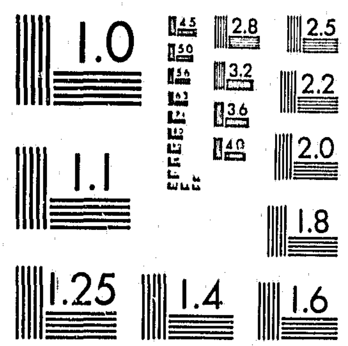

Presented at the 1980 meetings of the American Statistical Association

CORRECTING SAMPLE SELECTION BIAS FOR BIVARIATE LOGISTIC DISTRIBUTION OF DISTURBANCES

Subhash C. Ray
Department of Economics
University of California at Santa Barbara

Richard A. Berk
Department of Sociology
University of California at Santa Barbara

William T. Bielby
Department of Sociology
University of California at Santa Barbara

March 1981

The authors wish to acknowledge the National Institute of Justice for support of this project.

CORRECTING SAMPLE SELECTION BIAS

FOR BIVARIATE LOGISTIC DISTRIBUTION OF DISTURBANCES

Subhash C. Ray, Richard A. Berk and William T. Bielby

ABSTRACT

Notwithstanding Tobin's famous paper in 1958, the problem of censored or truncated samples was largely ignored in standard econometric research during the 1960's. In the past decade, however, Amemiya, Heckman and others have studied in detail the properties of OLS estimators obtained from the non-randomly selected subsample. This paper extends their analyses to the case where the random disturbances have a bivariate logistic distribution instead of bivariate normal.

CORRECTING SAMPLE SELECTION BIAS

FOR BIVARIATE LOGISTIC DISTRIBUTION OF DISTURBANCES

Despite Tobin's famous paper on limited dependent variables written in 1958, the 1960's found very few attempts in standard econometric literature to address the methodological questions associated with censoring or truncation. In the past decade, however, Amemiya (1973), Heckman (1976, 1979), Olsen (1980) and others have considered in depth the inconsistency associated with OLS estimators obtained from a selective sample and have suggested alternative, consistent estimation procedures. This paper is an extension of Heckman's perspective to the case in which the joint distribution between random disturbances in the specified (substantive) equation and random disturbances in the sample selection equation is bivariate logistic (rather than bivariate normal). Indeed Heckman (1978) briefly considered the possibility of using a multivariate logistic distribution in place of the usual multivariate normal (1978, p. 953), but recognized that a full development of the logistic approach would require additional work.

Recall that sample selection bias problem begins with the following substantively derived equation:

$$y_{1i} = x_{1i}\beta_1 + u_{1i} \quad . \quad . \quad . \qquad (1)$$

where y_{1i} is the observation on the dependent variable for individual i. x_{1i} is the vector of observations on the explanatory variables for the individual and u_{1i} is a random disturbance. A typical and perhaps the most frequently cited example is one in which y_1 is the observed market wage for an individual and x_1 is the vector of relevant characteristics like education, experience, and the like.

Consider now a second sample-selection equation

$$y_{2j} = x_{2j}\beta_2 + u_{2j} \quad . \ . \ . \qquad (2)$$

where y_{2j} is the dependent variable observed for individual j, x_{2j} is the vector of explanatory variables, and u_{2j} is a random disturbance. The critical point is that for some individual s we shall have observations y_{1s}, x_{1s} only if $y_{2s} > 0$ or otherwise exceeds some arbitrary threshold. An individual's market wage, for instance, is observable only when that individual obtains employment. For persons whose reservation wage exceeds the market wage, no wage data are available. Similarly in Tobin's framework, an individual would be recorded as expressing positive demand for a consumer durable (say, an automobile) only when his/her utility maximization results in a solution that assigns a value in excess of zero to the consumption level of that good.

Given a sample of N individuals initially, it may be necessary to exclude all but N_1 cases for whom one has observations on y_{1i} and x_{1i}. These are the instances of which $x_{2i}\beta_2 + u_{2i} > 0$. Consequently, Equation (2) in our model represents a sample selection "rule." With respect to OLS estimators for parameters of Equation (1) for the subsample of N_1 individuals, Heckman points out: "The bias that arises from using least squares to fit models for limited dependent variables or models with censoring or truncation arises solely because the conditional mean of u_{1i} is not included as a regressor" (1979:478). The bias thus results from the omitted variable type of specification error.

In this paper we will apply the basic Heckman correction procedure assuming that the random disturbances have bivariate logistic distributions.

Specifically, we consider two alternative formulations suggested by Gumbel (1961). Each of these two distribution functions yields logistic marginals. However, the implied estimation procedures are quite different. In one instance, the corrected model yields homoscedastic errors. In the other, although heteroscedasticity surfaces, the variance function is parametrically known and it is possible to search for the optimal value (in terms of lowest SSE) within the admissible range of the parameter.

In Section I, the two different bivariate distribution functions are briefly outlined. Sections II and III present consistent estimators of the parameters of Equation (1) using each of the distribution functions. At this stage we make a choice and opt for the estimator that is implied by the Gumbel type II distribution. Section IV shows that the estimator derived is consistent and that the distribution of the estimates is asymptotically normal so that despite non-normality of equation disturbances, the standard statistical tests apply in large samples. Finally in Section V, simulation results are reported comparing the instrument used by Heckman in his probit model and the one implied by our logistic model. Also, we provide in this section an application drawn from a study of the employment experiences of ex-offenders after their release from prison (Rossi, Berk and Lenihan, 1980).

I. Two Bivariate Distribution Functions

To begin, consider the distribution function for two random variables (ω_1, ω_2) of the form:

$$\Pr[\omega_1 \le t_1,\ \omega_2 \le t_2] = F(\omega_1,\ \omega_2) = \frac{1}{1 + e^{-t_1} + e^{-t_2}} \qquad (3)$$

The random variables ω_1 and ω_2 are continuous and can assume any value on the real line. The marginal distribution functions are

$$F_1(t_1) = \Pr[\omega_1 \leq t_1,\ \omega_2 \leq \infty] = \frac{1}{1 + e^{-t_1}} \tag{4}$$

and

$$F_2(t_2) = \Pr[\omega_1 \leq \infty,\ \omega_2 \leq t_2] = \frac{1}{1 + e^{-t_2}} \tag{5}$$

which are themselves logistic.

The marginal densities are

$$f_1(t_1) = \frac{e^{-t_1}}{\left(1 + e^{-t_1}\right)^2}, \quad f_2(t_2) = \frac{e^{-t_2}}{\left(1 + e^{-t_2}\right)^2} \tag{6}$$

and joint density is

$$h(t_1,\ t_2) = \frac{\partial^2 F}{\partial t_1 \partial t_2} = \frac{2e^{-(t_1+t_2)}}{\left(1 + e^{-t_1} + e^{-t_2}\right)^3} \tag{7}$$

We call this distribution Gumbel's type I. It may be noted in passing that Nerlove and Press (1973) used this form of the multivariate logistic distribution.

Gumbel has derived the conditional mean and variance from the conditional moment generating function $G_{\omega_1}(t_1|\omega_2)$.

From the conditional density function

$$f(\omega_1|\omega_2) = \frac{2e^{-\omega_1}\left(1 + e^{-\omega_2}\right)^2}{\left(1 + e^{-\omega_1} + e^{-\omega_2}\right)^3}, \tag{8}$$

the conditional moment generating function is derived as

$$= F_2(\omega_2)^{t_1}\ \Gamma(2 + t_1)\ \Gamma(1 - t_1). \tag{9}$$

Using this conditional generating function Gumbel further derives the conditional mean

$$E(\omega_1|\omega_2) = 1 + \ln F_2(\omega_2) \tag{10}$$

and the conditional variance

$$\sigma^2(\omega_1|\omega_2) = \frac{\pi^2}{3} - 1. \tag{11}$$

The distribution function considered above is one amongst several that yield logistic marginals. We now consider Gumbel's type II which can be written as

$$F(\omega_1,\ \omega_2) = F(\omega_1)F(\omega_2)[1 + \alpha(1 - F(\omega_1))(1 - F(\omega_2))] \quad \text{where } -1 < \alpha \leq 1. \tag{12}$$

Note that as before

$$F(\omega_1,\ \infty) = F(\omega_1) = \frac{1}{1 + e^{-\omega_1}}$$

and

$$F(\infty,\ \omega_2) = F(\omega_2) = \frac{1}{1 + e^{-\omega_2}}$$

which are by assumption logistic marginals. This distribution function is more flexible than the type I. In the particular case where $\alpha = 0$, $F(\omega_1,\ \omega_2) = F(\omega_1)F(\omega_2)$ which indicates independence.

The joint density function for Gumbel's type II is

$$f(\omega_1, \omega_2) = \frac{e^{-(\omega_1+\omega_2)}}{\left(1 + e^{-\omega_1}\right)^2\left(1 + e^{-\omega_2}\right)^2}\left[1 + \alpha\left(\frac{1 - e^{-\omega_1}}{1 + e^{-\omega_1}}\right)\left(\frac{1 - e^{-\omega_2}}{1 + e^{-\omega_2}}\right)\right]. \qquad (13)$$

Ord (1972) reports that the conditional moment generating function is

$$G_{\omega_1}(t_1|\omega_2) = \frac{\pi t_1(1 + \beta t_1)}{\sin \pi t_1} \qquad (14)$$

where $\beta = 2\alpha F_2(\omega_2) - \alpha$. The conditional mean

$$E(\omega_1|\omega_2) = \beta = 2\alpha F_2(\omega_2) - \alpha \qquad (15)$$

and the conditional variance

$$\sigma^2(\omega_1|\omega_2) = \frac{\pi^2}{3} - \left[\alpha(2F_2(\omega_2) - 1)\right]^2. \qquad (16)$$

It should be remembered that for Gumbel type II distribution functions the coefficient of correlation between ω_1 and ω_2 is $\rho = 3\alpha/\pi^2$. Since $1 \leq \alpha \leq 1$, $|\rho| \leq .30396$. For the Nerlove-Press or Gumbel type I bivariate distribution, $\rho = 1/2$. We shall consider the implications of such restrictions in the context of our estimation procedures.

II. Consistent Estimators for Gumbel Type I Bivariate Error Distributions

Arguing exactly as Heckman does, we can write

$$\begin{aligned} E(y_1|x_1;\ \text{sample selection rule}) &= E(y_1|x_1;\ y_2 > 0) \\ &= E(y_1|x_1;\ u_2 > -x_2\beta_2) \\ &= x_1\beta_1 + E(u_1|u_2 > -x_2\beta_2). \end{aligned} \qquad (17)$$

Except in the case where u_1 and u_2 are independent, the conditional expectation $E(u_1|u_2 > -x_2\beta_2)$ will not be zero.

Assume at this stage that $u_1 = k_1\omega_1$ and $u_2 = k_2\omega_2$ where ω_1 and ω_2 have a Gumbel type I bivariate distribution function. In the special case, any $u_i = \omega_i$, $k_i = 1$. The reason why we introduce the scale factors k_1 and k_2 would become apparent below. For convenience assume $k_2 = 1$ in this case so that $u_2 = \omega_2$.

Since the conditional mean of ω_1 given ω_2 is $E(\omega_1|\omega_2) = 1 + \ell n\ F_2(\omega_2)$ we can write

$$\omega_1 = 1 + \ell n\ F_2(\omega_2) + \varepsilon \qquad (18)$$

where ε has a univariate logistic distribution with mean 0 and variance $\left(\frac{\pi^2}{3} - 1\right)$.

Now

$$\begin{aligned} E(u_1|u_2 > -x_2\beta_2) &= E(k_1\omega_1|\omega_2 > -x_2\beta_2) \\ &= k_1 E(\omega_1|\omega_2 > -x_2\beta_2). \end{aligned} \qquad (19)$$

Define $-x_2\beta_2 = z$. Then

$$E(u_1|u_2 > -x_2\beta_2) = k_1\Big[E(\omega_1|\omega_2 > z)\Big].$$

Now

$$E(\omega_1|\omega_2 > z) = E\left(1 + \ln\frac{1}{1+e^{-\omega_2}} + \varepsilon\,\middle|\,\omega_2 > z\right)$$

$$= 1 + E\left(\ln\frac{1}{1+e^{-\omega_2}}\,\middle|\,\omega_2 > z\right). \tag{20}$$

Consider in particular

$$E\left(\ln\frac{1}{1+e^{-\omega_2}}\,\middle|\,\omega_2 > z\right) = \frac{\int_z^\infty \ln\frac{1}{1+e^{-\omega_2}} f(\omega_2)\,d\omega_2}{1 - F_2(z)}$$

$$= \frac{\int_z^\infty \ln\left(\frac{1}{1+e^{-\omega_2}}\right)\cdot\frac{e^{-\omega_2}}{\left(1+e^{-\omega_2}\right)^2}\,d\omega_2}{1 - F_2(z)}. \tag{21}$$

Define

$$\frac{1}{1+e^{-\omega_2}} = y$$

then as ω_2 varies from z to ∞, y varies from $\frac{1}{1+e^{-z}}$ to 1. Also

$$dy = \frac{e^{-\omega_2}}{\left(1+e^{-\omega_2}\right)^2}\,d\omega_2.$$

Therefore (21) above can be written as

$$\frac{\left[\int_{\frac{1}{1+e^{-z}}}^{1} \ln y\,dy\right]}{1 - F_2(z)} = \left[\frac{1}{1 - F_2(z)}\right]\cdot\left[y\ln y\Big|_{\frac{1}{1+e^{-z}}}^{1} - \int_{\frac{1}{1+e^{-z}}}^{1} y\,d(\ln y)\right] \tag{22}$$

$$= -1 - \frac{1}{e^{-z}}\ln\frac{1}{1+e^{-z}}. \tag{23}$$

We can therefore write

$$E(u_1|u_2 > z) = k[E(\omega_1|\omega_2 > z)]$$

$$= k_1\left[1 + E\left(\ln\frac{1}{1+e^{-\omega_2}}\,\middle|\,\omega_2 > z\right)\right]$$

$$= k_1\left[1 - 1 - \frac{1}{e^{-z}}\ln\frac{1}{1+e^{-z}}\right]$$

$$= -k_1\cdot\frac{1}{e^{-z}}\ln\frac{1}{1+e^{-z}}. \tag{24}$$

Thus

$$E(y_1|x_1;\ SSR) = x_1\beta_1 - k_1\left(e^z\ \ln\frac{1}{1+e^{-z}}\right). \tag{25}$$

If we define z as the threshold $\frac{1}{1 + e^{-z}} = P$ which is the probability that an individual is censored

$$1 + e^{-z} = \frac{1}{P} \quad \text{or} \quad e^{-z} = \frac{1 - P}{P} \quad \text{and} \quad e^{z} = \frac{P}{1 - P}$$

which is the odds ratio.

We can therefore write

$$E(y_1 | x_1, \text{SSR}) = x_1\beta_1 - \frac{P}{1 - P} \ln(P) \cdot k_1. \qquad (26)$$

Consequently, one should in practice estimate a logistic regression for Equation (2) to find the value of $\hat{P}$ for each individual (case). Then for the subsample of N_1 individuals, one should construct the variable

$$\left(\frac{\hat{P}_i}{1 - \hat{P}_i}\right) \ln(\hat{P}_i) = \phi_i$$

(say) and include it as an additional explanatory variable. The regression coefficient of ϕ_i is our estimate of k_1. As we have already seen, the conditional variance of $\omega_1|\omega_2$ is $k_1^2(\frac{\pi^2}{3} - 1)$ so that one has homoscedastic errors. Note however that we have more than one estimate of k_1, and the uniqueness of the estimated value, may be regarded as test of the validity of the model.

Strictly speaking, since the substantive equation's error (ε) has a logistic distribution, one should not apply t or F tests. However, since the logistic (univariate) and normal distributions are very close to one another, the usual tests can be applied without serious inaccuracy.

It is useful to recall here that in the absence of a scale factor k_1, the variance of ε has a specific value equal to $\frac{\pi^2}{3} - 1$. However, introducing the scale factor k_1 permits any value of error variance and in this sense destandardizes the distribution. Note also that since k_1 serves only to change the value of σ_ε^2, homoscedasticity is strictly maintained.

Unfortunately, the Gumbel type I bivariate distribution presupposes a correlation between u_1 and u_2 equal to 1/2, clearly a limitation of this approach.[1]

III. Consistent Estimators for Gumbel Type II Bivariate Error Distributions

As in the earlier section we assumed $u_{1i} = k_1\omega_{1i}$ and $u_{2i} = \omega_{2i}$ and that marginal distribution functions for ω_1 and ω_2 were logistic. Now however, the bivariate distribution function is represented by (12).

As before

[1] It may be instructive to note that Heckman computes his additional regressors λ_i (which is the hazard rate) from his Probit equation. He states, "In samples in which selectivity problem is unimportant (i.e., sample selection rule ensures that all potential population observations are samples), λ_i becomes negligibly small so that least squares estimates of the coefficients of 1(a) have optimal properties" (p. 479). Yet his corrected Equation (13a) is

$$y_{1i} = x_{1i}\beta_1 + \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \lambda_i + v_{1i}.$$

Clearly sample selectivity cannot be tested within this model. When λ_i goes to zero the model is not of full rank and is no longer estimable.

$$E(y_1|x_1;\ SSR) = x_1\beta_1 + E(u_1|u_2 > -x_2\beta_2)$$
$$= x_1\beta_1 + k_1E(\omega_1|\omega_2 > z).$$

Now however

$$E(\omega_1|\omega_2 > z) = E(2\alpha F_2(\omega_2) - \alpha + \mu \mid \omega_2 > z) \qquad (27)$$

by virtue of (15) and $E(\mu) = 0$ by assumption.

We can write

$$E(\omega_1|\omega_2 > z) = \frac{\left[\int_z^\infty \alpha(2F_2(\omega_2) - 1)f(\omega_2)\ d\omega_2\right]}{[1 - F_2(z)]}$$

$$= \frac{2\alpha\int_z^\infty F_2(\omega_2)f(\omega_2)\ d\omega_2}{1 - F_2(z)} - \alpha\ . \qquad (28)$$

In particular

$$\int_z^\infty F_2(\omega_2)f(\omega_2)\ d\omega_2 = \int_z^\infty \left(\frac{1}{1 + e^{-\omega_2}}\right)\frac{e^{-\omega_2}}{\left(1 + e^{-\omega_2}\right)^2}\ d\omega_2$$

$$= \int_z^\infty \frac{e^{-\omega_2}}{\left(1 + e^{-\omega_2}\right)^3}\ d\omega_2\ . \qquad (29)$$

Define $\theta = 1 + e^{-\omega_2}$ so that $d\theta = -e^{-\omega_2}\ d\omega_2$. (29) can then be expressed as

$$\int_{1+e^{-z}}^{1} \left(-\frac{1}{\theta^3}\right) d\theta = \left[\frac{1}{2\theta^2}\right]_{1+e^{-z}}^{1}$$

$$= \frac{1}{2}\left[1 - \frac{1}{(1 + e^{-z})^2}\right]. \qquad (30)$$

If we replace $\frac{1}{1 + e^{-z}}$ by P we can express (30) as

$$\frac{1}{2}(1 - P^2). \qquad (31)$$

Going back to (28) and replacing $F_2(z)$ in the denominator by P we obtain

$$E(\omega_1|\omega_2 > z) = 2\alpha\cdot\frac{1}{2}\frac{(1 - P^2)}{(1 - P)} - \alpha$$
$$= \alpha(1 + P) - \alpha$$
$$= \alpha P$$

(ruling aside the case $P = 1$ where there is complete censoring).

We therefore can express

$$E(y_1|x_1;\ SSR) = x_1\beta_1 + k_1\alpha P \qquad (32)$$

and the underlying stochastic formulation should be $y_1 = x_1\beta_1 + k_1\alpha P + \eta$ where η is k_1 times μ.

What we require then is an estimate of $\hat{P}$ from a logistic regression model for the full sample to insert as an additional regressor in Equation (1). The linear probability model will typically be a good approximation for the logistic regression. In this context, the coefficient of $\hat{P}$ in the reformulated model (32) has a simple interpretation. It is the rate at which y_{1i}

increases as the probability that an individual i would not cross the threshold increases.[2]

For Gumbel type II distributions, the coefficient of correlation between u_1 and u_2 is within $\pm .304$. Yet, this restriction may not be very serious in practice. Indeed, as reported in the next section, simulation experiments indicate that the variables ϕ_i (for Gumbel type I), $\hat{P}_i$ (for type II) and Heckman's λ_i are highly correlated $(r > .9)$. It does not therefore appear to make much difference whichever instrument one uses.

A far more serious problem for Gumbel type II error distribution is that unlike in type I case, we have heteroscedasticity. However the variance of η_i is known from (16) and

[2]The "hazard rate" which enters Heckman's model as an additional regressor is $\frac{\phi(i)}{1-\Phi(i)}$ where ϕ is the density and $1-\Phi$ is the complement of the cumulative probability for a normal distribution. It is interesting to compute this "hazard rate" using a univariate logistic instead of the normal. If logistic density and cumulative are denoted by f and F, the "misconstrued" hazard rate is $\frac{f(i)}{1-F(i)}$ where

$$F(i) = \frac{1}{1 + e^{-X_i}} \quad \text{and} \quad f(i) = \frac{e^{-X_i}}{(1 + e^{-X_i})^2}$$

$$1 - F(i) = \frac{e^{-X_i}}{1 + e^{-X_i}} .$$

Therefore

$$\frac{f(i)}{1 - F(i)} = \frac{1}{1 + e^{-X_i}} .$$

Note that the hazard rate coincides with $F(i)$ in the case of the logistic distribution. Also in this formulation $F(i)$ is estimated by $\hat{P}_i$ in our model and is used as a regressor. Essentially, we are using the Heckman correction for a logistic distribution by inserting our instrument.

$$\sigma^2(\eta_i|\omega_i) = k_1^2\left[\frac{\pi^2}{3} - \alpha^2(2P_i - 1)^2\right].$$

Once α is fixed at α_0 the coefficient of $\hat{P}_i$ in (32) gives k_1. One can therefore use this $\hat{k}_1$ and α_0 to find relevant transformations for GLS. It is also possible to search for the best α within (± 1.0) in terms of the minimum SSE. Again, uniqueness of estimated k_1 at the final iteration could be a test of the validity of Bivariate Logistic specification.

IV. Consistency and Asymptotic Normality of the Estimators

The final version of the model implied by correction based on Gumbel's type II distribution of the random disturbances was reported as Equation (32) above. It should be noted that for any individual i, the probability P_i that the individual is selected is unknown. What we have to use instead is the estimate $\hat{P}_i$ computed from the logistic regression of the selection equation.

Recall that Equation (2) was estimated to yield

$$\ell n \left(\frac{\hat{P}_j}{1 - \hat{P}_j}\right) = x_{2j}\hat{\beta}_2 .$$

Clearly, $\hat{P}_j$ obtained as

$$\frac{1}{1 + e^{-x_{2j}\hat{\beta}_2}}$$

is stochastic. Use of $\hat{P}_j$ as an explanatory variable in (32) therefore creates a problem. We can demonstrate, however, that least-squares estimators of the parameters of (32) are consistent.

We know from standard results that $\hat{\beta}_2$ as a maximum-likelihood estimator of β_2 is consistent. This in turn implies that

$$\ell n\left(\frac{\hat{P}_j}{1-\hat{P}_j}\right)$$

is a consistent estimator of $\ell n(\frac{P_j}{1-P_j})$. Therefore, by Slutsky's theorem, $\hat{P}_j$ is a consistent estimation of P_j. In other words, if we express $\hat{P}_j = P_j + \psi_j$ where ψ_j is a random component, as we increase the sample size ψ_j goes to 0 for all j. It should be remembered that since both $\hat{P}_j$ and P_j lie between (0,1) ψ_j will lie between ($\pm$.1) even in small samples.

Let us define $z = (x_1 \,\vdots\, \hat{P})$ and $\delta = \binom{\beta_1}{k}$. We can write our model as

$$y_1 = z\delta + \tau, \text{ where } \tau = k\psi + \eta.$$

(k above is $k_1\alpha$.)
The OLS estimator $\hat{\delta} = \delta + (z/z)^{-1}z'\tau$. We now show that $\underset{T\to\infty}{p\ell im}\,\hat{\delta} = \delta$, where T is the sample size.

The argument for consistency of the OLS estimator in this equation can first be stated intuitively. We noted at the beginning of the paper that the sample selection bias arises because the disturbance terms in the

two equations of our model are interdependent. However, as we have demonstrated in detail in earlier sections, the residual error term η in Equation (32) is obtained after the conditional mean of u_1 is explicitly taken into account. It can be argued, therefore, that this random term η should be statistically independent of u_2 or any of its components.

Let us now take a closer look at the random error $\psi_i = P_i - \hat{P}_i$. We can write the logistic regression of the selection equation as

$$\widehat{\ell n\left(\frac{P}{1-p}\right)} = \ell n\left(\frac{P}{1-p}\right) + x_2(x_2'x_2)^{-1}x_2'\varepsilon_2$$

where ε_2 is the random disturbance of that model. Now define $x_2(x_2'x_2)^{-1}x_2' = W'$ which is an $N \times N$ matrix of non-stochastic elements (N is the number of observations in the full sample). More explicitly, for any i

$$\ell n\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = \ell n\left(\frac{P_i}{1-P_i}\right) + w_i'\varepsilon_2$$

where w_i' is the ith row of W'.

We can write

$$\frac{\hat{P}_i}{1-\hat{P}_i} = \left(\frac{P_i}{1-P_i}\right)e^{w_i'\varepsilon_2}$$

or

$$\frac{1}{1-\hat{P}_i} = \frac{P_i e^{w_i'\varepsilon_2} + (1-P_i)}{1-P_i}.$$

Therefore

$$\hat{P}_i = \frac{P_i e^{w_i'\varepsilon_2}}{P_i e^{w_i'\varepsilon_2} + (1-P_i)}$$

and

$$\psi_i = \hat{P}_i - P_i = \frac{P_i(1 - P_i)(e^{w_i'\varepsilon_2} - 1)}{P_i e^{w_i'\varepsilon_2} + (1 - P_i)} . \qquad (33)$$

As we argued before, η_i and ψ_i are statistically independent.

Now we specifically consider

$$\underset{T\to\infty}{p\ell im}\left(\frac{z'z}{T}\right)^{-1} \quad \text{and} \quad \underset{T\to\infty}{p\ell im}\left(\frac{z'\tau}{T}\right)$$

$$z'z = \begin{pmatrix} x_1'x_1 & x_1'\hat{P} \\ \hat{P}'x_1 & \hat{P}'\hat{P} \end{pmatrix} = \begin{pmatrix} x_1'x_1 & x_1'P + x_1'\psi \\ P'x_1 + \psi x_1 & P'P + 2P'\psi + \psi'\psi \end{pmatrix}$$

therefore

$$\underset{T\to\infty}{p\ell im}\left(\frac{z'z}{T}\right) = \underset{T\to\infty}{p\ell im}\begin{pmatrix} \frac{x_1'x_1}{T} & \frac{x_1'P}{T} + \frac{x_1'\psi}{T} \\ \frac{p'x_1}{T} + \frac{\psi'x_1}{T} & \frac{P'P}{T} + \frac{2P'\psi}{T} + \frac{\psi'\psi}{T} \end{pmatrix} .$$

This converges to finite moment matrix

$$M = \underset{T\to\infty}{p\ell im}\begin{pmatrix} \frac{x_1'x_1}{T} & \frac{x_1'P}{T} \\ \frac{P'x}{T} & \frac{P'P + \psi'\psi}{T} \end{pmatrix} .$$

In other words, $\underset{T\to\infty}{p\ell im}(\frac{z'z}{T})^{-1} = M^{-1}$ and is finite. Now consider that

$$\underset{T\to\infty}{p\ell im}\left(\frac{z'\tau}{T}\right) = \underset{T\to\infty}{p\ell im}\begin{pmatrix} \frac{x_1'\eta}{T} + \frac{kx_1'\psi}{T} \\ \frac{(P+\psi)'(\eta + k\psi)}{T} \end{pmatrix}$$

$$= p\ell im\begin{bmatrix} \frac{x_1'\eta}{T} + \frac{x'\psi}{T} \\ \frac{kP'\psi + k\psi'\psi + P'\eta + \psi'\eta}{T} \end{bmatrix}$$

Since x_1 and P are non-random and ψ can be regarded as satistically independent of η, the probability limit above is zero.

In other words,

$$\underset{T\to\infty}{p\ell im}\left(\frac{z'z}{T}\right)^{-1} \underset{T\to\infty}{p\ell im}\left(\frac{z'\tau}{T}\right) = 0.$$

That means $\underset{T\to\infty}{p\ell im}\,\hat{\delta} = \delta + \underset{T\to\infty}{p\ell im}(z'z)^{-1}z'\tau = \delta$ and consistency of least-squares estimators of coefficients in Equation (32) is established.

Next we demonstrate that estimated parameters from the augmented model represented by Equation (32) above have asymptotically normal distributions. We noted in the earlier section that correction for selection bias based on Gumbel II distribution results in heteroscedasticity of the error term. If we consider the asymptotic version, $\hat{P}$ can be replaced by P and (32) can be written as $y_1 = x_1\beta_1 + kP_1 + \eta$. However, $E(\eta\eta') = \Sigma$ is a diagonal matrix. A typical diagonal element of Σ is

$$\sigma_{ii} = \frac{k_1^2\pi^2}{3} - \alpha^2(2P_i - 1)^2 k_1^2$$

which is strictly greater than zero, but is finite because k and α are both finite. Now consider the transformation D such that $DD' = \Sigma^{-1}$. Again D is a diagonal matrix where

$$d_{ii} = \left[\frac{k_1^2\pi^2}{3} - k_1^2\alpha^2(2P_i - 1)\right]^{-1/2}$$

which is finite.

Consider the transformed system $Dy_1 = DZ\delta + D\eta$ which can be re-defined as $y_1^* = z^*\delta + \eta^*$. Note that $E(\eta^*\eta^{*\prime}) = I$ and our GLS estimator $\tilde{\delta} = (z^{*\prime}z^*)^{-1}z^{*\prime}y_1^* = \delta + (z^{*\prime}z^*)^{-1}z^{*\prime}\eta^* = \delta + B'\eta^*$, where

$$B' = (z^{*\prime}z^*)^{-1}z^{*\prime}. \qquad (34)$$

If we consider any typical estimator $\tilde{\delta}_s$ we obtain

$$\tilde{\delta}_s = \delta_s + \sum_{j=1}^{N_1} b_{sj}\eta_j^*$$

where b_{sj} are elements of B. We can prove that $\tilde{\delta}_s$ has an asymptotic normal distribution by proving that

$$\sum_{j=1}^{N_1} b_{sj}\eta_j^*$$

has an asymptotic normal distribution.

First we state the classical limit theorem and Lindeberg condition, which is sufficient for the theorem to be valid.

<u>The Classical Limit Theorem</u>

Let $h_1, h_2, \ldots, h_r, \ldots$ be a sequence of mutually independent random variables each having finite expectation $E(R_j) = a_j$ and variance $E(h_j - E(h_j))^2 = d_j^2$. Define

$$V_n^2 = \sum_{j=1}^{n} d_j^2 .$$

Then the sum

$$S_n = \frac{1}{V_n}\sum_{j=1}^{n} (h_j - a_j)$$

has a distribution which converges to the normal distribution if the <u>Lindberg condition</u> is satisfied.

Gnedenko (1967:365) has shown that the <u>Lindeberg condition</u> essentially stipulates that for any j $(j = 1, 2, \ldots, n)$, $\frac{h_j - a_j}{V_n}$ should be small.

We now prove a lemma which will be used to establish asymptotic normality of the GLS estimators of the parameters in (32).

<u>Lemma</u>. The sum

$$S_n = \sum_{j=1}^{n} \theta_j,$$

where

$$\theta_j = \frac{C_j\eta_j^*}{\left(\sum_{j=1}^{n} C_j^2\right)^{1/2}}$$

has an asymptotically normal distribution provided that C_j $(j = 1, 2, \ldots, n)$ is uniformly bounded.

<u>Proof</u>. The proof of the lemma is in two stages.

(i) First we prove that the lemma is a particular case of the classical limit theorem.

(ii) Second, we show that the Lindeberg condition is almost surely satisfied.

(i) Considering the fact that the C_j are non-stochastic

$$E(\theta_j) = \frac{C_j}{\left[\sum_{j=1}^{n} C_j^2\right]^{1/2}} E(\eta_j^*) = 0$$

and

$$\mathrm{Var}(\theta_j) = \frac{C_j^2}{\left[\sum_{j=1}^{n} C_j^2\right]} \mathrm{Var}(\eta_j^2) = \frac{C_j^2}{\sum_{j=1}^{n} C_j^2} .$$

Therefore

$$\mathrm{Var}(S_n) = \mathrm{Var}\left(\sum_{j=1}^{n} \theta_j\right) = \frac{\sum_1^n C_j^2}{\sum_1^n C_j^2} = 1.$$

Therefore, $S_n = \sum_{j=1}^{n} \theta_j$ is a special case of the classical limit theorem where $D_n = 1$.

(ii) The Lindeberg condition in this situation requires that θ_j be small for $j = 1, 2, \ldots, n$. This requirement may be seen as $\theta_j < \gamma$, where γ is an arbitrarily small number. Therefore the probability that the Lindeberg condition holds is equivalent to the probability that $[\theta_j < \gamma]$ for all j. That is,

$$\mathrm{prob}\left[\eta_j^* < \frac{\gamma}{C_j} \left(\sum_{j=1}^{n} C_j^2\right)^{1/2}\right] .$$

Now since C_j are uniformly bounded, the right-hand expression inside the brackets can be made arbitrarily large by increasing n. Consequently, this probability can be made as close to 1 as desired. We can therefore state that the Lindeberg condition is almost surely satisfied. This completes the proof of the lemma.

We have seen that

$$S_n = \frac{\sum_{j=1}^{n} C_j \eta_j^*}{\left[\sum_{t=1}^{n} C_t^2\right]^{1/2}}$$

has an asymptotic normal distribution. Therefore

$$LS_n = \sum_{j=1}^{n} C_j \eta_j^*,$$

where

$$L = \left(\sum_{t=1}^{n} C_t^2\right)^{1/2},$$

also has an asymptotic normal distribution. Now redefine $C_j = b_{sj}$. Note that b_{sj} is an element of the matrix $B = (z^{*\prime}z^*)^{-1}z^{*\prime}$ and is bounded therefore by virtue of the lemma above.

$$\tilde{\delta}_s = \sum_{j=1}^{n} b_{sj}\eta_j^*$$

has an asymptotically normal distribution.

V. Some Simulation Results for Different Censoring Adjustments

While the formal properties of our two logit-based estimators differ from one another and from Heckman's estimator, one might nevertheless wonder if in practice a choice among these three is likely to affect one's results. In order to explore this question we undertook the following simulation.

One thousand observations were drawn at random from a normal distribution with a mean of zero and a variance of one. (The sample was checked to confirm that chance factors had not produced an anomalous realization.) The data were then used to construct for each observation the probability value, our adjusted log odds and Heckman's hazard rate. With these in hand we calculated the linear correlation among the three adjustment variables.

The probability value correlated .959 with the hazard rate and -.978 with the adjusted log odds. The hazard rate and the adjusted log odds correlated .906. Clearly, these correlations suggest that the three adjustment variables will likely yield very similar results when the underlying distribution of the errors is normal. Note also that the linear correlations are high despite the fact that the justifications for the use of the probability value and the adjusted log odds rest on the logistic and not the normal form. Put in stronger terms, even if one mistakenly applies one of our logistic adjustments in situations where the underlying error distribution is bivariate normal, one should not be seriously misled. Moreover, this implies that should software not be readily available to apply Heckman's adjustment, one may rely on logistic estimation procedures that are found in such common statistical packages as SAS.

VI. An Illustration

It is widely recognized that individuals released from American prisons face a host of obstacles in their efforts to rejoin society. While these obstacles can take a number of forms, financial hardship is among the most important.

In this context, the Manpower Development and Training Act of 1962 (and subsequent amendments) gave to the Department of Labor a mandate to

programs responding tc the economic needs of ex-prisoners. The Department of Labor initially launched projects to provide manpower training within prisons and after release. Unfortunately, evaluation studies of these efforts found little evidence of success. The Department of Labor then turned to short-term income-support strategies for the period immediately after release from prison. These experimental efforts rested broadly on two kinds of interventions: special job placement and counselling for ex-prisoners, and the provision of unemployment insurance eligibility immediately upon release[3] (Lenihan, 1975; Mallar and Thornton, 1978; Rossi, Berk and Lenihan, 1980). In the first instance, special palcement and counseling services were thought to speed an ex-offender's re-entry into the labor force and through the income derived and the reduction in leisure time, decrease participation in criminal activities. In the second instance, modest transfer payments of about $65 a week for periods of up to 26 weeks (for those who could not find jobs) were thought to reduce the need to resort to crime as an alternative source of income and increase the opportunity costs of incarceration for any new crimes. Hence, recidivism would decline.[4]

After some initial pilot testing, the two interventions were field tested through a randomized experiment conducted in Baltimore. Approximately 400 men about to be released from state prisons were randomly assigned to

---

[3]Given the usual eligibility rules based on employment for the preceding five quarters (or so), most ex-prisoners were automatically excluded from unemployment insirance payments even if they had been working immediately prior to their incarceration.

[4]Actually, the theoretical mechanisms beneath the interventions (which include more than those briefly described in the text) were not fully clear at first and have only been recently explicitly articulated. Also, the nature of the interventions is somewhat more complicated than we have indicated here (Rossi, Berk and Lenihan, 1980).

either treatment, both treatments, or neither treatment (thus, three experimental groups and one control group in a fully crossed design). Ex-prisoners were then followed for up to two years in order to guage immediate and longer-term effects. The placement and counselling intervention apparently had no impact on re-arrests, but the benefits-eligibility intervention reduced re-arrests for property crimes about 8% (the difference between the two groups eligible for payments and the two groups that were not eligible). While these findings were statistically significant and encouraging, they were hardly definitive (Lenihan, 1975; Mallar and Thornton, 1978).

In order to better document the apparent success of the LIFE experiment and explore the mechanisms by which the reduction in re-arrests occurred, a second experiment was launched. Known as the TARP experiment (Transitional Aid Research Project), approximately 1000 men and women in each of the states of Texas and Geogria who were about to be released from state prisons were randomly assigned to treatment and control groups much like those used in the LIFE experiment. After a 12-month followup, there was no evidence that membership in any of the treatment groups by itself reduced the re-arrest rate. However, when the financial treatments were embedded in a set of structural equations depicting a range of related factors in the re-arrest process, substantial effects were found. In brief, unemployment benefits provided a work disincentive that markedly reduced labor force participation. Unemployment, in turn, increased the re-arrest rate for both property and non-property crimes. However, holding the impact of unemployment constant, every $100 of TARP payments reduced the <u>number</u> of re-arrests for property and non-property crimes about .02. Since payments to ex-prisoners

often exceeded $1500, reductions of .30 arrests were common[5] (see Rossi, Berk and Lenihan, 1980, for a full discussion of the findings). In short, the apparent null findings within an analysis of variance framework occurred because the indirect, "detrimental" effect of payments (through unemployment) obscured the direct "beneficial" effect of payments.[6]

In addition to the concern with re-arrests and employment, there was substantial interest in the impact of the TARP payments on "quality" of jobs eventually obtained by ex-prisoners. Presumably, TARP payments would subsidize a more thorough job search, perhaps leading to jobs with higher wages (among other things). However, in order to estimate equations for the wages of TARP participants (experimentals and controls), the possible effects of sample selection bias had to be addressed. We will use one of the analyses undertaken for Texas ex-prisoners to illustrate the use of our logit-based adjustment for sample selection bias.

In Table 1 we show the sample selection equation and results of logit estimation procedures in which the endogenous variable represents (in dummy form) whether any wages were "observed" during the first 12 months after release from prison. Evidence of wages came from Social Security files and from interviews with the ex-prisoners themselves 3 months, 6 months and 12 months after release. However, here we will rely on the former since, for a variety of reasons, the Social Security earnings were more accurate. The exogenous variables in the table are pretty much self-explanatory. The

---

[5]Since the mean number of arrests was well under 1, the reduction was clearly nontrivial.

[6]The job-counseling treatment had no effects either in the analysis of variance or structural-equation framework.

TABLE 1

Sample Selection Equation Using Logit Formulation

Variable	Coefficient	t-Value
Constant	-1.68	-2.78
Age 21-40 (dummy) -- x_1	0.30	1.58
Handicapped (dummy) -- x_2	-0.19	-1.14
Employed at time of arrest (dummy)[a] -- x_3	0.17	1.18
Number of previous arrests (integers) -- x_4	-0.02	-3.81
Released on parole (dummy) -- x_5	0.42	2.75
Job arranged before release (4-point scale) -- x_6	0.21	2.54
Male (dummy) -- x_7	1.41	4.19
Bexar County (dummy)[b] -- x_8	-0.33	-1.08
Dallas County (dummy) -- x_9	0.19	0.89
Harris County (dummy) -- x_{10}	0.44	2.21
Tarrant County (dummy) -- x_{11}	0.44	1.53
Black (dummy) -- x_{12}	-0.59	-3.66
Chicano (dummy) -- x_{13}	-0.44	-1.89
Vocational training in prison (dummy) -- x_{14}	0.15	0.87
Education (years) -- x_{15}	0.05	1.47
Driver's license (dummy) -- x_{16}	0.06	0.43
26 weeks/100% tax (dummy)[c] -- x_{17}	-0.57	-2.78
13 weeks/100% tax (dummy) -- x_{18}	-0.11	-0.58
13 weeks/25% tax (dummy) -- x_{19}	-0.49	-2.52
$F = 5.22$	$N = 975$	$p < .0001$

$$y_{2j} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \ldots + \beta_{19} x_{19,j} + u_{2j}$$

[a]Before the incarceration preceding the TARP study.

[b]County to which the individual was released. Rural counties are the residual category.

[c]The tax refers to the rate by which each dollar of payments was reduced for each dollar earned.

biographical variables reflect human capital attributes (e.g., being of prime working-force age, having physical handicaps, years of education), and attributes on which employers might "discriminate" (e.g., sex, race, criminal record). In addition, we included a dummy variable indicating whether an individual was released on parole (seeking work is often a parole condition), a scale indicating how likely prior to release a genuine job placement had been arranged, four dummy variables reflecting major urban counties in Texas (to capture part of the demand side), and three dummy variables for the three treatment groups.[7]

From Table 1, one can see that a number of exogenous variables predict the observation of wages at better than conventional chance levels. For one-tail tests (at the .05 level) we find that blacks, Chicanos, and individuals with a larger number of previous arrests were less likely to find employment. Male ex-prisoners, individuals on parole, individuals who are more likely to have a job arranged, residents in Harris County (i.e., Houston) were more likely to find employment. Finally, members of two of the three treatment groups (compared to the control groups not eligible for benefits) were less likely to find employment.[8] In short, there are no particular anomalies with perhaps the exception of null findings for the 13 week/100% tax treatment group (see Rossi, Berk and Lenihan, 1980, for a further discussion).

[7]The three groups varied in the maximum number of weeks of possible support and the rate at which earned dollars were "taxed." Thus, the third treatment group, for example, was eligible for 13 weeks of support with each earned dollar reducing each benefits dollar 25%.

[8]It was clear from the interviews that few individuals understood the implications of the tax rate. Thus, there may have been few perceived differences between the income treatments. In this context the t-value for the pooled treatment effect is well over 1.64. Finally, the job-counselling treatment had no impact and is therefore included among the control groups.

Table 2 shows the wage equation and the results from efforts to estimate wages for TARP ex-prisoners. These equations rest on 372 of the original 975 cases so that the potential for substantial selection bias exists. The specification for the wage equations are almost identical to the specification for the selection equation. Sex was dropped since only about 10% of the cases were female to begin with, and after the selection process, even less variance remained. Parole status was dropped because there was little reason to believe that being on parole had much impact on wages per se. Finally, the variable measuring whether an individual had a job arranged prior to release was deleted since that too seemed causally irrelevant to wages.

The first two columns of coefficients show the metric regression coefficients and t-values for the unadjusted equation. Using one-tail tests and the .05 level for the directional hypotheses, individuals of prime working age and with job experience prior to their recent incarceration earned higher wages: about $24 more per week and $16 more per week respectively. Ex-prisoners who returned to Harris County earned about $35 more a week. Finally, members in the first and third treatment groups earned approximately $27 more per week and $37 more per week respectively.

From the selection equation we estimated the predicted probability of "observing" wages and constructed both of our adjustment variables: the predicted probability and the adjusted log odds. They correlated .99 so that for all practical purposes either could be used. In the second two columns of Table 2 we show the results for the probability adjustment. To begin, the adjustment variable is statistically significant and clearly has a nontrivial impact. Every change of .10 in the probability of observing wages increases wages by over $8.00. Thus, selection processes that affect whether a job is obtained also affect weekly wages.

TABLE 2

Wage Equation Comparing Adjusted and Unadjusted Results

(Wages in Dollars per Week)

Variable	Unadjusted Coefficient	Unadjusted t-Value	Adjusted Coefficient	Adjusted t-Value
Constant	41.77	1.96	17.61	0.75
Probability adjustment--x_1	--	--	83.83	2.49
Age 21-40 (dummy)--x_2	24.09	2.39	18.98	1.86
Handicapped (dummy)--x_3	15.20	1.70	17.87	1.99
Employed at time of arrest (dummy)--x_4	15.58	1.95	10.50	1.38
Bexar County (dummy)--x_5	- 1.70	-0.09	5.70	0.31
Dallas County (dummy)--x_6	- 0.06	-0.01	0.01	0.00
Harris County (dummy)--x_7	35.12	3.39	27.34	2.55
Tarrant County (dummy)--x_8	- 4.85	-0.33	-15.17	-1.01
Black (dummy)--x_9	-11.44	-1.43	0.37	0.04
Chicano (dummy)--x_{10}	-14.99	-1.21	- 5.90	-0.45
Vocational training in prison (dummy)--x_{11}	-16.33	-1.85	-21.04	-2.35
Education (years)--x_{12}	2.03	1.15	0.73	0.40
Driver's licence (dummy)--x_{13}	2.53	0.32	0.96	0.12
26 weeks/100% tax (dummy)--x_{14}	26.55	2.40	35.79	3.09
13 weeks/100% tax (dummy)--x_{15}	6.20	0.66	7.40	0.78
13 weeks/25% tax (dummy)--x_{16}	37.32	3.63	44.27	4.18
	R^2 = .16 F = 4.12 N = 372 p < .001		R^2 = .17 F = 4.30 N = 372 p < .0001	

$$y_{1i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{16} x_{16,i} + u_{1i}$$

Again using one-tailed tests, it is apparent that in broad terms, the substantive story in the adjusted equation is much like the substantive story in the unadjusted equation. Yet, if for policy purposes one wanted to take the point estimates of the regression coefficients seriously, some nontrivial alterations surface. In particular, the impact of membership in the treatment groups has been enhanced. Members of the first treatment group are earning about $36.00 per week more (compared to $27.00 per week more) and members of the third treatment group earn about $44.00 per week more (compared to $37.00 per week more). In the first instance, the change represents a 33% increase, and in the second instance, the change represents a 19% increase. Moreover, if such differences were projected over the course of a year or more, substantial sums are involved.

Using two-tailed tests, it is also apparent that the adjustment equation yields two anomalous findings. Having a physical handicap <u>increases</u> earnings nearly $18.00 a week and vocational training in prison <u>decreases</u> earnings by about $21.00 a week. While one could certainly construct one or more post hoc explanations for these effects, the point here is that the sample selection adjustment does make a difference. Perhaps most important, it suggests the kinds of vocational training available in prison has no impact on whether a job is obtained, but may prepare ex-prisoners for lower paying jobs.

Finally, the sample selection adjustment definitively puts to rest the suggestion in the unadjusted equation that blacks and Chicanos receive lower wages. The point estimates from the unadjusted equation indicate that minorities receive about $15 a week less with t-value in excess of -1.20. Given strong priors one might be tempted to make something of such deficits. However, once the rather strong negative selection effects of minority status on

obtaining a job (see Table 1) are taken into account, the wage differences evaporate. That is, the wage disparities are really attributable to the initial likelihood of obtaining employment.

The results reported in Tables 1 and 2 represent but a sample of many analyses we have recently undertaken in which our logit-based sample-selection adjustments have been applied. From this experience, several broader points of possible importance to practitioners may be of interest.

First, the sample selection adjustments have rarely made an important difference in any substantive conclusions. This is <u>not</u> to argue that adjustments should not be used, but to argue that perhaps previous work in which sample-selection biases have been ignored may not be terribly misleading.

Second, the major computational costs for our estimators stem from the logit procedures employed in the selection equation. With this in mind, we have frequently reconstructed our adjustment variables from the results of linear probability models estimated with ordinary least squares. We have yet to find important disparities in the adjusted equations as a consequence. For the analysis reported here, regression coefficients rarely differed by more than one dollar from the adjusted results reported in Table 2 and the t-values typically varied in the first decimal place; clearly nothing of any importance changed. Thus, when one is working with large samples and situations in which the split on the dummy selection variable is no worse than about 90% to 10%, the linear probability model will probably suffice. It is at least clear from other work (Goodman, 1975) that the results from the logit and linear probability approaches are virtually the same under these circumstances.

Third, if a relatively large proportion of cases fail to exceed the selection boundary, it is critical to closely examine the qualities of the data that remain. One has in essence made the sample more homogeneous on a number of exogenous variables with the consequence that some exogenous variables may contain little variance and some exogenous variables may be highly collinear.

Finally, one has to be alert to the possibility that the adjustment variable will produce near singularity in the variance-covariance data matrix of the selected subset. This can result when most cases exceed the boundary and when the selection equation and the substantive equations are similarly specified, and it can also occur when one or more especially dominant variables in the selection equation also appear in the substantive equation.

REFERENCES

Amemiya, T.

1973 "Regression analysis when the dependent variable is truncated normal," Econometrica.

Gnedenko, B.V.

1967 The Theory of Probability (4th Ed.). New York: Chelsea.

Goodman, Leo A.

1975 "The relationship between modified and usual multiple-regression approaches to the analysis of dichotomous variables." Pp. 83-110 in D.R. Heise, ed., Sociological Methodology 1976. San Francisco: Jossey-Bass.

Gumbel, E.J.

1961 "Bivariate logistic distributions," Journal of the American Statistical Association.

Heckman, 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.

1979 "Sample selection bias as a specification error," Econometrica.

Lenihan, Kenneth J.

1975 Unlocking the Second Gate. R and D Monograph 45, U.S. Department of Labor, Employment and Training Division.

Mallar, Charles D., and Thornton, C.V.D.

1978 "Transitional aid among released prisoners: evidence from the Life Experiment," The Journal of Human Resources, Vol. 13, No. 2, pp. 208-236.

Nerlove, M., and Press, S.J.

1973 Univariate and Multivariate Log-Linear Models. Rand Corporation.

Olsen, R.J.

1980 "A least squares correction for selectivity bias," Econometrica, Vol. 48, pp. 1815-1820.

Ord, J.K.

1972 Families of Frequency Distributions. London: Griffin.

Rossi, Peter H., Berk, Richard A., and Lenihan, Kenneth J.

1980 Money, Work and Crime: Some Experimental Results. New York: Academic Press.

Tobin, J.

1958 "Estimation of relationship for limited dependent variables," Econometrica.

END