the equation (1.4) is to compute the conditional mean of ε conditional on $\varepsilon \leq z\gamma$ so as to derive the correct conditional regression equation for the observed samples. For different probability models, there are correspondingly different expressions for the selectivity bias term.[2/] For the linear probability choice model, there is a linear probability correction of the selectivity bias. For the logistic probability model, a corresponding modification of the selectivity bias term is presented in Hay [1980]. However, there are some questions on the general applicability of this approach. For the linear probability model, ε is uniformly distributed on [0, 1] rather than normally distributed together with the assumption that the conditional expectation of u given ε is linear in ε may implicitly impose an outrageous distribution upon u. In the extreme case when $|\rho|$ is closed to one, the density of u is closed to be uniform which is most unlikely distribution for a regression model.[3/] The selectivity bias terms in the regression equation may be quite sensitive to the specific probability models even though there may be only slight differences in the probability models. As pointed out in Domencich and McFadden ([1975],p. 58), the three popular probability models, namely, probit, arctan and logit models, are virtually indistinguishable except at arguments yielding probabilities extremely close to zero or one, and they concluded that, within the range of most data, the three models provide essentially equivalent probability functions, and except for computational reasons, there is little to choose among them. However, the selectivity bias terms for the regression equation will not have the similarity. As the arctan probability model is generated based on the distribution of ε being Cauchy, the conditional mean for the dependent variable y_1 does not exist. The Olsen's approach is thus restrictive because the specified probability model dictates the correction of the selectivity bias. Another problem remained unsolved in Olsen [1980] is to provide a rigorous statistical inference procedure to discriminate between his linear probability correction of the selectivity bias and the correction based on normal distribution. Under the Olsen's approach, it is not clear how that can be done since any specific probability model will lead to a specific selectivity bias term.

In this article, we attempt to overcome the above restrictions in Olsen's approach and suggest a more flexible approach. Under our generalized approach, any specific probability model need not restrict the expression of selectivity bias term in the regression equation and hence a much wider class of models can be derived. Rigorous statistical inference procedure can also be derived to choose among the various corrections of selectivity bias under a commonly specified probability model. Olsen's approach has lately been extended to the polychotomous choice case in Dubin and McFadden [1980] and Hay [1980]. In this article, we also propose some approaches to the correction of selectivity bias in the polychotomous choice models. Our approach is much flexible and the models are much easier to be implemented than theirs. Statistical procedure will also be provided to choose among the models and compare their approach with ours.

2. A Class of Dependence Models

Consider the two equations model,

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

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

where x and z are exogenous variables, $E(u|x,z) = 0$, $var(u|x,z) = \sigma^2$, $E(\varepsilon|x,z) = \mu_\varepsilon$ and $var(\varepsilon|x,z) = \sigma_\varepsilon^2$. The mean μ_ε and the variance σ_ε of ε are assumed to be known.[4/] The observability of the dependent variable y_1 and the dichotomous indicator I are indicated as in the previous section. Let J be a specified strictly increasing transformation. Since

$$\begin{aligned} I = 1 &\Longleftrightarrow z\gamma > \varepsilon \\ &\Longleftrightarrow J(z\gamma) > J(\varepsilon), \end{aligned}$$

the model with equations (2.1) and (2.2) is equivalent to

$$y_1 = x\beta + u \tag{2.3}$$

$$y^{**} = J(z\gamma) - J(\varepsilon) \tag{2.4}$$

where $y^{**} = J(y^*)$.

The class of dependence models that will be considered is based on the specification that u is a convolution of two independent random variables and one of them is proportional to $J(\varepsilon)$. Specifically, we assume

$$u = \lambda(J(\varepsilon) - \mu_J) + v \tag{2.5}$$

where v and $J(\varepsilon)$ are independent and $\mu_J = E(J(\varepsilon))$. The disturbances ε and u in the choice equation and the regression equation are correlated if $\lambda \neq 0$ and uncorrelated if $\lambda = 0$. The correlation of u and ε is derived by transforming independent random variables. This specification can be regarded as special cases in the construction of bivariate distributions due to Steffensen [1922]. This approach provides a way to generalize a large class of models with selectivity. By specifying different transformations, we can allow different implied implicit distributions on u and thus any specific probability choice model need not dictate the way of correcting the selectivity bias term. When the transformation J is the identity mapping, it corresponds to the Olsen's approach. In practice, the appropriate transformation in (2.5) is hardly known. If the transformation J could be estimated within the class of strictly increasing transformations for given samples, it would be desirable. Unfortunately, that does not seem to be possible. However, at least one can try different transformations and select the ones that provide the reasonable results.

When there were some priori information available on u, it might also be useful in providing some suggestions on the specification of the transformations. For example, if the marginal distribution of u is normal and if v were assumed to be normal, the selection of the transformation J such that $J(\varepsilon)$ is a normal random variable seems appropriate. Of course, it is not necessarily true that this can be done for any marginal distribution of u, which is known a priori, under this approach. For models with specific marginal distributions on u and ε, the alternative approaches based on the translation method and the contingency distribution method which generate bivariate distributions with specified marginal distribution in Lee [1980] are more appropriate than this approach. When the specified marginal distribution of u is normal, this approach and the translation method in Lee [1980] are similar. However, the approach in this paper is slightly more general in the correction of the selectivity bias as the bivariate normality is a sufficient condition for the results to hold, but not necessary.

3. The Correction of Selection Bias, Estimation and Model Selection

Let σ_J^2 and ρ be the variance of $J(\varepsilon)$ and the correlation cofficient of u and $J(\varepsilon)$. The equation (2.5) is equivalent to $u = \rho\sigma_u(J(\varepsilon) - \mu_J)/\sigma_J + v$ where $\sigma_v^2 \equiv \mathrm{var}(v) = \sigma_u^2(1-\rho^2)$. The two equations model becomes

$$y_1 = x\beta + \rho\sigma_u(J(\varepsilon) - \mu_J)/\sigma_J + v \tag{3.1}$$

$$y^* = z\gamma - \varepsilon \tag{3.2}$$

The selectivity bias term for the observed dependent variable y is $E(J(\varepsilon)|z\gamma \geq \varepsilon)$, or equivalently, $E(\varepsilon^*|J(z\gamma) \geq \varepsilon^*)$ where $\varepsilon^* = J(\varepsilon)$. Assume that the distribution of ε is known or completely be specified. Let $f_J(.)$ be the implied density function of ε^* which is assumed to exist under the transformation J. Let $\mu(J(z\gamma)) = \int_{J(-\infty)}^{J(z\gamma)} \varepsilon^* f_J(\varepsilon^*)d\varepsilon^*$ denote the incomplete first moment of the random variable ε^* evaluated at $J(z\gamma)$. Let $F(z\gamma) = \Pr(z\gamma \geq \varepsilon)$ be the probability that the event $I = 1$ occurs. Conditional on the sample y being observed, the regression equation (3.1) after the correction of the selectivity bias becomes

$$y = x\beta + \rho\sigma_u(\mu(J(z\gamma))/F(z\gamma) - \mu_J)/\sigma_J + \xi \tag{3.3}$$

where $\xi = \rho\sigma_u(J(\varepsilon) - \mu(J(z\gamma))/F(z\gamma))/\sigma_J + v$ has zero conditional mean, i.e., $E(\xi|x,z, I = 1) = 0$. The conditional variance of ξ is

$$\begin{aligned}\mathrm{var}(\xi|x,z, I=1) &= \frac{\rho^2\sigma_u^2}{\sigma_J^2}[E(J(\varepsilon)^2|z\gamma \geq \varepsilon) - E(J(\varepsilon)|z\gamma \geq \varepsilon)^2] + \sigma_u^2(1-\rho^2) \\ &= \frac{\rho^2\sigma_u^2}{\sigma_J^2}\left[\frac{\mu_2(J(z\gamma))}{F(z\gamma)} - \left(\frac{\mu(J(z\gamma))}{F(z\gamma)}\right)^2\right] + \sigma_u^2(1-\rho^2)\end{aligned} \tag{3.4}$$

where $\mu_2(J(z\gamma)) = \int_{J(-\infty)}^{J(z\gamma)} \varepsilon^{*2} f_J(\varepsilon^*)d\varepsilon^*$ is the incomplete second moment around zero of ε^* evaluated at $J(z\gamma)$. Since the distribution of ε and the transformation J have been completely specified, μ_J and σ_J are known parameters and the remaining unknown parameters of the model are β, ρ, γ and $\sigma_u{}^2$.

The nonlinear equation (3.3) can be estimated by similar two stage method as discussed in the literature, see for example, Amemiya [1974], Heckman [1976] and Lee [1976], among others. In the first stage estimation, γ can be estimated by the maximum likelihood method for the implied probability choice model under the assumed distribution F(.) for the disturbance ε. Let $\hat{\gamma}$ denote the maximum likelihood estimate of γ. The second stage estimation is to estimate the modified equation of (3.3) with the noncensored observations,

$$y_i = x_i\beta + \rho\sigma_u(\mu(J(z_i\hat{\gamma}))/F(z_i\hat{\gamma}) - \mu_J)/\sigma_J + \tilde{\xi}_i \quad , \; i=1,\ldots,N_1 \qquad (3.5)$$

by the ordinary least squares procedure (OLS). Under very general conditions, the OLS estimate of β, $\rho\sigma_u$ can be shown to be consistent and asymptotically normal for random samples under the specification (3.1) as in Lee and Trost [1978]. Correct asymptotic variance matrix for the estimates can also be derived as in Lee et. al. [1980] with slight modifications to take into account the presence of the transformation J. The parameter $\sigma_u{}^2$ can then be estimated with the estimated residuals of $\tilde{\xi}$ by several methods as described in Lee and Trost [1978]. The detail derivations are refered to those articles and are omitted here.

The above paragraphs outline the correction of the selectivity bias in the regression equation and the simple two stage estimation method for those models. In practice, whether the method is really simple or not will depend on the specified transformation J. A general class of transformations that is rich enough is to specify the transformation $J = G_o^{-1}F$ where G is an absolutely continuous distribution function. As ε is specified to have the distribution function $F(\varepsilon)$, the transformed variable $\varepsilon^* = G^{-1}(F(\varepsilon))$ will be a random variable with distribution function $G(\varepsilon^*)$.

The distribution of the random variable u under the convolution formulae (2.5) can take on various shapes as the distribution functions G(.) vary, while the probability model can be chosen to be a specific model and remains unchanged. Some popular random variables in the literature of probability theory will be rich enough to serve our purpose; consider, for example, the continuous univariate distributions in the two volumes of Johnson and Kotz [1970]. The correction of the selectivity bias term and the conditional variance of ξ require the derivation of the first two incomplete moments of some popular random variables. As a convenient reference, we provide a list of the formulae for these two incomplete moments for many popular random variables in the appendix. Thus, for example, if the probability choice model is a linear probability model as considered in Olsen [1980], we can have the uniform distribution correction for selectivity bias in Olsen when J(.) is an identity mapping, as well as the normal distribution correction for selectivity bias when J(.) is chosen to be $\Phi^{-1}(.)$, where $\Phi(.)$ is the standard normal distribution function.

Since different transformations J lead to different regression equations after correcting the selectivity bias term, we have a model selection problem. Since, in our approach, the probability choice model can be fitted separately with the samples of dichotomous indicators and presumably can be chosen according to some goodness of fit criteria as derived in Domencich and McFadden [1976], it will remain unchanged in the estimation of the remaining outcome regression equation. Suppose there was a finite number of transformations J. The problem of selecting the regression equations in (3.3) or (3.5) can be regarded as a special case in the problem of selection of regressors considered in Theil [1961], Mallows [1973] and, most recently, Amemiya [1980], among others. Amemiya's Prediction Criteria (PC) which is applicable to linear or nonlinear regression models with general variance-covariance matrix without a specified distribution seems to be an interesting criteria for our problem since the disturbances in the regression equation (3.3) are heteroscedastic. Since, in our models, all the regression equations have the same number of regressors, the PC will select the equation with the smallest average variances of the disturbances ξ, i.e., the equation with the smallest estimated value of $N_1^{-1} \Sigma_{i=1}^{N_1} \mathrm{var}(\xi_i | x_i, z_i, I_i=1)$ where N_1 is the number of observations on Y_1. The PC is convenient to be used since it provides a single index. The PC is derived based on the principle of minimizing an estimate of the mean square prediction error but as explicitly pointed out in Amemiya, all this kind of criteria considered in the literature are based on a somewhat arbitrary assumption which cannot be fully justified. It can best be used with other knowledge of the underlying economic problem. In the selectivity models, priori theoretical consideration such as the possibility of positive self-selection, i.e., conditional on the exogenous variables, the observed outcome should be greater than population mean, will indicate that the selectivity bias term multiplied by the coefficient, i.e., $\rho\sigma_u(\mu(J(z\gamma))/F(z\gamma) - \mu_J)/\sigma_J$, should be non-negative. Knowledge of this sort will allow us to reject some of the estimated equations.

Another entirely different approach that can be useful for our problem is to nest the competitive models in a generalized equation. This latter approach provides a probabilistic statement regarding the choice between any two competing models and is in the spirit of procedures due to Cox [1961, 1962]. Consider two regression equations with different transformations J_1 and J_2; the first model is

$$y_i = x_i\beta + \rho\sigma_u[\mu(J_1(z_i\hat{\gamma}))/F(z_i\hat{\gamma}) - \mu_{J_1}]/\sigma_{J_1} + \tilde{\xi}_i \tag{3.6}$$

and the second model is

$$y_i = x_i\beta + \rho\sigma_u[\mu(J_2(z_i\hat{\gamma}))/F(z_i\hat{\gamma}) - \mu_{J_2}]/\sigma_{J_2} + \tilde{\xi}_i \tag{3.7}$$

where $i = 1,\ldots, N_1$. These two equations can be nested into a general equation as

$$y_i = x_i\beta + \lambda_1[\mu(J_1(z_i\hat{\gamma}))/F(z_i\hat{\gamma}) - \mu_{J_1}] + \lambda_2[\mu(J_2(z_i\hat{\gamma}))/F(z_i\hat{\gamma}) - \mu_{J_2}] + \tilde{\xi}_i$$

$$i = 1,\ldots, N_1 \tag{3.8}$$

The latter equation contains the two models in (3.6) and (3.7). When $\lambda_2 = 0$, it reduces to the first model and it reduces to the second model when $\lambda_1 = 0$. The discrimination of the two models is related to the test of the significance of the coefficients λ_1 and λ_2. The hypothesis that the first model is the correct one is equivalent to the hypothesis that $\lambda_2 = 0$. The equation (3.8) can be estimated by the OLS procedure. Under general conditions, an asymptotic normal statistic can be derived as follows. Let $\tilde{X}$ be the $N_1 x(k+2)$ data matrix of the regressors in (3.8) where k is the dimension of the parameter β, i.e.,

$$\tilde{X}' = \begin{bmatrix} x_1' & \cdots\cdots & x_{N_1}' \\ \mu(J_1(z_1\hat{\gamma}))/F(z_1\hat{\gamma}) - \mu_{J_1} & \cdots\cdots & \mu(J_1(z_{N_1}\hat{\gamma}))/F(z_{N_1}\hat{\gamma}) - \mu_{J_1} \\ \mu(J_2(z_1\hat{\gamma}))/F(z_1\hat{\gamma}) - \mu_{J_2} & \cdots\cdots & \mu(J_2(z_{N_1}\hat{\gamma}))/F(z_{N_1}\hat{\gamma}) - \mu_{J_2} \end{bmatrix}$$

Let d_{ji} be the gradient vector of $\frac{\partial}{\partial\gamma}(\mu(J_j(z_i\gamma))/F(z_i\gamma))$ and $D_j' = [d_{ji}', \ldots d_{jN_1}']$, $j = 1, 2$. Furthermore, let V_j be the $N_1 x N_1$ diagonal matrix defined as

$$V_j = \mathrm{Diag}[\sigma_v^{\,2} + \lambda_j^2 \left(\frac{\mu_2(J_j(z_i\gamma))}{F(z_i\gamma)} - \left(\frac{\mu_1(J_j(z_i\gamma))}{F(z_i\gamma)}\right)^2\right)],\ j = 1, 2.$$

It follows that under the null hypothesis H_0: $\lambda_2 = 0$, the OLS estimates $\hat{\beta}$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are asymptotically normal and their asymptotic variance-covariance matrix is

$$\mathrm{var}\begin{pmatrix} \hat{\beta} \\ \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'(V_1 + D_1\Omega_{\hat{\gamma}}^{-1}D_1')\tilde{X}(\tilde{X}'\tilde{X})^{-1} \tag{3.9}$$

where $\Omega_{\hat{\gamma}}$ denotes the asymptotic variance-covariance matrix of the estimate $\hat{\gamma}$ for the choice equation. Similarly, if we like to test whether the second model is the correct one, we can test the hypothesis that $\lambda_1 = 0$. Under this hypothesis, the asymptotic variance-covariance matrix in (3.9) should be changed with D_2 and V_2 replacing D_1 and V_1, respectively, in the expression. Similar to the suggestions in Cox [1961, 1962], each model should be tested once as the null hypothesis. It is possible to reject both the models as both of them are not necessarily the correct ones. In the event that both the models will be accepted, it would be likely that both models provide similar results.

4. Polychotomous Choice Models and Selectivity Bias

The approach discussed in the previous paragraphs can be generalized to the case with polychotomous choices. There are at least two possible ways for the generalization. The first approach generalizes slightly the approach in Hay [1980] and Dubin and McFadden [1980]. This approach is based on the point of view that polychotomous choice model can be formulated as models with multiple binary choice rules with partial observations. The second approach is motivated by the formulation of order statistics in the polychotomous choice models.

Consider the following polychotomous choice model with M categories and one potential outcome regression equation in each category.

$$\begin{aligned} y_{si} &= x_{si}\beta_s + u_{si} & s &= 1,\ldots, M \\ y^*_{si} &= z_{si}\gamma + \eta_{si} & i &= 1,\ldots, N \end{aligned} \qquad (4.1)$$

where i refers to the ith observation, all the variables x_s, z_s are exogeneous, $E(u_s|x_1,\ldots,x_M, z_1,\ldots,z_M) = 0$ and the joint distribution of $(\eta_1,\ldots,\eta_M)$ has been completely specified. The dependent variable or potential outcome y_s in the sth category is observed if and only if the sth category is chosen. Let I be a polychotomous variable with values 1 to M and I = s if the sth category is chosen.

$$I = s \text{ if and only if } z_s\gamma - z_j\gamma > \eta_j - \eta_s \text{ for all } j \in \{1,\ldots,M\}-\{s\}$$

This formulation is to relate the polychotomous choice model as model with M-1 binary decision rules with partial observations.[5/] An alternative formulation is that

$$I = s \quad \text{if and only if} \quad y_s^* > \max_{\substack{j=1,\ldots,M \\ j\neq s}} y_j^*$$

Let

$$\varepsilon_s = \max_{\substack{j=1,\ldots,M \\ j\neq s}} y_j^* - \eta_s \tag{4.2}$$

It follows that I=s if and only if $z_s\gamma > \varepsilon_s$. This formulation is to relate the choice of the sth alternative as a binary decision, i.e., the sth alternative will either be chosen or not, mutually exclusively.

Based on the first formulation, one approach to specify the regression equations y_s with u_s correlated with the choice equations is to assume that

$$u_s = \Sigma_{\substack{j=1 \\ j\neq s}}^{M} \lambda_{sj}(J_{sj}(\eta_j-\eta_s) - \mu_{J_{sj}}) + v_s \quad s=1,\ldots,M \tag{4.3}$$

where all the J_{sj} are some strictly increasing transformations, $\mu_{J_{sj}}=E(J_{sj}(\eta_j-\eta_s))$ and for each s, v_s is assumed to be independent with $J_{sj}(\eta_j-\eta_s)$ for all $j\varepsilon\{1,\ldots,M\}-\{s\}$ Equivalently, the formulation in (4.3) can be rewritten as

$$u_s = \Sigma_{u_sJ_s}\Sigma_{J_s}^{-1}(J_s(\eta)-\mu_{J_s}) + v_s \tag{4.4}$$

where $J_s(\eta)\equiv(J_{s1}(\eta_1-\eta_s),\ldots,J_{ss-1}(\eta_{s-1}-\eta_s),\ J_{ss+1}(\eta_{s+1}-\eta_s),\ldots,J_{sM}(\eta_M-\eta_s))'$ and $\mu_{J_s} \equiv (\mu_{J_{s1}},\ldots,\mu_{J_{ss-1}},\ \mu_{J_{ss+1}},\ldots,\mu_{J_{sM}})'$ are two column vectors, $\Sigma_{u_sJ_s}$ is the covariance vector of u_s and $J_s(\eta)'$ and Σ_{J_s} is the variance-covariance matrix of the vector $J_s(\eta)$. When all the transformation J_{sj} are chosen to be the identity mapping, this specification is the approach in Dubin and McFadden [1980] and Hay [1980]. Under the specification in (4.3), it implies that the observed dependent variable of the outcome equation, conditional on the sth category being chosen, will satisfy the following equation after the correction of selectivity bias term,

$$y_s = x_s\beta_s + \Sigma_{\substack{j=1\\ j\neq s}}^{M} \lambda_{sj}(T_{J_{sj}}(z_1\gamma,\ldots,z_M\gamma) - \mu_{J_{sj}}) + \xi_s \tag{4.5}$$

where $T_{J_{sj}}(z_1\gamma,\ldots,z_M\gamma) \equiv E(J_{sj}(\eta_j-\eta_s)|z_s\gamma-z_j\gamma > \eta_j-\eta_s,\ j\varepsilon\{1,\ldots,M\}-\{s\})$ is the selectivity bias term and $\xi_s = v_s + \Sigma_{\substack{j=1\\ j\neq s}}^{M} (J_{sj}(\eta_j-\eta_s) - T_{J_{sj}}(z_1\gamma,\ldots,z_M\gamma))$. It follows that, conditional on I=s, the disturbances ξ_s have zero mean but are heteroscedastic errors. The variances of ξ_s involve expressions of the incomplete second moments around zero of the transformed random variables $J_{sj}(\eta_j-\eta_s)$ and their incomplete cross second moments. The equation (4.5) can be estimated by some two stage method. With the parameter vector γ estimated from the polychotomous choice model as $\hat{\gamma}$, the modified equation

$$y_s = x_s\beta_s + \Sigma_{\substack{j=1\\ j\neq s}}^{M} \lambda_{sj}(T_{J_{sj}}(z_1\hat{\gamma},\ldots,z_M\hat{\gamma}) - \mu_{J_{sj}}) + \tilde{\xi}_s$$

can then be estimated by the OLS procedure.

Whether the above approach is really computationally simple or not depends on the evaluations of the first two incomplete moments of the random variables $J_{sj}(\eta_j-\eta_s)$ and the specification of the polychotomous choice model. One of the widely used polychotomous choice model is the conditional logit model in McFadden [1973]. The conditional logit model is derived under the utility maximization hypothesis; the assumption that the η_j, j=1,...,M, are independently identically distributed (i.i.d.) with Gumbel distribution (with parameter 0), i.e., $\text{Prob}[\eta_j \leq \eta] = \exp(-\exp(-\eta))$, and other minor conditions. This distributional assumption implies that the M-1 random variables $\omega_{sj}=\eta_j-\eta_s$, $j\varepsilon\{1,\ldots,M\}-\{s\}$ will have a multivariate logistic distribution of Gumbel [1961], i.e., the joint distribution is

$$F_L(\omega_{s1},\ldots,\omega_{ss-1},\omega_{ss+1},\ldots,\omega_{sM}) = (1 + \Sigma_{\substack{j=1\\ j\neq s}}^{M} e^{-\omega_{sj}})^{-1}. \tag{4.6}$$

Let $\ell'=(1,\ldots,1)$ be a M-1 dimensional vector with all ones. The variance-covariance matrix of the M-1 vector ω_s is

$$\Sigma_\omega = \frac{\pi^2}{6}(I+\ell\ell').$$

More detail description of this distribution and its properties can be found in Chapter 42; Johnson and Kotz [1972]. Let us now consider in more detail the implementation of the selectivity model with the conditional logit model. Since η_j, $j=1,\ldots,M$ are i.i.d., we assume that $E(u_s J_s(\eta))=\sigma_{sJ_s}\ell$. As the joint distribution of $J_s(\eta)$ has been completely specified, Σ_{J_s} is a known matrix and μ_{J_s} is a known vector. It follows that

$$u_s = \sigma_{sJ_s}\ \ell'\Sigma_{J_s}^{-1}(J_s(\eta) - \mu_{J_s}) + v_s \tag{4.7}$$

and the number of parameters in (4.4) has been reduced to two. When the transformations J_{sj} are identities, this model is exactly the model considered in Dubin and McFadden [1980] and Hay [1980].

However, even for the conditional logit model, this approach does not provide analytical closed form expressions for the selectivity bias terms for the general class of transformations considered in the previous section. Consider for example, the selectivity bias term for j=1,

$$\begin{aligned} &T_{J_{s1}}(z_1\gamma,\ldots,z_M\gamma) \\ &= E(J_{s1}(\omega_{s1})\,|\,t_{sj} > \omega_{sj},\ j\varepsilon\{1,\ldots,M\}-\{s\}) \\ &= \int_{-\infty}^{z_s\gamma-z_1\gamma} J_{s1}(\omega)\frac{\partial F_L}{\partial\omega}(\omega,\ t_{s2},\ldots,t_{ss-1},\ t_{ss+1},\ldots,t_{sM})d\omega / \\ &\qquad F_L(t_{s1},\ldots,t_{ss-1}\ t_{ss+1},\ldots,t_{sM}) \end{aligned}$$

$$= \int_{-\infty}^{z_s\gamma - z_1\gamma} J_{s1}(\omega)\, \frac{e^{-\omega}}{(C_{s1}+e^{-\omega})^2}\, d\omega / F_L(t_{s1},\ldots,t_{ss-1},t_{ss+1},\ldots,t_{sM}) \quad (4.8)$$

where $t_{sj} \equiv z_s\gamma - z_j\gamma$ and $C_{s1} = 1 + \Sigma_{\substack{j=1 \\ j\neq s}}^{M} e^{-t_{sj}}$. The integral in (4.8) does not seem to have closed form expressions except for some simple transformations such as the identity transformation. The evaluation of the variances of ξ_s in (4.5) are even more complicated and involve double integrals. When the transformations are identities, closed form expression for the selectivity bias term can be derived and as shown in the appendix, see also Dubin and McFadden [1980] and Hay [1980], we have

$$E(\omega_{si} | t_{sj} > \omega_{sj},\ j\varepsilon\{1,\ldots,M\}-\{s\})$$

$$= (1 - e^{-t_{sj}} F_L(t_s))^{-1} [\ell n F_L(t_s) - t_{si} e^{-t_{si}} F_L(t_s)],\ i\varepsilon\{1,\ldots,M\}-\{s\} \quad (4.9)$$

where $F_L(t_s) \equiv F_L(t_{s1},\ldots,t_{ss-1},t_{ss+1},\ldots,t_{sM})$ in (4.6). Unfortunately, even for this case, the evaluation of the second moments of ω_s does not seem to have closed form expressions, see the appendix. For more general polychotomous choice model such as the generalized extreme value distribution in McFadden [1977], this approach will not be simpler. Thus this approach does not seem to be able to generate large class of computational simple selectivity models.

Let us now consider an alternative approach based on the second formulation. Under the second formation, I=s if and only if $z_s\gamma > \varepsilon_s$, where ε_s is defined in (4.2). Let $F_s(.)$ denote the implied distribution function of ε_s. For example, if η_j, j=1,...,M, are i.i.d. Gumbel distributed, $F_s(\varepsilon)$ will be a logistic distribution with $F_s(\varepsilon) = \exp(\varepsilon)/[\exp(\varepsilon) + \Sigma_{\substack{j=1 \\ j\neq s}}^{M} \exp(Z_j\gamma)]$. Let J_s be a strictly increasing transformation of ε_s which transforms ε_s to a random variable $J_s(\varepsilon_s)$ with constant mean and variance. The alternative approach is to assume that

$$u_s = \lambda_s(J_s(\varepsilon_s) - \mu_{J_s}) + v_s \quad s=1,\ldots,M \tag{4.10}$$

where v_s and $J_s(\varepsilon_s)$ are independent and μ_{J_s} denotes the mean of $J_s(\varepsilon_s)$. This approach is almost exactly the approach for the binary choice case. The class of transformation $J_s = G_o^{-1}F_s$ where G is any popular distribution function will generate a large class of interesting and computational simple selectivity model by the same arguments for the binary choice model. This approach seems to be more flexible than the first one and also generalizes the approach in Lee [1980] without imposing marginal normal distributional assumption on u_s. For the case that u_s and η_s for all $s=1,\ldots,M$ are multivariate normal, it implies the relation (4.3) with all the transformation J_{sj} being identities and the first approach will be the proper one. Except for those cases, there does not seem to have theoretical reasons to prefer one approach over the other. From the computational point of view, the second approach will be simpler.

Finally, we note that the model selection procedures discussed in the previous sections are also applicable to the polychotomous choice models. Thus we can compare the selectivity models generated from the same approach or models generated from the two different approaches.

5. Conclusions

This article has considered the specification of some econometric models with selectivity. Our approaches generalize the approach in Olsen [1980], and allow us to relax much of the restrictions imposed on the potential outcome regression equation by the distributional assumption imposed on the probability discrete choice equation. Our approaches provide various ways to specify and correct the selectivity bias in the observed outcomes in the regression models. Statistical procedures are suggested so that one can select the best fitted model among many competitive models that one may like to consider. The models can all be estimated by simple consistent two stage methods similar to those suggested in the limited and censored dependent variables literature. Simplified Cox type model discrimination procedure is also suggested so that one can test the competitive model hypothesis. This provides a rigorous procedure to discriminate the corrections of selectivity bias based on the normal distribution and some non-normal distributions. We have also generalized our approaches to models with polychotomous discrete choices. The corrections of the selectivity bias in our approaches are also very simple and the problem of estimation is much simpler in our models than the model specified in Dubin and McFadden [1980] and Hay [1980]. Simple two stage methods for the estimation and the model selection procedures are also available. The model selection procedures provide ways to discriminate our models with theirs.

Appendix: List of Truncated First and Second Moments for Some Distributions

Let us define some common notations to simplify the expressions. Let $f(\varepsilon)$ denote the density function, $F(\varepsilon)$ be the distribution function, $\mu_1(x)=E(\varepsilon|\varepsilon\leq x)$ be the truncated mean and $\mu_2(x)=E(\varepsilon^2|\varepsilon\leq x)$ be the truncated second moment around zero. The following list of distributions cover most the continuous univariate distributions listed in Johnson and Kotz [1970a, 1970b]. The detail derivations of the expressions are straightforward and will be omitted.

Normal Distribution: $f(\varepsilon)=(2\pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}\varepsilon^2)$; $-\infty < \varepsilon < \infty$.

$$\mu_1(x) = -f(x)/F(x), \quad \mu_2(x) = 1 - xf(x)/F(x)$$

The first expression can be found in Raiffa and Schlaifer [1961], p. 231 and both expressions can be found in Johnson and Kotz [1970a] pp. 81-83.

Student Distribution: $f(\varepsilon) = \frac{\nu^{\frac{1}{2}\nu}}{B(\frac{1}{2},\frac{1}{2}\nu)} (\nu+\varepsilon^2)^{-\frac{1}{2}(\nu+1)}$; $\nu>2$, $-\infty<\varepsilon<\infty$,

where $B(a,b)$ is the Beta function with parameters a and b.

$$\mu_1(x) = -\frac{\nu+x^2}{\nu-1} f(x)/F(x),$$

$$\mu_2(x) = \frac{\nu B(\frac{1}{2},\frac{1}{2}\nu-1)}{2B(\frac{1}{2},\frac{1}{2}\nu)} [1 + \mathrm{sgn}(x) F_\beta(\frac{x^2}{\nu+x^2}|\tfrac{1}{2},\tfrac{1}{2}\nu-1)]/F(x) - \nu$$

where $F_\beta(u|a,b)$ is the Beta distribution function with parameters a and b evaluated at u, and sgn(x) is a sign function defined as

$$\mathrm{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

The first expression can be found in Raiffa and Schlaifer [1961], p. 233.

Logistic Distribution: $f(\varepsilon) = e^{-\varepsilon}/(1 + e^{-\varepsilon})^2 \quad -\infty<\varepsilon<\infty$

$$\mu_1(x) = x + \ell n(1 - F(x))/F(x),$$

$$\mu_2(x) = [\frac{\pi^2}{3} + \mathrm{sgn}(x)\ \Sigma_{j=1}^{\infty}\ (-1)^{j-1} j^{-2} \Gamma_{j|x|}(3)]/F(x)$$

where $\Gamma_{j|x|}(a) = \int_0^{j|x|} \varepsilon^{a-1} e^{-\varepsilon} d\varepsilon$ is the incomplete Gamma function with parameter a. The first expression has been derived in Goldberger [1980] and Hay [1980]. The second expression can also be found in Hay [1980].

Laplace Distribution: $f(\varepsilon) = \frac{1}{2}e^{-|\varepsilon|}, \quad -\infty<\varepsilon<\infty$

$$\begin{aligned}\mu_1(x) &= x - 1 && \text{for } x \leq 0,\\ &= -(x+1)\ f(x)/F(x) && \text{for } x \geq 0\end{aligned}$$

$$\begin{aligned}\mu_2(x) &= x^2-2x+2 && \text{for } x \leq 0,\\ &= [3/2-(x^2+x+1)f(x)]/F(x) && \text{for } x \geq 0.\end{aligned}$$

The expression $\mu_1(x)$ has been derived in Goldberger [1980].

Uniform Distribution: $f(\varepsilon) = 1, \quad 0\leq\varepsilon\leq1$

$$\mu_1(x) = \frac{1}{2}x, \qquad \mu_2(x) = x^2/3$$

where $0\leq x\leq 1$.

Beta Distribution: $f(\varepsilon|p,q) = \frac{1}{B(p,q)}\ \varepsilon^{p-1}(1-\varepsilon)^{q-1}; \quad p,q>0, \quad 0\leq\varepsilon\leq1$

$$\mu_1(x) = \frac{p}{p+q}\ F_\beta(x|p+1,q)/F_\beta(x|p,q),$$

$$\mu_2(x) = \frac{p(p+1)}{(p+q)(p+q+1)}\ F_\beta(x|p+2,q)/F_\beta(x|p,q)$$

The expression $\mu_1(x)$ can be found in Raiffa and Schlaifer [1961], p. 216.

Lognormal Distribution: $\varepsilon = e^u$ where u is a standard normal random variable.

$$\mu_1(x) = e^{\frac{1}{2}}\Phi(\ell n\ x-1)/F(x),$$

$$\mu_2(x) = e^2\Phi(\ell n\ x-2)/F(x)$$

where $\Phi(z)$ is the standard normal distribution function evaluated at z.

Exponential Distribution: $f(\varepsilon) = \frac{1}{\sigma} e^{-\varepsilon/\sigma}, \quad \sigma > 0, \quad \varepsilon > 0$

$$\mu_1(x) = \sigma\Gamma_{x/\sigma}(2)/F(x), \quad \mu_2(x) = 2\sigma^2\Gamma_{x/\sigma}(3)/F(x)$$

Gamma Distribution: $f(\varepsilon|\alpha) = \varepsilon^{\alpha-1}e^{-\varepsilon}/\Gamma(\alpha), \quad \alpha > 0, \quad \varepsilon \geq 0.$

$$\mu_1(x) = \alpha F_\gamma(x|\alpha+1)/F_\gamma(x|\alpha), \quad \mu_2(x) = \alpha(\alpha+1)F_\gamma(x|\alpha+2)/F_\gamma(x|\alpha)$$

where $F_\gamma(z|a)$ is the standard Gamma distribution function with parameter a. The first expression can be found in Raiffa and Schlaifer [1961], p. 222.

Chi-square Distribution: $f(\varepsilon|\nu) = \dfrac{\varepsilon^{\frac{1}{2}\nu-1}e^{-\varepsilon/2}}{2^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}\nu)}\ ; \ \nu > 0, \quad \varepsilon \geq 0$

$$\mu_1(x) = \nu F_\gamma(\tfrac{x}{2}|\tfrac{1}{2}\nu+1)/F(x), \quad \mu_2(x) = \nu(\nu+2)F_\gamma(\tfrac{x}{2}|\tfrac{1}{2}\nu+2)/F(x).$$

The first expression can be found in Raiffa and Schlaifer [1961], p. 227.

F Distribution: $f(\varepsilon|\nu_1,\nu_2) = \dfrac{1}{B(\frac{1}{2}\nu_1,\frac{1}{2}\nu_2)} \dfrac{\varepsilon^{\frac{1}{2}\nu_1-1}(\nu_2/\nu_1)^{\frac{1}{2}\nu_2}}{(\varepsilon+\nu_2/\nu_1)^{\frac{1}{2}(\nu_1+\nu_2)}}\ ; \ \nu_1>0, \ \nu_2>4, \ \varepsilon\geq 0.$

$$\mu_1(x) = \frac{\frac{1}{2}\nu_2}{\frac{1}{2}\nu_2-1} F_\beta(\frac{x}{x+\nu_2/\nu_1}|\tfrac{1}{2}\nu_1 + 1, \tfrac{1}{2}\nu_2 \quad - 2)/F(x),$$

$$\mu_2(x) = \frac{\frac{1}{2}\nu_1(\frac{1}{2}\nu_1+1)}{(\frac{1}{2}\nu_2-1)(\frac{1}{2}\nu_2-2)} \quad (\frac{\nu_2}{\nu_1})^2 F_\beta(\frac{x}{x+\nu_2/\nu_1}|\tfrac{1}{2}\nu_1 + 2, \tfrac{1}{2}\nu_2 - 2)/F(x).$$

Weibull Distribution: $f(\varepsilon) = c\, \varepsilon^{c-1}\, e^{-\varepsilon^{c}}, \quad c > 0, \quad \varepsilon > 0$

$$\mu_1(x) = \Gamma_{x^c}\left(\frac{1}{c} + 1\right)/F(x), \qquad \mu_2(x) = \Gamma_{x^c}\left(\frac{2}{c} + 1\right)/F(x)$$

Multivariate Logistic Distribution: $F(v_1,\ldots,v_J) = [1 + \Sigma_{j=1}^{J}\, e^{-v_j}]^{-1}$,

$$E(v_1 \mid v_1 < x_1,\ldots,v_J < x_J)$$

$$= \frac{1}{1-e^{-x_1}F(x_1,\ldots,x_J)} \{\ell n F(x_1,\ldots,x_J) - x_1 e^{-x_1} F(x_1,\ldots,x_J)\}$$

$$E(v_1^{\,2} \mid v_1 < x_1,\ldots,v_J < x_J)$$

$$= \int_{-\infty}^{x_1+\ell n c(x)} \frac{(w-\ell n c(x))^2}{c(x)} \frac{e^{w}}{(1+e^{w})^2}\, dw, \qquad c(x) = 1 - e^{-x_1}F(x_1,\ldots,x_J)$$

$$= \frac{1}{c(x)F(x_1,\ldots,x_J)} \{\frac{\pi^2}{3} + \mathrm{sgn}(x_1 + \ell n c(x))\, \Sigma_{j=1}^{\infty}\, (-1)^{j-1} j^{-2} \Gamma_{j\,|x_1+\ell n c(x)|} \tag{3}$$

$$-2\, \ell n c(x)\, [(x_1 + \ell n c(x))\, G(x_1 + \ell n c(x)) + \ell n\, (1 - G(x_1 + \ell n c(x)))]$$

$$+ (\ell n c(x))^2\, G(x_1 + \ell n c(x))\}$$

where $c(x) = 1 - e^{-x_1}F(x_1,\ldots,x_J)$ and $G(z) = \frac{e^{z}}{1+e^{z}}$ is the standard logistic distribution.

$$E(v_1 v_2 \mid v_1 < x_1,\ldots,v_J < x_J)$$

$$= 2\int_{-\infty}^{x_2} v_2 e^{-v_2}(1+e^{-v_2}+\Sigma_{j=3}^{J} e^{-x_j})^{-1} [\ell n F(x_1,v_2,x_3,\ldots,x_J) - x_1 \frac{e^{-x_1}}{1+e^{-x_1}+e^{-v_2}+\Sigma_{j=3}^{J} e^{-x_j}}]$$

$$dv_2/F(x_1,\ldots,x_J)$$

Footnotes

(*) This author is an Associate Professor, Department of Economics, University of Minnesota, Minneapolis, 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.

(1) We have formally adopted these notations so as to allow other distributional assumptions on ε which need not necessarily imply zero mean and variance one.

(2) Specifically, the 'selectivity bias term' or 'selectivity bias' terminologies in this article are referred to the conditional expectation $E(u|\varepsilon\leq z\gamma)$ for the binary choice case and $E(u_s|$sth category is chosen) for the polychotomous choice case where u_s is the disturbance in the outcome equation in the sth category.

(3) This problem has been pointed out in Olsen [1980].

(4) In general, for the identification of the choice equation, either μ_ε and σ_ε^2 are known constants or will be appropriately normalized to some specific values.

(5) For each s, there are M-1 binary decision rules which can be defined as $D_{sj} = z_s\gamma - z_j\gamma + \eta_s - \eta_j$, $j\varepsilon\{1,\ldots,M\}-\{s\}$, I=s if and only if $D_{sj} > 0$ for all $j\varepsilon\{1,\ldots,M\}-\{s\}$.

References

Amemiya, T. (1974), "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal", Econometrica 41, pp. 997-1016.

Amemiya, T. (1980), "Selection of Regressors", International Economic Review, Vol. 21, pp. 331-354.

Cox, D. R. (1961), "Tests of Separate Families of Hypotheses", in J. Neyman, ed., Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 105-123, University of California Press, Berkeley.

Cox, D. R. (1962), "Further Results on Tests of Separate Families of Hypotheses", Journal of the Royal Statistical Society, Series B, Vol. 24, pp. 406-424.

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, MIT.

Goldberger, A. S. (1980), "Abnormal Selection Bias", Discussion Paper No. 8006, Social Systems Research Institute, University of Wisconsin, Madison.

Gumbel, E. J., (1961), "Bivariate Logistic Distributions", Journal of the American Statistical Association, 56, pp. 335-349.

Hay, J. (1980), "Selectivity Bias in a Simultaneous Logit-OLS Model", Manuscript, University of Southern California.

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.

Johnson, N. L., and S. Kotz (1970a), Continuous Univariate Distributions-1, Houghton-Mifflin Co., Boston.

Johnson, N. L. and S. Kotz (1970b), Continuous Univariate Distributions-2, Houghton Mifflin Co., Boston.

Johnson, N. L. and S. Kotz (1972), Distributions in Statistics: Continuous Multivariate Distributions, John Wiley and Sons, Inc., New York.

Lee, L. F. (1976), Two-Stage Estimations of Limited Dependent Variable Models, Ph.D. Thesis, University of Rochester, New York.

Lee, L. F., and R. P. Trost (1978), "Estimation of Some Limited Dependent Variable Models with Application to Housing Demand", Journal of Econometrics, Vol. 8, No. 3., pp. 357-383.

Lee, L. F., G. S. Maddala and R. P. Trost (1980), "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Models for Simultaneous Equations Models with Selectivity", Econometrica 48, pp. 977-996.

Lee, L. F. (1980), "Generalized Econometric Models with Selectivity", Manuscript, Center for Econometrics and Decision Sciences, University of Florida.

Mallows, C. L. (1973), "Some Comments on C_p", Technometrics 15, pp. 661-676.

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

McFadden D., (1977), "Quantitative Methods for Analyzing Travel Behavior of Individuals: Some Recent Developments", Proc. 3rd, Int. Conf. on Behavioral Travel Modelling, Adelaide, Australia (in press).

Olsen, R. J. (1980), "A Least Squares Correction for Selectivity Bias", Econometrica, Vol. 48, No. 7, pp. 1815-1820.

Raiffa, H. and R. Schlaifer (1961), Applied Statistical Decision Theory, Graduate School of Business Admin., Harvard U. Press, Boston.

Steffensen, J. F. (1922), "A Correlation-Formula", Skand. Aktuar., 5, pp. 73-91.

Theil, H. (1961), Economic Forecasts and Policy, 2nd Edition, North-Holland Pub. Co., Amsterdam.