

FIML Estimation of Sample Selection Models for Count Data

by

William H. Greene

EC-97-02

July 1997

FIML Estimation of Sample Selection Models for Count Data

William H. Greene¹

Stern School of Business, New York University

July, 1997

Abstract

This paper presents an estimator for a model of sample selection for count data. The model is an extension of the standard sample selectivity treatment for the linear regression model. To develop the model, we first review some received results on unobserved heterogeneity in the Poisson regression model for count data. The model is then extended to encompass an endogenous sample selection mechanism. Previous papers have developed sequential, single equation, limited information estimation techniques. This paper presents a full information maximum likelihood (FIML) estimator for the model. Two techniques for computation of the sort of log-likelihood we analyze are described, simulation and numerical quadrature. An application to a problem in credit scoring is presented to illustrate the techniques.

Keywords: Count data; Poisson; Selectivity; FIML; Unobserved heterogeneity; Quadrature; Credit scoring

JEL classifications: C12; C13; Cross Section Econometrics

1. Introduction

The econometric issue of *sample selection* concerns the possible biases that arise when a nonrandomly sampled set of observations from a population is used as if the sample were random to make inferences about that population. Current literature, with a few exceptions noted below, has focused on, and finely tuned, the known results relating to this issue in the framework of the linear regression model and analysis of a continuous dependent variable, such as hours worked or wages. This paper will examine an extension of the sample selection model to the Poisson regression model for discrete, count data,

¹ Address for correspondence, Department of Economics, Stern School of Business, New York University, 44 West 4th St. New York, NY, 10012; wgreene@stern.nyu.edu. Helpful comments of Rainer Winkelmann and participants in department seminars at Amherst College, Washington University, and the University of Umea are gratefully acknowledged.

2. Models for Selection and The Poisson Regression Model

2.1. Sample Selection in the Linear Regression Model

Models for sample selection have become a standard body of technique in econometrics.³ The linear regression framework which forms the core of the technique is formulated as follows:⁴ A classical *normal* linear regression model is specified as

$$y_i = \beta' \mathbf{x}_i + \varepsilon_i, \varepsilon_i \sim N[0, \sigma^2].$$

The variables in the model are observed only when a threshold variable, z_i , equals 1;

$$(1) \quad \begin{aligned} z_i^* &= \alpha' \mathbf{w}_i + u_i, u_i \sim N[0, 1], \\ z_i &= \text{sgn}(z_i^*). \end{aligned}$$

When u_i , the unobserved effect in the observation mechanism, is correlated with ε_i , the unobserved individual heterogeneity in the regression model, then $E[y_i | \mathbf{x}_i, z_i=1]$ is not equal to $\beta' \mathbf{x}_i$, and the widely cited problems of “selection bias” in linear least squares regression arise. Linear regression of y_i on \mathbf{x}_i in the selected subpopulation with $z_i = 1$ estimates not β , but a hash of β and a nonlinear function of α and the moments of the variables in \mathbf{w}_i . Interest then centers on more detailed formulations of the inconsistency and on alternative, consistent estimation techniques.

Heckman's (1979) estimator for the linear model is a two step procedure based on the result that

$$(2) \quad E[y_i | \mathbf{x}_i, z_i=1] = \beta' \mathbf{x}_i + E[\varepsilon_i | z_i = 1] = \beta' \mathbf{x}_i + \theta M_i,$$

where $M_i = \phi(\alpha' \mathbf{w}_i) / \Phi(\alpha' \mathbf{w}_i)$, $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, $\theta = \rho \sigma$, and $\rho = \text{Corr}(\varepsilon, u)$.⁵ The two steps are (1) probit estimation of α in the model in (1) followed by computation of M_i for all observations for which z_i equals 1, then (2) linear regression of y_i on \mathbf{x}_i and M_i to estimate (β, θ) in (2) followed by an adjustment of the estimated asymptotic covariance matrix for the estimates which accounts for the use of the estimated regressor.

Although used somewhat less frequently, the technique of full information maximum likelihood of $(\beta, \alpha, \rho, \sigma)$ can also be employed based on the joint distribution of observations $(z_i = 0, \mathbf{w}_i)$ and $(z_i=1, \mathbf{w}_i, y_i, \mathbf{x}_i)$. (See Greene (1997b).) One noteworthy

³ Most of the current literature is based on Heckman's pioneering work (1979). Although a spate of recent literature, e.g. Manski (1990) and Newey et al. (1990) has questioned the implications of the fully parametrized nature of the model, the applied literature remains dominated by Heckman's and “Heckman-like” (e.g., Lee (1983) formulations. Nonparametric and semiparametric approaches are outside the scope of this paper.

⁴ We will only sketch the model formulation. There are more detailed treatments in many standard references, such as Heckman (1979) and Greene (1997a, Chapter 20).

⁵ The use of the symbol λ_i rather than M_i is more familiar in the literature, but this would conflict with another standard notation that will appear in the discussion to follow.

1 and variance α .⁶ When $P(y_i|\varepsilon_i)$ is Poisson with mean $\lambda_i(\varepsilon_i)$, we can find the unconditional distribution by integrating ε_i out of the conditional distribution. The now standard result is $P(y_i) = E_\varepsilon P(y_i|\varepsilon_i) =$ a negative binomial variate. (The full derivation appears in several references, including Cameron and Trivedi (1986) and Greene (1997a, pp. 939-940).) The resulting negative binomial has provided a mainstay in this literature.

A shortcoming for our purposes is that the negative binomial model does not lend itself to the sorts of extensions that will allow for a model of sample selection. In the same fashion as other similar applications (e.g., Winkelmann (1997), Crepon and Duguet (1997)), we respecify the model with lognormal instead of log-gamma heterogeneity. That is, $\varepsilon_i \sim N[0, \sigma^2]$; $f(\varepsilon_i) = (1/\sigma)\phi(\varepsilon_i/\sigma)$.⁷ The exact distribution of y_i after integrating out the heterogeneity will now be unknown, but that, in itself, is not an obstacle.

The conditional probability distribution is

$$P(y_i|\varepsilon_i) = \exp[-\lambda_i(\varepsilon_i)]\lambda_i(\varepsilon_i)^{y_i} / y_i!$$

The unconditional probability distribution is

$$P(y_i) = \int_\varepsilon P(y_i|\varepsilon_i) (2\pi\sigma^2)^{-1/2} \exp[-1/2 (\varepsilon_i/\sigma)^2] d\varepsilon_i.$$

Let $v_i = \varepsilon_i/(\sigma\sqrt{2})$, $\theta = \sigma\sqrt{2}$, and $\lambda_i(v_i) = \exp(\beta'\mathbf{x}_i + \theta v_i)$. With the change of variable,

$$P(y_i) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left[\exp(-\lambda_i(v_i))\lambda_i(v_i)^{y_i} / y_i! \right] e^{-v_i^2} dv_i,$$

The integral has no closed form but can be closely approximated by using Hermite quadrature for the integration

$$P^*(y_i) = \frac{1}{\sqrt{\pi}} \sum_{h=1}^H w_h \left[\exp[-\lambda_i(v_h)]\lambda_i(v_h)^{y_i} / y_i! \right] \approx P(y_i).$$

For our applications, we have found that a 20 point integration provides a sufficiently accurate approximation.⁸ The approximation to the log-likelihood is, then,

$$\log-L^* = \sum_i \log P^*(y_i) \approx \log-L.$$

Optimization and computation of the BHHH estimator of the asymptotic covariance matrix for the estimates will use the approximation to the first derivatives vector,

⁶ For an innovative alternative approach for panel data modelling, see Nagin and Land (1993).

⁷ Yet another approach is suggested by Gourieroux and Visser (1997).

⁸ Tables of the weights, w_h and nodes, v_h for the Hermite quadrature can be found in Abramovitz and Stegun (1971).

$$\text{Var}[y_i|\mathbf{x}_i] = E[y_i|\mathbf{x}_i] \{1 + \kappa E[y_i|\mathbf{x}_i]\}.$$

The lognormal model has a similar characteristic. Conditioned on ε_i , $y_i|\mathbf{x}_i, \varepsilon_i$ is Poisson. Let $\lambda_i(\varepsilon_i) = \exp(\beta' \mathbf{x}_i + \varepsilon_i) = E[y_i|\mathbf{x}_i, \varepsilon_i]$. Then, using properties of the lognormal distribution,

$$E[y_i|\mathbf{x}_i] = E_\varepsilon[E[y_i|\mathbf{x}_i, \varepsilon_i]] = \lambda_i^* = \exp(\beta' \mathbf{x}_i + \sigma^2/2).$$

Since $\text{Var}[y_i|\mathbf{x}_i, \varepsilon_i] = E[y_i|\mathbf{x}_i, \varepsilon_i]$, it follows from a bit of algebra that

$$\text{Var}[y_i|\mathbf{x}_i] = \text{Var}[E[y_i|\mathbf{x}_i, \varepsilon_i]] + E[\text{Var}[y_i|\mathbf{x}_i, \varepsilon_i]] = \lambda_i^* \{1 + [\exp(\sigma^2) - 1] \lambda_i^*\} = \lambda_i^* (1 + \omega \lambda_i^*)$$

Likewise, if the dispersion of the heterogeneity distribution (σ) goes to zero, we revert to the Poisson model.

The conditional mean functions in the heterogeneity models are $E[y_i|\mathbf{x}_i] = \exp(\beta' \mathbf{x}_i)$ for the negative binomial model and $\exp(\beta' \mathbf{x}_i + \frac{1}{2}\sigma^2)$ for the lognormal model. In both cases, the marginal effects are

$$\delta_i = \partial E[y_i|\mathbf{x}_i] / \partial \mathbf{x}_i = E[y_i|\mathbf{x}_i] \times \beta.$$

Estimation of the effects can be done at the sample means. Standard errors for the effects can easily be obtained with the delta method. We do note, because of the particular form of the conditional mean function, rather different estimates for the marginal effects are likely to be obtained in a small sample if they are computed, instead, by evaluating the effects, themselves, at each observations, then averaging the sample values.

2.4. Sample Selectivity in the Poisson Model: 2 Step Approaches

We now build a selection model upon the heterogeneity model. Consistent with standard applications, suppose that the primary model and observation mechanism are

$$P(y_i|\varepsilon_i) = \text{Poisson}[\lambda_i(\varepsilon_i)] = \exp[-\lambda_i(\varepsilon_i)] \lambda_i(\varepsilon_i)^{y_i} / y_i!$$

$$z_i^* = \alpha' \mathbf{w}_i + u_i, \quad u_i \sim N[0, 1]$$

$$z_i = \text{sgn}(z_i^*)$$

$$[\varepsilon_i, u_i] \sim N_2[(0, 0), (\sigma^2, \rho\sigma, 1)]$$

$$(y_i, \mathbf{x}_i) \text{ observed iff } z_i = 1.$$

Thus, the modelling framework is the same as Heckman's as specified in Section 2.1. What remains is to construct an appropriate estimation technique.

Greene (1994, 1997b) suggests a direct analog to Heckman's, two step procedure:¹¹

¹¹ The technique was applied in Freund, et al. (1997).

$$\mathbf{H} = \sum_i e_i^2 \mathbf{x}_i^0 \mathbf{x}_i^{0'}$$

\mathbf{V}_α is the estimated asymptotic covariance matrix of the probit estimates, $\hat{\alpha}$, and \mathbf{G} is the sum of cross products of \mathbf{x}_i^0 and

$$\mathbf{w}_i^0 = \partial E[y_i|\mathbf{x}_i]/\partial \alpha = E[y_i|\mathbf{x}_i] \left(c_i - \frac{\phi(a_i)}{\Phi(a_i)} \right) \mathbf{w}_i$$

Let $Q_i(\theta) = \log \left(\frac{\Phi(a_i + \theta)}{\Phi(a_i)} \right)$. In Terza's formulation, $E[y_i|\mathbf{x}_i] = \exp[\beta_i' \mathbf{x}_i + Q_i(\theta)]$.

Expand the function $Q_i(\theta)$ in a linear Taylor series around the point $\theta = \rho\sigma = 0$ (or, $\rho = 0$, since σ is positive by construction). The result is $Q_i(\theta) \approx \theta M_i$ where $M_i = \phi(a_i)/\Phi(a_i)$ as defined earlier. Thus, Greene's (1994) formulation could be viewed as this approximation to Terza's model.¹⁵ This suggests another two step approach: First, as usual, estimate the probit model by maximum likelihood, then compute \hat{M}_i as before. The second step consists of nonlinear least squares, where now the conditional mean function is $E^0(y_i|\mathbf{x}_i) = \exp(\beta_i' \mathbf{x}_i + \theta M_i)$. As before, it is necessary to adjust the estimated asymptotic covariance matrix of the estimator of (β, θ) . The end result is a minor modification of Terza's results:

$$\mathbf{x}_i^0 = \partial E^0[y_i|\mathbf{x}_i]/\partial \begin{pmatrix} \beta \\ \theta \end{pmatrix} = E^0[y_i|\mathbf{x}_i] \begin{pmatrix} \mathbf{x}_i \\ M_i \end{pmatrix},$$

and

$$\mathbf{w}_i^0 = \partial E^0[y_i|\mathbf{x}_i]/\partial \alpha = E^0[y_i|\mathbf{x}_i] \{ \theta M_i (-a_i - M_i) \}$$

with other calculations the same as before.¹⁶

Marginal effects in the selection model can be obtained from the conditional mean given earlier. By simple differentiation, we obtain

$$\partial E[y_i|\mathbf{x}_i, \mathbf{w}_i, z_i=1]/\partial \mathbf{x}_i = E[y_i|\mathbf{x}_i, \mathbf{w}_i, z_i=1] \times \beta,$$

$$\partial E[y_i|\mathbf{x}_i, \mathbf{w}_i, z_i=1]/\partial \mathbf{w}_i = E[y_i|\mathbf{x}_i, \mathbf{w}_i, z_i=1] \times \alpha \times (c_i - a_i).$$

When \mathbf{x}_i and \mathbf{w}_i have variables in common, the effects are added in the marginal effect, with the first part constituting the direct effect on the conditional mean and the second part constituting the indirect effect on the probability of selection into the sample.

¹⁵ Since the conditional mean is an approximation, it remains to be shown that the nonlinear least squares estimator based on the approximation is consistent for the same parameters as is that based on the true function. Kmenta's (1967) approximation to the CES production function is an early application that has the same characteristics.

¹⁶ The computations will be straightforward with most current econometrics computer packages. A short LIMDEP program to do them is given in the appendix. Gauss code for this model would be likewise brief.

When $z_i = 0$, only (z_i, \mathbf{w}_i) are observed. The contribution to the likelihood function is

$$\text{Prob}[z_i = 0 | \mathbf{w}_i] = E_{\varepsilon}[1 - \text{Prob}[u_i > -\alpha' \mathbf{w}_i | \mathbf{w}_i, \varepsilon_i]] = E_{\varepsilon}[\text{Prob}[u_i \leq -\alpha' \mathbf{w}_i | \mathbf{w}_i, \varepsilon_i]].$$

This provides the probability needed to construct the likelihood function.

$$\text{Prob}[z_i = 0 | \mathbf{w}_i, \varepsilon_i] = 1 - \Phi[\gamma' \mathbf{w}_i + \tau \varepsilon_i / (\sqrt{2} \sigma)]$$

so
$$\text{Prob}[z_i = 0 | \mathbf{w}_i] = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-v^2) \Phi[-(\gamma' \mathbf{w}_i + \tau v)] dv.$$

As before, quadrature or simulation is used to evaluate the integral.

Maximum likelihood estimates of $[\beta, \gamma, \theta, \tau]$ are obtained by maximizing

$$\log L = \sum_{z=0} \log \text{Prob}[z_i = 0 | \mathbf{w}_i] + \sum_{z=1} \log P[y_i, z_i = 1 | \mathbf{x}_i, \mathbf{w}_i].$$

The log-likelihood and its derivatives are obtained as follows: For observation i , if $z_i = 0$,

$$\log L_i | z_i = 0 = \log \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-v^2) \Phi[-(\gamma' \mathbf{w}_i + \tau v)] dv = \log T_i,$$

$$\partial \log L_i | z_i = 0 / \partial \begin{pmatrix} \gamma \\ \tau \end{pmatrix} = -(1/T_i) \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-v^2) \phi[-(\gamma' \mathbf{w}_i + \tau v)] \begin{pmatrix} \mathbf{w}_i \\ v \end{pmatrix} dv,$$

and

$$\partial \log L_i | z_i = 0 / \partial \begin{pmatrix} \beta \\ \theta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

If $z_i = 1$,

$$\log L_i | z_i = 1 = \log \left[\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-v^2) \frac{\exp[-\lambda_i(v)] \lambda_i(v)^{y_i}}{y_i!} \Phi(\gamma' \mathbf{w}_i + \tau v) dv \right]$$

$$= \log \left[\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-v^2) H(\beta' \mathbf{x}_i + \theta v) \Phi(\gamma' \mathbf{w}_i + \tau v) dv \right]$$

$$= \log P_i,$$

and

$$\partial \log P_i / \partial \begin{pmatrix} \beta \\ \theta \end{pmatrix} = (1/P_i) \left[\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-v^2) [\partial H(b_i) / \partial \begin{pmatrix} \beta \\ \theta \end{pmatrix}] \Phi(\gamma' \mathbf{w}_i + \tau v) dv \right], \quad b_i = \beta' \mathbf{x}_i + \theta v.$$

But,
$$\partial H(b_i) / \partial \begin{pmatrix} \beta \\ \theta \end{pmatrix} = H(b_i) \partial \log H(b_i) / \partial \begin{pmatrix} \beta \\ \theta \end{pmatrix} = H(b_i) (y_i - \lambda_i(v)) \begin{pmatrix} \mathbf{x}_i \\ v \end{pmatrix}.$$

The same logic and construction gives

$$P(y_i|\mathbf{x}_i, \mathbf{w}_i, \varepsilon_i) = \text{Poisson with } \lambda_i = \exp(\beta' \mathbf{x}_i + \varepsilon_i) \Phi \left(\frac{1}{\sqrt{1-\rho^2}} (\alpha' \mathbf{w}_i + (\rho/\sigma) \varepsilon_i) \right)$$

Estimation of $(\beta, \alpha, \rho, \sigma)$ is based on the likelihood function for the observed data, which is

$$\log-L = \sum_i \log P(y_i|\mathbf{x}_i, \mathbf{w}_i) = \sum_i \log E_\varepsilon P(y_i|\mathbf{x}_i, \mathbf{w}_i, \varepsilon_i)$$

Once again, the techniques are the same.

3. An Application

Greene (1992) examines a model of sample selection in the setting of a credit application model. The variable of primary interest in that study is the probability of default on a credit card loan in the first year of activity.¹⁷ The conditioning variable for the sample selection is acceptance of the individual's application for the credit card. (The model is largely similar to that in Boyes, Hoffman, and Low (1989). Thus, the default model is constructed to describe the probability that an individual would default on a loan *if they were given a credit card* (if they were given a loan), but is based on data for individuals to whom credit cards (loans) were already granted. Thus, there is a reasonable question as to the possibility of sample selectivity of the sort discussed earlier.

In passing, it is noted that an important predictor of whether a credit card application will be accepted is the number of "major derogatory reports," (MDRs) in an individual's credit reporting files at agencies such as TRW. An MDR is a sixty day delinquency in payment to a major credit account, such as one of the major bank cards or a major department store. At any point in time, most people have zero MDRs in their files. Observed values usually range from zero to three or four, but are sometimes much higher; the largest value in our sample was 14. In this study, we view MDRs, which is clearly a candidate for a count data model, as the behavioral variable of interest. The data analyzed in Greene (1992) are a sample of 13,777 observations on applications and account activity for a major credit card vendor. Of the 13,777 applications represented in the sample, approximately 76% (a choice based sample) were accepted. The default behavior and expenditure patterns in the first twelve months of holding were observed for the cardholders in the sample. Thus, whether or not the individual has the credit card in question is the sample selection rule. To illustrate the techniques described above, we used a random subsample of 1,319 observations from the full sample, including 1,023 cardholders. The variables used in the study are described in Tables 1-3.¹⁸ A histogram of the outcome variable for the subgroups is listed in Table 3. The relationship between MDRs and application acceptance is strongly suggested by the data in Table 3. In fact, we do have observations on all variables for both cardholders and noncardholders. The sociodemographic data were obtained from the credit card applications, themselves. The

¹⁷ The identity of the vendor is, at their request, not revealed. Since the behavioral variable in this study pertains to the general finances of the individual, the credit card, itself, is immaterial.

¹⁸ The data set can be downloaded from <http://www.stern.nyu.edu/~wgreene/poisson.selection.dat>.

4. Conclusions

This paper has presented an estimator of a model for count data which extends the lognormal heterogeneity model which has appeared elsewhere in the literature. The lognormal model has proved useful in several settings, such as those in Section 2.6. Winkelmann (1996) suggests some others, and, given recent developments, further extensions such as to the random effects model for panel data (Greene (1997b)) are equally straightforward. The pessimism expressed in Cameron and Trivedi's recent survey (1996, pp. 305-306) is clearly unwarranted.

It is difficult to draw general conclusions from the single application. The different formulations of the model discussed here do present three consistent estimators of the parameters of the model, so large differences would be surprising. The fact that the selection itself is producing relatively little movement in the estimates may be an artifact of this data set, since our dependent variable is a crucial determinant of the selection variable. A more appropriate specification might depart from a probit model such as

$$\begin{aligned} z_i^* &= \alpha' \mathbf{w}_i + \gamma y_i, \quad u_i, u_i \sim N[0,1] \\ z_i &= \text{sgn}(z_i^*) \\ [\varepsilon_i, u_i] &\sim N_2[(0,0), (\sigma^2, \rho\sigma, 1)] \end{aligned}$$

But, is it is not possible to proceed from here to an internally consistent selection model in which z_i is the mechanism that determines whether y_i is observed. On the other hand, with a full set of observations on all variables, such as we do have here, estimating the heterogeneity model

$$P(y_i|\varepsilon_i) = \text{Poisson}[\lambda_i(\varepsilon_i)] = \exp[-\lambda_i(\varepsilon_i)] \lambda_i(\varepsilon_i)^{y_i} / y_i!$$

and this binary choice model jointly would be straightforward. The likelihood function would be built up from the joint probabilities

$$P(z_i, y_i | \mathbf{x}_i, \mathbf{w}_i) = E_\varepsilon P(z_i = j | y_i, \mathbf{w}_i, \varepsilon_i) P(y_i | \mathbf{x}_i, \varepsilon_i), \quad j = 0, 1$$

using exactly the methods we considered earlier.

FIML estimation of the selection model is quite simple. Its advantages over the two step procedures would stem from the asymptotic efficiency of a joint estimator, which is inherent, and from the approximation in Greene's estimator. It could be argued that the nonlinear least squares estimators are more robust to misspecification, as they require only the specification of the conditional mean function. But, the extent of this advantage seems speculative.

Table 3. Counts of MDRs for All Observations and Cardholders

MDRs	All	Card Holders
0	1060	915
1	137	90
2	50	13
3	24	4
4	17	1
5	11	0
6	5	0
7	6	0
8	0	0
9	2	0
10	1	0
11	4	0
12	1	0
13	0	0
14	1	0

Table 6. Estimated Poisson Model Ignoring Selectivity: Cardholders Only.

Log likelihood function -407.9441

Variable	Coefficient Estimates		Marginal Effects	
	Coefficient	t-ratio	Coefficient	t-ratio
Constant	-3.615542	-8.574		
AGE	.1880018E-01	2.154	.2425830E-02	1.592
INCOME	.1341672	2.470	.1731189E-01	1.741
EXP_INC	1.985568	1.570	.2562023	1.262
AVGEXP	.4826625E-04	.122	.6227904E-05	.108
MAJOR	.2416640	.900	.3118245E-01	.768

Table 7. Estimated Probit Model for Sample Inclusion (Cardholder Status)

Variable	Single Equation ML		FIML Estimated with Poisson Model	
	Coefficient	t-ratio	Coefficient	t-ratio
Constant	0.542	2.95	0.305	2.191
AGE	-0.0086	-1.72	-0.0039	-1.26
INCOME	0.092	1.73	0.0523	1.62
MAJOR	0.212	2.06	0.114	1.65
OWNRENT	0.349	3.46	0.199	2.33
DEPDNT	-0.131	1.90	-0.0726	-1.62
INC_PER	-0.015	-0.21	-0.144	-0.39
SELFEMPL	-0.201	-1.23	-0.121	-1.24
OPEN	-0.286	-11.67	0.165	3.08
CURADD	-0.0004	-0.58	-0.0004	-1.03
ACTIVE	-0.230	-10.75	-0.136	-3.06

Table 8. Estimates of Selection Models

Variable	Greene, NLSQ		Terza NLSQ		FIML	
	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio
Constant	-5.345	-7.22	-4.068	-6.83	-4.700	-8.16
Age	0.0128	1.16	0.0142	1.34	0.0170	1.45
Income	0.191	3.20	0.136	2.32	0.161	2.02
Exp. Inc.	1.775	1.88	1.734	1.61	1.718	0.75
Avg. Exp.	-0.0000268	-0.09	-0.0000362	-.09	0.0000179	-0.09
Major	1.376	2.33	0.811	1.65	0.333	1.03
M_i	1.969	6.72				
θ			3.365	0.11		
ρ					0.966	7.29
σ					1.268	6.09
$e'e$	165.319		168.262		177.183	

Appendix. LIMDEP Computations for the Sample Selection Estimators

```

/* Computation of estimators and appropriate asymptotic covariance
   matrices for Terza and Greene nonlinear least squares estimators.
   The routine is general - different applications change only the
   namelist definitions and the variable names given to Y and Z.
*/
? Define lists of variables used in the computations.
Namelist ; W=one,age,income,major,ownrent,depndt,inc_per,selfempl,
          accounts,cur_add,active
          ; X=one,age,income,exp_inc,avgexp,major$
? LHS variables in regression and probit model.
Create   ; Y = Majordrg
          ; Z = Cardhldr $
? Probit estimates. Mills ratio is kept for Greene estimator.
Probit   ; Lhs = Z ; Rhs=W ; Hold(IMR=Mi)$
? Retain estimators for later.
Matrix   ; Alpha = b ; Valpha = VARB $
? Uncorrected estimates, for starting values
Poisson  ; Lhs = Y ; Rhs = X $
Matrix   ; BPois = b $
? Heckman form of mean corrected Poisson
Poisson  ; Lhs = Y ; Rhs = X ; Selection $
? FIML estimator is internal:
Poisson  ; Lhs = Y ; Rhs = X ; Selection ; MLE $
? 2 Step estimators - covariance matrices must be constructed.
? Use selected subsample
Reject   ; Z = 0$
? AI appears in conditional mean function, uses first step estimates
Create   ; Ai=Alpha'W$
? Nonlinear Least Squares
Calc     ; K = Col(X) $ ? Number of variables in X.
NLSQ     ; Lhs = Y
          ; Fcn = exp(b1'x) * Phi(ai+t) / Phi(ai)
          ; Labels = K b,t
          ; Start = BPois,0 $
? (For Greene's estimator, change Fcn to exp(b1'x+t*mi))
Matrix   ; Beta = Part(b,1,K) $
Calc     ; Theta= b(kreg) $ (Kreg=#parameters, left by NLSQ)
Create   ; bi=beta'x
          ; ey=exp(bi)*phi(ai+theta)/phi(ai) ? conditional mean function
          ; gi=n01(ai+theta)/phi(ai+theta) ? ey*gi=dey/dtheta
          ; ui=y-ey ? residual
          ; pj=gi-mi ? dey/dalpha'z
          ; wb=ey*ey ? invS = <D'[wb]D>
          ; wh=ey*ey*ui*ui ? H = D'[wh]D
          ; wp=ey*ey*pj$ ? G = D'[wp]Z
? (For Greene's estimator, the only changes needed are
? ey to exp(bi + theta*mi)
? gi to gi=mi and pj to
? pj to -theta * mi * (ai+mi).
? Asymptotic covariance matrix, in two parts.
Namelist ; D=x,gi$
Matrix   ; V = <D'[wb]D> * D'[wh]D * <D'[wb]D>
          + <D'[wb]D> * D'[wp]Z * Valpha * Z'[wp]D * <D'[wb]D>
          ; Stat (B,V) $

```

- Maddala, G.S., *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, 1983.
- Manski, C., "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80, 2, 1990, pages 319-323.
- Murphy, K., and Topel, R., "Estimation and Inference in Two Step Econometric Models," *Journal of Business and Economic Statistics*, 3, 1985, pp. 370-379.
- Nagin, D., and Land, K., "Age, Criminal Careers, and Population Heterogeneity: Specification and Estimation of a Nonparametric, Mixed Poisson Model," *Criminology*, 31, 1993, pp. 327-362.
- Newey, W., Powell, J., and Walker, J., "Semiparametric Estimation of Selection Models: Some Empirical Results," *American Economic Review*, 80, 2, 1990, pages 324-328.
- McFadden, D., and Ruud, P., "Estimation by Simulation," *Review of Economics and Statistics*, 76, 1994, pp. 591-608.
- Smith, V., "Selection and Recreation Demand," *American Journal of Agricultural Economics*, 70, 1988, pp. 29-36.
- Terza, J., "Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects," Working Paper No. 4-95-15, Department of Economics, Penn State University, 1995.
- Vuong, Q., "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 1989, pp. 307-334.
- White, H., "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica*, 48, 1980, pp. 817-838.
- Winkelmann, R., "Count Data Models with Selectivity," Discussion Paper No. 9603, Department of Economics, University of Canterbury, New Zealand, 1996.
- Winkelmann, R., *Econometric Analysis of Count Data*, 2nd. edition, Berlin, Springer Verlag, 1997.