ON THE ESTIMATION OF PROBIT CHOICE MODEL WITH CENSORED DEPENDENT VARIABLES AND AMEMIYA'S PRINCIPLE

by

Lung-Fei Lee

Discussion Paper No. 78-99, April 1978

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

On The Estimation of Probit Choice Model with Censored Dependent Variables and Amemiya's Principle [1]

Lung-Fei Lee

1. Introduction

In Lee (1976), we have introduced a general class of econometric models involving dichotomous and censored dependent variables. Estimation methods which are computationally simple and consistent are introduced. These consistent estimates are then used as initial estimates in maximum likelihood procedures. However when there are too many parameters involved, maximum likelihood methods are still complicated and one may prefer to use the consistent estimates as final ones. To make statistical inferences, distributions of those estimates are also required. In this paper we will derive the correct asymptotic variance matrix for our estimates and compare it empirically with some approximates used. We also investigate a principle proposed by Amemiya (1977a, 1977b). The Amemiya principle is a general principle to derive structural parameter estimates from estimated reduced form parameters. Amemiya proved in the Nelson-Olson model with one limited dependent and one continuous endogenous variables (Amemiya (1977a)) and in Heckman's model with a continuous and a dummy variable (Amemiya (1977b)) that his principle can lead to simple estimators which are more efficient than the estimators derived in Nelson-Olson (1977) and Heckman (1977). Since Amemiya has only demonstrated the efficiency

---

[1] Prepared for the NBER-NSF Conference on Decision Rules and Uncertainty, Carnegie-Mellon University, March 29-April 1, 1978. I would like to thank Jerry Hausman, James Heckman, G. S. Maddala, Randall Olsen and Richard Westin for their comments. Any errors are solely my own.

of his approach in a case-by-case basis, one may wonder whether it still holds in more general and complicated models. We will point out that more efficient estimates can be derived from his prinicple in the most general model.

The paper is organized as follows. In the second section, we consider the asymptotic properties of some two stage estimators in a binary choice model with censored dependent variables. The exact covariance matrices are compared with some approximations. In the third section an empirical example is presented to illustrate the comparisons. In the fourth section, we investigate an estimation principle proposed by Amemiya in a general simultaneous equation model. The simultaneous equation model contains our binary choice model as well as Nelson-Olson and Heckman models as special cases. In the fifth section, censored simultaneous equation model and switching simultaneous equation model are considered. Various estimation methods are compared.

2. The Probit Choice Model with Censored Dependent Variables and the Two Stage Procedures

In Lee (1977), a class of probit choice models with limited dependent variables is introduced. One of the simple version of such models is

$$I_i^* = R_i\delta_0 + (W_{1i}^* - W_{2i}^*)\delta_1 + \varepsilon_i \tag{2.1}$$

where I_i^* is an unobservable index which determines the dichotomous observed choice I_i in the following way;

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

$$I_i = 0 \text{ otherwise.}$$

The disturbances ε_i are assumed to be independently and identically normally distributed with zero mean and finite variance. The vector of variables R_i is exogenous. If the variables W_{1i}^* and W_{2i}^* were exogenous and completely observed, equation (2.1) would be the usual probit model. In our model, we generalize it by allowing that W_{1i}^* and W_{2i}^* are exogenous and the observations are censored. Instead of observing W_{1i}^* and W_{2i}^*, W_i is observed and related to the underlying variables W_{1i}^* and W_{2i}^* as follows,

$$W_i = W_{1i}^* \quad \text{if } I_i = 1$$

$$W_i = W_{2i}^* \quad \text{otherwise.}$$

The exogenous variables W_{1i}^* and W_{2i}^* are assumed to be explained by other observable exogenous variables.

$$W_{1i}^* = X_{1i}\alpha_1 + \varepsilon_{1i} \tag{2.2}$$

(2.3) $$W_{2i}^{*} = X_{2i}\alpha_2 + \varepsilon_{2i}$$

The disturbances in (2.2) and (2.3) are also assumed to be jointly normal. Specifically, let

(2.4) $$I_i^{*} = X_i C - \varepsilon_{0i}$$

be the reduced form equation derived from (2.1), (2.2) and (2.3) where ε_{0i} is normalized to be a standard normal variate. We assume

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{0i} \end{pmatrix} \sim N(0, \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{10} \\ \sigma_{21} & \sigma_2^2 & \sigma_{20} \\ \sigma_{01} & \sigma_{02} & 1 \end{bmatrix})$$

and are independently and identically distributed. For the system as a whole, equations (2.1), (2.2) and (2.3) form a simultaneous equations model with qualitative and censored dependent variables. Equations (2.2) and (2.3) are always identifiable. Equation (2.1) will be identified under the exclusion of exogenous variables condition which is similar to the usual rank condition (a detail can be found in Lee (1977)).

Consistent estimates of equation (2.2) and (2.3) can be derived from some two stage procedures which are originated by Amemiya (1974) and Lewis (1974) and are generalized by Heckman (1976) and Lee (1976). The reduced form equation (2.4) is estimated by the probit maximum likelihood procedure and equations (2.5) and (2.6) are then estimated by ordinary least squares based on observed subsamples of W_1^{*} and W_2^{*} separately.

$$(2.5)\qquad W_{1i} = X_{1i}\alpha_1 - \sigma_{10}\frac{\phi(X_i\hat{C})}{\Phi(X_i\hat{C})} + \eta_{1i}$$

$$(2.6)\qquad W_{2i} = X_{2i}\alpha_2 + \sigma_{20}\frac{\phi(X_i\hat{C})}{1-\Phi(X_i\hat{C})} + \eta_{2i}$$

where ϕ and Φ denote the standard normal density function and distribution respectively, $\hat{C}$ is the probit estimate and

$$(2.7)\qquad \begin{aligned} \eta_{1i} &= \xi_{1i} - \sigma_{10}\left(\frac{\phi(X_iC)}{\Phi(X_iC)} - \frac{\phi(X_i\hat{C})}{\Phi(X_i\hat{C})}\right), \quad \xi_{1i} = \varepsilon_{1i} + \sigma_{10}\frac{\phi(X_iC)}{\Phi(X_iC)} \\ \eta_{2i} &= \xi_{2i} + \sigma_{20}\left(\frac{\phi(X_iC)}{1-\Phi(X_iC)} - \frac{\phi(X_i\hat{C})}{1-\Phi(X_i\hat{C})}\right), \quad \xi_{2i} = \varepsilon_{2i} - \sigma_{20}\frac{\phi(X_iC)}{1-\Phi(X_iC)} . \end{aligned}$$

To estimate equation (2.1) , a probit maximum likelihood procedure is applied to

$$(2.4)'\qquad I_i^* = R_i\delta_0^* + (X_{1i}\hat{\alpha}_1 - X_{2i}\hat{\alpha}_2)\delta_1^* - \tilde{\varepsilon}_{oi}$$

as if $\tilde{\varepsilon}_{0i}$ were a standard normal variable, where δ_0^*, δ_1^* are the identifiable parameters.

The asymptotic covariance matrices of the estimates for equations (2.5) and (2.6) have been derived in Lee et al (1977) in detail. For the sake of completeness, we mention these briefly.[2] Consider equation (2.5) and without loss of generalization, assume that the first N_1 observations corresponding to $I_i = 1$ in the whole samples with sample size N. Evidently,[3]

---

[2]The asymptotic normality of the estimators is taken to be granted. This can be easily shown under standard assumptions such as the boundedness of the exogenous variables and compactness of the parameter space.

[3] $\underset{=}{\Delta}$ means that the two expressions have the same asymptotic distribution.

$$(2.8) \qquad \binom{\hat{\alpha}_1}{\sigma_{10}} - \binom{\alpha_1}{\sigma_{10}} \triangleq (\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'\eta_1$$

$$\triangleq (\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'(\xi_1 + D_1(\hat{C}-C))$$

where $\tilde{X}_1$ is a matrix with the i^{th} row as $(X_{1i}, -\phi(X_iC)/\Phi(X_iC))$,

$$D_1 = -\sigma_{10}\begin{bmatrix} (X_1C\frac{\phi_1}{\Phi_1} + (\frac{\phi_1}{\Phi_1})^2)X_1 \\ \vdots \\ (X_{N_1}C\frac{\phi_{N_1}}{\Phi_{N_1}} + (\frac{\phi_{N_1}}{\Phi_{N_1}})^2)X_{N_1} \end{bmatrix}$$

with $\phi_i = \phi(X_iC)$ and $\Phi_i = \Phi(X_iC)$, the standard normal density function and distribution evaluated at X_iC respectively. $\hat{C}$ is the probit estimator of C in (2.4) and hence

$$(2.9) \qquad \hat{C} - C \triangleq [\Sigma_{i=1}^N \frac{\phi_i^2}{\Phi_i(1-\Phi_i)}X_i'X_i]^{-1} \Sigma_{i=1}^N \frac{\phi_i}{\Phi_i(1-\Phi_i)}X_i'(I_i-\Phi_i)$$

It can be shown that $E[(I_i-\Phi_i)(\varepsilon_{1j} + \sigma_{10}\,\phi_j/\Phi_j)|I_j=1]=0$ and hence $\tilde{X}_1'\xi_1$ and $\hat{C}$ are asymptotically uncorrelated. It follows from equations (2.7), (2.8) and (2.9) that

$$(2.10) \qquad \mathrm{var}\binom{\hat{\alpha}_1}{\sigma_{10}} = (\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'(V_1 + D_1(X'\Lambda X)^{-1}D_1')\tilde{X}_1(\tilde{X}_1'\tilde{X}_1)^{-1}$$

where V_1 is a $N_1 \times N_1$ diagonal matrix with the i^{th} diagonal element

$$\mathrm{var}(\xi_{1i}|I_i=1) = \sigma_1^2 - \sigma_{10}^2\, X_iC\,\frac{\phi(X_iC)}{\Phi(X_iC)} - \sigma_{10}^2(\frac{\phi(X_iC)}{\Phi(X_iC)})^2 \quad ;$$

Λ is a $N \times N$ diagonal matrix with the i^{th} diagonal element

$$\frac{\phi_i^2}{\Phi_i(1-\Phi_i)} \quad \text{and} \quad X = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} .$$

Similarly, the asumptotic covariance matrix for the consistent estimates from (2.6) is

$$(2.11) \qquad \mathrm{var}\begin{pmatrix} \hat{\alpha}_2 \\ \sigma_{20} \end{pmatrix} = (\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2(V_2 + D_2(X'\Lambda X)^{-1}D_2')\tilde{X}_2(\tilde{X}_2'\tilde{X}_2)^{-1}$$

where $\tilde{X}_2 = \begin{bmatrix} X_{2N+1}, & \hat{\phi}_{N_1+1}/(1-\hat{\Phi}_{N_1+1}) \\ \vdots & \vdots \\ X_{2N}, & \hat{\phi}_N/(1-\hat{\Phi}_N) \end{bmatrix}$,

$$D_2 = -\sigma_{20} \begin{vmatrix} (X_{N_1+1}C \frac{\phi_{N_1+1}}{1-\Phi_{N_1+1}} - (\frac{\phi_{N_1+1}}{1-\Phi_{N_1+1}})^2)X_{N_1+1} \\ \vdots \\ (X_N C \frac{\phi_N}{1-\Phi_N} - (\frac{\phi_N}{1-\Phi_N})^2)X_N \end{vmatrix}$$

and V_2 is a $(N-N_1) \times (N-N_1)$ diagonal matrix with diagonal element

$$\mathrm{var}(\xi_{2i}|I_i = 0) = \sigma_2^2 + \sigma_{20}^2 X_i C \frac{\phi(X_iC)}{1-\Phi(X_iC)} - \sigma_{20}^2 (\frac{\phi(X_iC)}{1-\Phi(X_iC)})^2 .$$

Comparisons of these asymptotic covariance matrices with the approximations which regards the estimated regressors as if they were the exact ones are obvious. The approximations will always underestimate the true variances by a nonnegative definite matrix since the approximations used are $(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'V_1\tilde{X}_1(\tilde{X}_1\tilde{X}_1)^{-1}$ and $(\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'V_2\tilde{X}_2(\tilde{X}_2'\tilde{X}_2)^{-1}$. However, it is evident from (3.3) and (3.4) that the relevant approximation will be quite close when σ_{10}^2 and σ_{20}^2 are much smaller than σ_1^2 and σ_2^2 respectively.

It is evident that the disturbances η_{1i} and η_{2i} in (2.5) and (2.6) are both heteroscedastic and autocorrelated. To derive more efficient GLS estimates, both these features should be taken into account.

Now let us consider the two stage probit estimator for equation (2.4)'. The two stage probit estimates are derived as the values which maximize the following function,

$$\ell nL = \Sigma_{i=1}^{N}\{I_i \ell n\Phi(R_i\delta_0^* + (X_{1i}\hat{\alpha}_1 - X_{2i}\hat{\alpha}_2)\delta_1^*) + (1-I_i)\ell n(1-\Phi(R_i\delta_0^* + (X_{1i}\hat{\alpha}_1 - X_{2i}\hat{\alpha}_2)\delta_1^*))\} \tag{2.12}$$

Let $\Theta_1' = (\delta_0^{*'}, \delta_1^*)$, $\Theta_2' = (\alpha_1', \alpha_2')$ and $\hat{\Theta}_1$ be the two stage probit estimates. Following Amemiya (1977),

$$\hat{\Theta}_1 - \Theta_1 \stackrel{\Delta}{=} -[E(\frac{\partial^2 \ell nL(\Theta_1,\Theta_2)}{\partial\Theta_1\partial\Theta_1'})]^{-1}(\frac{\partial \ell nL}{\partial\Theta_1} + E(\frac{\partial^2 \ell nL(\Theta_1,\Theta_2)}{\partial\Theta_1\partial\Theta_2'})(\hat{\Theta}_2-\Theta_2)) \tag{2.13}$$

Let $S = \begin{vmatrix} R_1, & X_{11}\alpha_1 - X_{21}\alpha_2 \\ \vdots & \vdots \\ R_N, & X_{1N}\alpha_1 - X_{2N}\alpha_2 \end{vmatrix}$, $\quad P = \begin{vmatrix} X_{11}\delta_1^*, & -X_{21}\delta_1^* \\ & \\ X_{1N}\delta_1^*, & -X_{2N}\delta_1^* \end{vmatrix}$

It can be easily shown that

$$E(-\frac{\partial^2 \ell nL}{\partial\Theta_1\partial\Theta_1'}) = S'\Lambda S \tag{2.14}$$

$$E(\frac{\partial^2 \ell nL}{\partial\Theta_1\partial\Theta_2'}) = -S'\Lambda P \tag{2.15}$$

and

$$\frac{\partial \ell nL}{\partial\Theta_1} = \Sigma_{i=1}^{N} \frac{\phi_i}{\Phi_i(1-\Phi_i)} S'_i(I_i-\Phi_i) \tag{2.16}$$

where S_i is the i^{th} row of S. Equation (2.13) can be rewritten as

$$\hat{\Theta}_1-\Theta_1 \stackrel{\Delta}{=} (S'\Lambda S)^{-1}(\Sigma_{i=1}^{N}\frac{\phi_i}{\Phi_i(1-\Phi_i)} S_i'(I_i-\Phi_i) - S'\Lambda P(\hat{\Theta}_2-\Theta_2)) \tag{2.17}$$

From equations (2.8) and (2.11)

$$(2.18)\qquad \hat{\alpha}_1 - \alpha_1 \stackrel{\Delta}{=} L_1(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'(\xi_1+D_1(\hat{C}-C))$$

$$(2.19)\qquad \hat{\alpha}_2 - \alpha_2 \stackrel{\Delta}{=} L_2(\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'(\xi_2-D_2(\hat{C}-C))$$

where L_1 and L_2 are relevant identity matrices augmented with a zero column. It follows from (2.9), (2.10), (2.11), (2.18) and (2.19) that the asymptotic coveriance matrix for $\hat{\theta}_2$ is

$$(2.20)\qquad V_2^* = \begin{bmatrix} L_1(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1' & 0 \\ 0 & L_2(\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2' \end{bmatrix}$$

$$\begin{bmatrix} V_1 + D_1(X'\Lambda X)^{-1}D_1' & -D_1(X'\Lambda X)^{-1}D_2' \\ -D_2(X'\Lambda X)^{-1}D_1' & V_2 + D_2(X'\Lambda X)^{-1}D_2' \end{bmatrix}$$

$$\begin{bmatrix} \tilde{X}_1(\tilde{X}_1'\tilde{X}_1)^{-1}L_1 & 0 \\ 0 & \tilde{X}_2(\tilde{X}_2'\tilde{X}_2)^{-1}L_2' \end{bmatrix}$$

The asymptotic covariance of $\hat{C}-C$ and $\Sigma_{i=1}^{N} \frac{\phi_i}{\Phi_i(1-\Phi_i)} S_i'(I_i-\Phi_i)$ follows from (2.9) which is $(X'\Lambda X)^{-1}X'\Lambda S$. Hence the asymptotic covariance of $\hat{\theta}_2 - \theta_2$ and $\Sigma_{i=1}^{N} \frac{\phi_i}{\Phi_i(1-\Phi_i)} S_i'(I_i-\Phi_i)$ is

$$(2.21)\qquad A = \begin{bmatrix} L_1(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1' & 0 \\ 0 & L_2(\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2' \end{bmatrix} \begin{bmatrix} D_1 \\ -D_2 \end{bmatrix} (X'\Lambda X)^{-1}X'\Lambda S$$

With (2.17), (2.20) and (2.21) the asymptotic covariance for $\hat{\theta}_1$ is

$$(2.22)\qquad V = (S'\Lambda S)^{-1}[I - S'\Lambda P]\begin{bmatrix} S'\Lambda S & A' \\ A & V_2^* \end{bmatrix}\begin{bmatrix} I \\ -P'\Lambda S \end{bmatrix}(S'\Lambda S)^{-1} .$$

If one regards that the likelihood function ℓnL in (2.12) is the correct likelihood for I_i, the asymptotic covariance matrix will be $(S'\Lambda S)^{-1}$. An analytic comparison between this approximation and (2.22) does not seem possible. However, it is evident that when σ_{10} and σ_{20} are small and A is closed to zero, the approximation will likely underestimate the true variances in (2.22).

3. Unionism and Wage Rate Revisited

In Lee (1976), the three equations model (2.2), (2.3), and (2.4)' has been applied to study the simultaneous effects of unionism and wages. The data consist of 3720 observations on operatives from the 1967 SEO sample. The list of variables used in the model is as follows:

N.E. = Northern-Eastern Region dummy variable; 1 for N.E.

N.C. = Northern - Central Region dummy variable; 1 for N.C.

S = Southern Region dummy variable; 1 for S.

UR_1 = In SMSA dummy variable; 1 for SMSA size 500,000 or more.

UR_2 = Outside SMSA dummy variable; 1 for outside SMSA.

ED_1 = Highest Grade Completed, grade from 0 to 4, dummy variable; 1 for 0 to 4 grades.

ED_2 = Highest Grade Completed dummy variable; 1 for grade 5 to 7.

ED_3 = Highest Grade Completed dummy variable; 1 for grade 9 to 11.

ED_4 = Highest Grade Completed dummy variable; 1 for grade 12.

ED_5 = Highest Grade Completed dummy variable; 1 for grade at least 13.

ME = Years of market experience.

ME^2 = Square of years of market experience.

RACE = Race dummy variable; 1 for white worker.

SEX = Sex dummy variable; 1 for male worker.

HLT = Health limitation dummy variable; 1 for worker with health limitation.

WK_1 = Weeks worked from 1 to 26 dummy variable; 1 for the workers in this category

WK_2 = Weeks worked from 48 to 52 dummy variable; 1 for those workers in this category.

IND_1 = Mining industry dummy variable; 1 for mining industry.

IND_2 = Construction industry dummy variable; 1 for the construction industry.

IND_3 = Durable goods manufacturing industry dummy variable; 1 for the durable goods manufacturing.

IND_4 = Nondurable goods manufacturing industry dummy variable; 1 for nondurable goods manufacturing.

U = percentage of union coverage in the relevant industry.

CCR = Industrial Concentration Ratio

W_u = Union hourly wage rate, measured in cents.

W_n = Nonunion hourly wage rate, measured in cents.

I = Union status of individual; 1 for workers in labor union.

The total number of parameters to be estimated is 66. Since the model is complicated in structures and involves large number of parameters and observations, the efficient maximum likelihood procedure will be extremely expensive. The two stage methods are much simpler and are used to estimate the model.

In Table 1 and Table 2, the estimates for the union and nonunion wages are presented. The exact asymptotic T-values which are derived from the correct asymptotic covariance are in the last columns. A comparison between the last two columns indicates that the approximations used are excellent for these two equations. The approximate T-values are only a slightly larger in some coefficients. At the significance level 0.05, all the significance coefficients revealed by the approximations are really significant. Even though σ_{10} and σ_{20} of the selectivity variables ϕ_i/Φ_i and $\phi_i/(1-\Phi_i)$ are significant (especially σ_{20}), the approximation is still excellent. This is so, even though the ratios $\hat{\sigma}^2_{10}/\hat{\sigma}^2_1$ and $\hat{\sigma}^2_{20}/\hat{\sigma}^2_2$ are not very small.[4]

In Table 3, the two stage probit estimates and their standard errors are presented. A comparison does reveal that the approximations underreport the errors. Even many of the coefficients remain highly significant, the coefficients of N.C., UR_2, ED_1, ED_5 and HLT are not significant at the 0.05 level of significance. The evaluation of the correct asymptotic covariance matrices seems necessary to obtain reliable inferences for the two stage probit estimator.

[4] $\hat{\sigma}^2_{10}/\hat{\sigma}^2_1 = 0.30$ and $\hat{\sigma}^2_{20}/\hat{\sigma}^2_2 = 0.15$.

TABLE 1

The Union Wage Equation (W_u)

Exogenous Variables	Coefficients	Approximated T-values	Exact T-values
Constant	4.431	27.129 *	27.151 *
NE	-0.083	-3.369 *	-3.281 *
NC	-0.007	-0.240	-0.250
S	-0.172	-5.422 *	-5.292 *
UR_1	0.067	3.279 *	3.175 *
UR_2	-0.092	-3.667 *	-3.580 *
ED_1	-0.108	-2.666 *	-2.584 *
ED_2	-0.033	-1.330	-1.264
ED_3	0.052	2.600 *	2.476 *
ED_4	0.111	5.168 *	4.978 *
ED_5	0.139	4.112 *	3.927 *
ME	0.016	7.526 *	7.619 *
ME^2	-0.0002	-5.418 *	-5.168 *
RACE	0.095	6.367 *	6.090 *
SEX	0.317	14.915 *	14.541 *
IND_1	0.223	4.034 *	3.878 *
IND_2	0.169	3.722 *	3.588 *
IND_3	0.034	1.477	1.405
IND_4	0.018	0.722	0.692
U	0.662	6.168 *	6.147 *
HLT	-0.055	-2.105 *	-2.037 *
$-\phi_i/\Phi_i$	-0.168	-1.914 Δ	-1.922 Δ

* indicates significance under 0.05 level of significance; Δ under 0.1 level of significance.

$\hat{\sigma}_1^2 = 0.093$

TABLE 2

The Nonunion Wage Equation (W_n)

Exogenous Variables	Coefficients	Approximated T-value	Exact T-values
Constant	4.754	71.220 *	70.118 *
NE	-0.091	-2.975 *	-2.907 *
NC	-0.074	-2.238 *	-2.189 *
S	-0.139	-4.427 *	-4.344 *
UR_1	0.039	1.610	1.585
UR_2	-0.067	-2.933 *	-2.876 *
ED_1	-0.049	-1.187	-1.167
ED_2	-0.016	-0.526	-0.530
ED_3	0.087	3.380 *	3.296 *
ED_4	0.157	5.979 *	5.836 *
ED_5	0.282	6.747 *	6.589 *
ME	0.012	5.468 *	5.455 *
ME^2	-0.0002	-4.788 *	-4.398 *
RACE	0.186	10.205 *	10.000 *
SEX	0.267	13.501 *	13.218 *
IND_1	0.120	1.915 Δ	1.866 Δ
IND_2	0.130	2.433 *	2.372 *
IND_3	0.053	1.474	1.440
IND_4	0.058	1.594	1.547
HLT	-0.088	-2.961 *	-2.904 *
$\phi_i/(1-\Phi_i)$	0.136	3.152 *	3.261 *

$\hat{\sigma}_2^2 = 0.120$

TABLE 3

The Union Status Equation (The Structural Form)

Exogenous Variables	Coefficients	Approximated Standard Error	Exact Standard Error
Constant	-0.654	0.145 *	0.309 *
NE	0.227	0.076 *	0.105 *
NC	0.197	0.077 *	0.107 Δ
S	-0.296	0.077 *	0.1122 *
UR_1	0.129	0.063 *	0.084
UR_2	-0.174	0.067 *	0.089 Δ
ED_1	-0.269	0.116 *	0.152 Δ
ED_2	-0.098	0.084	0.107
ED_3	0.079	0.072	0.092
ED_4	0.119	0.074	0.096
ED_5	0.258	0.119 *	0.156 Δ
ME	0.0020	0.0022	0.0028
RACE	0.166	0.054 *	0.072 *
SEX	0.093	0.055 Δ	0.072
WK_1	-0.372	0.115 *	0.115 *
WK_2	-0.017	0.073	0.073
CCR	0.365	0.132 *	0.137 *
HLT	-0.185	0.087 *	0.112 Δ
W_u-W_n	2.455	0.205 *	0.4011 *

4. Two Stage Methods and Amemiya's Principle in a Generalized Simultaneous-Equation Model

To provide an unified framework, let us consider the following simultaneous equation model

$$(4.1) \qquad \underline{Y}_i = \underline{Y}_i B + X_i \Gamma + \varepsilon_i \qquad i=1,\ldots,N$$

where $\underline{Y}_i$ is a $1 \times G$ row vector of endogenous variables, X_i is a $1 \times k$ vector of exogenous variables, $I-B$ is a $G \times G$ nonsingular matrix, Γ is a $k \times G$ matrix, $\varepsilon_i \sim N(0,\Sigma)$ and are i.i.d. The model differs from the usual simultaneous equation model in that $\underline{Y}_i$ may consist of latent variable, limited and censored dependent variables as well as observable continuous variables. Without loss of generality, we assume that $0 \leq G_1 \leq G_2 \leq G_3 \leq G$ and

1) the first G_1 variables $Y_{1i},\ldots,Y_{G_1 i}$ are observable continuous variables,
2) the next G_2-G_1 variables $Y_{G_1+1i},\ldots,Y_{G_2 i}$ are limited dependent variables, i.e., only when $Y_{ji} > 0$, one can observe it,
3) the next $G_3 - G_2$ variables $Y_{G_2+1i},\ldots,Y_{G_3 i}$ are unobservable latent variables. However binary indicators I_{ji} are observable and are determined by the latent variable Y_{ji} as follows,

$$I_{ji} = 1 \qquad \text{iff} \qquad Y_{ji} > 0$$

$$I_{ji} = 0 \qquad \text{otherwise,} \qquad j = G_2+1,\ldots,G_3$$

4) the last $G-G_3$ variables are censored dependent variables. The variables $Y_{G_3+1i},\ldots,Y_{Gi}$ are censored by a subset of

latent variables in 3). Specifically, the index set $\{G_3+1,\ldots,G\}$ can be partitioned into finite mutually exclusive and exhausted non-empty subsets S_k i.e. $\{G_3+1,\ldots,G\} = \cup_{\ell=1}^{L} S_\ell$ where $L \leq G_3-G_2$. For each $1 \leq \ell \leq L$, there is a unique latent variable $Y_{G_2+\ell i}$ activating on it;

$$Y_{j_\ell i} \text{ is observed if } Y_{G_2+\ell i} > 0, \text{ for all } j_\ell \varepsilon S_\ell^* \quad ;$$

$$Y_{j_\ell i} \text{ is observed if } Y_{G_2+\ell i} > 0, \text{ for } j_\ell \varepsilon\ S_\ell \setminus S_\ell^* \quad ;$$

where S_ℓ^* is a subset of S_ℓ which may be empty or equal S_ℓ.

The model in (1) is well-defined and contains our models in section 2 as well as Heckman's models without structural shifts (1976, 1977) and Nelson and Olson (1977) models as special cases.

The parameters B and Γ can be identified under rank conditions and suitable normalization rules. However in general, only certain nonlinear transformation of Σ will be identifiable when $G_3 < G$ and $\phi \neq S_\ell^* \subsetneq S_\ell$ for some ℓ.

To estimate model (4.1), maximum likelihood methods are too complicated to be useful. However consistent methods proposed by Heckman (1976, 1977), Lee (1977), Maddala and Lee (1976), Nelson and Olson (1976) can be easily extended. Alternative estimates can also be derived from Amemiya's principle (Amemiya (1977a, 1977b)). All those methods require estimation of reduced form parameters in the first stage. For the model in (4.1), reduced form equations always exist which are

$$(4.2) \qquad \underline{Y}_i = X_i \Pi + u_i$$

where $\Pi = \Gamma(I-B)^{-1}$ and $u_i = \varepsilon_i(I-B)^{-1}$. Each equation in (4.2) can be estimated by single equation method such as Probit, Tobit Maximum likelihood methods etc., depending on the nature of the dependent variable. They are defined into four categories as specified. The second stage in the estimation procedures is to estimate the structural parameters. To simplify notations, each single equation will be specified as

$$(4.3) \qquad Y_i = R_i\delta_0 + \underline{Y}_i^*\delta_1 + \varepsilon_i$$

where $\underline{Y}_i^*$ is a subvector of endogenous variables other than Y_i in $\underline{Y}_i$. Equation (4.3) can be modified to

$$(4.4) \qquad Y_i = R_i\delta_0 + (X_i\Pi^*)\delta_1 + v_i$$

where $\underline{Y}_i{}^* = X_i\Pi^* + u_i^*$. With consistent estimates $\hat{\Pi}^*$ derived in the first stage. The second stage in the methods proposed (except Amemiya's method) is to estimate (δ_0, δ_1) from

$$(4.5) \qquad Y_i = R_i\delta_0 + (X_i\hat{\Pi}^*)\delta_1 + w_i$$

where $w_i = v_i + X_i(\Pi^* - \hat{\Pi}^*)\delta_1$. (4.5) is estimated by Probit, Tobit, etc., depending on the nature of Y_i in (4.5).

Instead of estimating (4.5), Amemiya suggests one should solve by regression methods the structural parameters from the estimated reduced form parameters. Based on this principle, one can derive alternative estimates. Let $R = XJ_1$ where J_1 consists of unit and zero column vectors. From (4.4), one has

$$(4.6) \qquad Y_i = X_i(J_1\delta_0 + \Pi^*\delta_1) + v_i$$

Let c be the corresponding reduced form parameter vector of Y_i in (4.2). It is obvious that

$$c = J_1\delta_0 + \Pi^*\delta_1 \tag{4.7}$$

The estimates suggested by Amemiya are OLS or GLS estimates derived from

$$\hat{c} = J_1\delta_0 + \hat{\Pi}^*\delta_1 + \xi \tag{4.8}$$

where $\xi = \hat{c} - c - (\hat{\Pi}^* - \Pi^*)\delta_1$.

Under general regular conditions, all these estimation methods give consistent and asymptotic normal estimates. Amemiya in the two mentioned cases showed that his GLS estimates are more efficient. The question remained is to compare his GLS estimates with the other consistent methods in the general model with arbitrary number of equations and different type of endogenous variables.

4.1 Structural equation with probit structure

The $G_3 - G_2$ equations have unobservable latent variables in the left-hand-side which belong to this category. In this case, the endogenous variable Y_i is an unobservable latent variable. The two stage estimates are derived from maximizing the function ℓnL in (4.9) w.r.t. $\Theta_1^{*\prime} = (\delta_0^{*\prime}, \delta_1^{*\prime})$ which are the identifiable parameters under the normalization $\sigma_v^2 = 1$.

$$(4.9) \qquad \ell nL = \Sigma_{i=1}^{N}\{I_i \ell n\Phi(R_i\delta_0^* + (X_i\hat{\Pi}^*)\delta_1^*) + (1-I_i)\ell n(1-\Phi(R_i\delta_0^* + (X_i\hat{\Pi}^*)\delta_1^*))\}$$

where I_i is the observed dichotomous indicator of Y_i, Φ is the standard normal c.d.f. Let $P = [X\delta_{11},\ldots,X\delta_{1M}]$, $\Theta_2' = (\Pi_1^{*\prime},\ldots,\Pi_M^{*\prime})$, $S = [R\ X\Pi^*]$ where $\delta_1^{*\prime} = [\delta_{11},\ldots,\delta_{1M}]$ and $\Pi^* = (\Pi_1^*,\ldots,\Pi_M^*)$. Let

$$\Lambda_1 = \begin{bmatrix} \frac{\phi_1}{1-\Phi_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\phi_N}{1-\Phi_N} \end{bmatrix} \quad \text{and} \quad \Lambda = \begin{bmatrix} \frac{\phi_1^2}{1-\Phi_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\phi_N^2}{1-\Phi_N} \end{bmatrix}$$

where ϕ_i and Φ_i are standard normal density and c.d. functions evaluated at $R_i\delta_0^* + X_i\Pi^*\delta_1^*$. Following Amemiya, the asymptotic distribution of this two stage estimates $\hat{\Theta}_1$ can be derived from

$$(4.10) \qquad \hat{\Theta}_1 - \Theta_1 \overset{\Delta}{=} (S'\Lambda S)^{-1}(S'\Lambda_1(I-\Phi) - S'\Lambda P(\hat{\Theta}_2 - \Theta_2))$$

where $\overset{\Delta}{=}$ means both sides have the same asymptotic distributions and $I - \Phi$ is a $N \times 1$ vector consisting of $I_i - \Phi_i$. The detailed expression for the asymptotic variance matrix is lengthy but can be derived in a straightforward manner. The two stage estimates can then be compared with Amemiya's GLS estimates.

<u>Proposition 1</u> For equation (4.3) with unobservable latent variable Y_i and its dichotomous realization I_1, the two stage estimate $\hat{\Theta}_1$ derived from maximizing equation (4.9) is asymptotically less efficient than the GLS estimate $\hat{\Theta}_1^A$ derived from Amemiya's principle.

<u>Proof</u>: From (4.10), the asymptotic variance of $\hat{\Theta}_1$ is

$$V_{\hat{\Theta}_1} = (S'\Lambda S)^{-1}\{S'\Lambda S + S'\Lambda P V_{\hat{\Theta}_2} P'\Lambda S - S'\Lambda_1 E_2' P'\Lambda S - S'\Lambda P E_2 \Lambda_1 S\}(S'\Lambda S)^{-1}$$

where E_2 is the asymptotic covariance of $(\hat{\Theta}_2 - \Theta_2)$ and $(I - \Phi)$. The asymptotic variance matrix of $\hat{\Theta}_1^A$ is

$$V_{\hat{\Theta}_1^A} = (Z'\Omega_\xi^{-1}Z)^{-1}$$

where $Z = [J_1 \quad \Pi^*]$ and Ω_ξ in this probit structural equation is

$$\Omega_\xi = [I \quad P_1] \begin{bmatrix} (X'\Lambda X)^{-1} & * \\ -E_2\Lambda_1 X(X'\Lambda X)^{-1} & V_{\hat{\Theta}_2} \end{bmatrix} \begin{bmatrix} I \\ P_1' \end{bmatrix}$$

with $P = XP_1$. The two expressions $V_{\hat{\Theta}_1}$ and Ω_ξ follow since c is the probit maximum likelihood estimate of the reduced form equation and

$\hat{c} - c \stackrel{\Delta}{=} (X'\Lambda X)^{-1}X'\Lambda_1(I-\Phi)$. On the other hand, $V_{\hat{\Theta}_1}$ can be rewritten as

$$V_{\hat{\Theta}_1} = (Z'X'\Lambda XZ)^{-1}Z'X'\Lambda X\Omega_\xi X'\Lambda XZ(Z'X'\Lambda XZ)^{-1} .$$

It follows that $V_{\hat{\Theta}_1} - V_{\hat{\Theta}_1 A}$ is nonnegative definite. Q.E.D.

4.2 Structural equations with Censored dependent variable

The last $G-G_3$ equations belong to this category. The variable Y_i is censored. When S_ℓ^* and its complement set $S_\ell \setminus S_\ell^*$ are nonempty, there are switching systems. For the variable Y in S_ℓ^* which is observed when $Y_{G_2+\ell,i} > 0$, based on observed subsamples, equation (4.4) can be rewritten

$$(4.11) \qquad Y_i = R_i\delta_0 + (X_i\Pi^*)\delta_1 + \lambda\frac{\phi(X_i\alpha)}{\Phi(X_i\alpha)} + \xi_i$$

where $E(\xi_i|I_i) = 0$, I_i is the dichotomous indicator of the underlying latent variable which activates the censoring and α is the reduced form parameters of that latent variable. The two stage estimator is to estimate α and Π^* in the first stage from Probit maximum likelihood and similar equations in (4.16) and estimate $\Theta_1' = (\delta_0', \delta_1', \lambda)$ in the second stage from

$$(4.12) \qquad Y_i = R_i\delta_0 + (X_i\hat{\Pi}^*)\delta_1 + \lambda\frac{\phi(X_i\hat{\alpha})}{\Phi(X_i\hat{\alpha})} + \eta_i$$

$$\text{where} \qquad \eta_i = \xi_i - \lambda\left(\frac{\phi(X_i\hat{\alpha})}{\Phi(X_i\hat{\alpha})} - \frac{\phi(X_i\alpha)}{\Phi(X_i\alpha)}\right) - X_i(\hat{\Pi}^* - \Pi^*)\delta_1 \quad .$$

For the variable Y in $S_\ell \setminus S_\ell^*$, (4.12) should be modified to

$$(4.12)' \qquad Y_i = R_i\delta_0 + (X_i\hat{\Pi}^*)\delta_1 + \lambda\frac{\phi(X_i\hat{\alpha})}{1-\Phi(X_i\hat{\alpha})} + \eta_i$$

$$\text{where} \quad \eta_i = \xi_i - \lambda\left(\frac{\phi(X_i\hat{\alpha})}{1-\Phi(X_i\hat{\alpha})} - \frac{\phi(X_i\alpha)}{1-\Phi(X_i\alpha)}\right) - X_i(\hat{\Pi}^* - \Pi^*)\delta_1 \quad .$$

All these expressions can be represented as

$$(4.13) \qquad Y_i = R_i\delta_0 + (X_i\hat{\Pi}^*)\delta_1 + \lambda G(X_i\hat{\alpha}) + \eta_i$$

where $G(X_i\alpha)$ is a nonlinear function of $X_i\alpha$ and

$$\eta_i = \xi_i - \lambda(G(X_i\hat{\alpha}) - G(X_i\alpha)) - X_i(\hat{\Pi}^* - \Pi^*)\delta_1 .$$

Let $S = [\, R \quad X\Pi^* \quad G]$ and $\tilde{S}$ its estimated value where G is the vector consisted of $G(X_i\alpha)$. The two stage estimator $\hat{\Theta}_1$ from (4.13) is

$$(4.14) \qquad \hat{\Theta}_1 = (\tilde{S}'\tilde{S})^{-1}\tilde{S}'Y$$

Let $P = X(\delta_1' \otimes I)$ and $\rho = \xi + \lambda D_1(\hat{\alpha}-\alpha)$ where D_1 is the gradient matrix of G evaluated at α. The asymptotic covariance matrix can be derived from

$$(4.15) \qquad \hat{\Theta}_1 - \Theta_1 \triangleq (\tilde{S}'\tilde{S})^{-1}\tilde{S}'(\rho - P(\hat{\Theta}_2 - \Theta_2)).$$

To derive the Amemiya's GLS, one has to estimate the reduced form parameter $c' = (c_1', c_2')$ for Y_i from

$$(4.16) \qquad Y_i = X_i c_1 + G(X_i\hat{\alpha})c_2 + \rho_i \quad .$$

It is $\hat{c} = (W'W)^{-1}W'Y$ where $W = [X \quad G]$. Let $Z = [J_1 \quad J_2\Pi^* \quad J_3]$ be defined from $S = W[J_1 \quad J_2\Pi^* \quad J_3]$. The GLS derived from Amemiya's principle is

$$(4.17) \qquad \hat{\Theta}_1{}^A = (\tilde{Z}'\tilde{\Omega}_\omega^{-1}\tilde{Z})^{-1}\tilde{Z}'\tilde{\Omega}_\omega^{-1}\hat{c} \quad ,$$

where Ω_ω is the asymptotic covariance matrix of $\omega = \hat{c}-c + J_2(\Pi^*-\hat{\Pi}^*)\delta_1$. The comparisons of $\hat{\Theta}_1$ and $\hat{\Theta}_1{}^A$ follow from the following proposition.

<u>Proposition 2</u>. In the equation with censored dependent variable Y_i, the GLS estimate (4.17) derived from Amemiya's principle is asymptotically more efficient than the two stage estimate $\hat{\Theta}_1$ in (4.14).

<u>Proof</u>: From (4.15), the asymptotic covariance matrix of $\hat{\Theta}_1$ is

$$V_{\hat{\Theta}_1} = (Z'W'WZ)^{-1}Z'W'(V_\rho+PV_{\hat{\Theta}_2}P'-PE-E'P')WZ(Z'W'WZ)^{-1}.$$

where V_ρ is the asymptotic variance matrix of ρ, E is the asymptotic covariance matrix of $\hat{\Theta}_2 - \Theta_2$ and ρ On the other hand, $V_{\hat{\Theta}_1{}^A} = (Z'\Omega_\omega^{-1}Z)^{-1}$ from (4.17). Since $\hat{c}-c \triangleq (W'W)^{-1}W'\rho$, it implies

$$\Omega_\omega = (W'W)^{-1}W'V_\rho W(W'W)^{-1} + J_2(\delta_1'\otimes I)V_{\hat{\Theta}_2}(\delta_1\otimes I)J_2'$$
$$- J_2(\delta_1'\otimes I)EW(W'W)^{-1} - (W'W)^{-1}W'E'(\delta_1\otimes I)J_2'$$

Since $P = WJ_2(\delta_1\otimes I)$, $V_{\hat{\Theta}_1}$ can be rewritten as

$$V_{\hat{\Theta}_1} = (Z'W'WZ)^{-1}Z'W'W\Omega_\omega W'WZ(Z'W'WZ)^{-1} .$$

It follows that $V_{\hat{\Theta}_1} - V_{\hat{\Theta}_1{}^A}$ is nonnegative definite. Q.E.D.

It is interesting to point out that the conclusion applies not only to the censored dependent variables in (4.12) and

(4.12)'. It can also be applied to other models. In Amemiya (1977b), he derived the asymptotic covariance matrix for a two equations Heckman model with structural shift but he failed to apply his principle to that model. For the following equation with structural shift,

$$(4.18) \qquad Y_{2i} = R_i\delta_0 + Y_{1i}\delta_1 + I_i\delta_2 + \varepsilon_i$$

where Y_{1i} is an unobservable latent variable with dichotomous realization I_i, Y_{2i} is observable continuous variable. From the reduced form equation for Y_{1i},

$$(4.19) \qquad Y_{1i} = X_i c + u_i$$

equation (4.18) is modified to

$$(4.20) \qquad Y_{2i} = R_i\delta_0 + (X_i c)\delta_1 + \Phi(X_i c)\delta_2 + \xi_i$$

where Φ is standard normal c.d.f. With consistent estimates $\hat{c}$ available, Heckman's two stage estimator for $\Theta_1' = (\delta_0', \delta_1')$ is derived by least squares applied to

$$(4.20)' \qquad Y_{2i} = R_i\delta_0 + (X_i\hat{c})\delta_1 + \Phi(X_i\hat{c})\delta_2 + \eta_i \ .$$

It is obvious that equation (4.20)' is a special case in (4.13). Hence it is possible to apply Amemiya's principle to equation (4.18) and more efficient estimates can be derived.

The efficiency of the estimators derived from Amemiya's principle in the structural equations with continuous endogenous variable and Tobit structures can also be shown. The details are in the Appendix.

5. Switching and Censored Simultaneous Equations Models with Sample Separation Information.

Switching and censored simultaneous equations systems with sample separation are special cases of the general model introduced in section 4. These models correspond to the cases in which $G_2 = 0$, $G_3 = 1$, $S_1 = \{G_2+1,\ldots,G\}$ and the unobserved latent variable Y_1 will not appear explicitly in other equations. The system is a switching simultaneous equation model when S_1^* is nonempty, $S_1^* \neq S_1$, the endogenous variables in S_1^* and $S_1 \setminus S_1^*$ form a complete simultaneous equation system in each regime and the endogenous variables in one regime will not appear in another one. The model discussed in section 1 is a special case in which there is only one single equation in each regime. This switching simultaneous equation model differs from the models studied in Goldfeld and Quandt (1972, 1973, 1976) in that the sample separation information is available.[5] The censored simultaneous equation model introduced by Heckman (1974 , 1976) can also be regarded as a special case in which either S_1^* is empty or $S_1^* = S_1$.

In this section, we would consider procedures such as two stage least squares and instrumental variables methods,[6] and Amemiya's principle in the estimation of single structural equations in the

[5] These two approaches have many different aspects in identification, estimation and empirical applications. In our model, the structural equations in each regime can be identified under the usual rank conditions. This is not the case when the sample separation information is not available, see Goldfeld and Quandt (1973). More discussions on the value of sample separation can be found in Goldfeld and Quandt (1975), Kiefer (1978) and Lee (1977).

[6] Instrumental variables methods on the estimation of usual simultaneous equation models can be found in Theil (1971), Sargan (1958), Brundy and Jorgensen (1974) and Henry (1976), etc.

the system. For each structural equation (4.3) in regime S_1^*, based on observed subsamples corresponding to that regime, equation (4.3) can be rewritten as

$$(5.1) \qquad Y_i = R_i\delta_0 + \underline{Y}_i^*\delta_1 + \lambda \frac{\phi(X_i\alpha)}{\Phi(X_i\alpha)} + \varepsilon_i^*$$

where $E(\varepsilon_i^*|I_i = 1) = 0$. For the structural equation in the other regime,

$$(5.2) \qquad Y_i = R_i\delta_0 + \underline{Y}_i^*\delta_1 + \lambda \frac{\phi(X_i\alpha)}{1-\Phi(X_i\alpha)} + \varepsilon_i^*$$

where $E(\varepsilon_i^*|I_i = 0) = 0$. These two expressions can be represented by

$$(5.3) \qquad Y_i = R_i\delta_0 + \underline{Y}_i^*\delta_1 + \lambda G(X_i\alpha) + \varepsilon_i^*$$

and so it is enough to consider the estimation of (5.3). Let $\hat{\alpha}$ be the probit maximum likelihood estimator of the reduced form Y_1. (5.3) can be modified to

$$(5.4) \qquad Y_i = R_i\delta_0 + \underline{Y}_i^*\delta_1 + \lambda G(X_i\alpha) + \eta_i$$

Let H be a matrix consisting of $(R_i, \underline{Y}_i^*, G(X_i\alpha))$ as its ith row and $\tilde{H}$ be its estimated value evaluated at $\hat{\alpha}$. In matrix form, equation (5.4) is

$$(5.5) \qquad Y = \tilde{H}\Theta + \eta$$

where $\Theta' = (\delta_0', \delta_1', \lambda)$. The disturbances η_i in (5.5) are heteroscedastic and autocorrelated as pointed out in section 2. There are several methods to estimate the above equation.

Method 1: Let $\tilde{X}$ be a matrix with $(X_i, G(X_i\hat{\alpha}))$ as its rows. Premultiply (5.5) by $\tilde{X}$,

$$\tilde{X}'Y = \tilde{X}'\tilde{H}\Theta + \tilde{X}'\eta \tag{5.6}$$

This equation can then be estimated by GLS with $\tilde{X}'\tilde{X}$ as the "covariance" matrix. The estimator is

$$\hat{\Theta}_{(1)} = [\tilde{H}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{H}]^{-1}\tilde{H}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y$$

with asymptotic covariance matrix

$$V(\hat{\Theta}_{(1)}) = [\tilde{H}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{H}]^{-1}\tilde{H}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{V}_\eta\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{H}[\tilde{H}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{H}]^{-1}$$

where V_η is the covariance matrix of η and $\tilde{V}_\eta$ is a consistent estimate of V_η. Method 1 is similar to usual two stage least squares procedures. Method 1 has been discussed in Heckman [1976] and Lee et al [1977].

Method 2: Equation (5.6) is estimated by GLS with covariance matrix $\tilde{X}'\tilde{V}_\eta\tilde{X}$. The estimator is

$$\hat{\Theta}_{(2)} = [\tilde{H}'\tilde{X}(\tilde{X}'\tilde{V}_\eta\tilde{X})^{-1}\tilde{X}'\tilde{H}]^{-1}\tilde{H}'\tilde{X}(\tilde{X}'\tilde{V}_\eta\tilde{X})^{-1}\tilde{X}'Y$$

with asymptotic covariance matrix $V(\hat{\Theta}_{(2)}) = [\tilde{H}'\tilde{X}(\tilde{X}'\tilde{V}_\eta\tilde{X})^{-1}\tilde{X}'\tilde{H}]^{-1}$. Method 2 differs from method 1 in that the correct asymptotic covariance matrix of $\tilde{X}'\eta$ is used.

Method 3: Premultiply equation (5.5) by $\tilde{X}'\tilde{V}_\eta^{-1}$,

$$\tilde{X}'\tilde{V}_\eta^{-1}Y = \tilde{X}'\tilde{V}_\eta^{-1}\tilde{H}\Theta + \tilde{X}'\tilde{V}_\eta^{-1}\eta \tag{5.7}$$

GLS procedure is then applied to (5.7). The estimator is

$$\hat{\Theta}_{(3)} = [\tilde{H}'\tilde{V}_\eta^{-1}\tilde{X}(\tilde{X}'\tilde{V}_\eta^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{V}_\eta^{-1}\tilde{H}]^{-1}\tilde{H}'\tilde{V}_\eta^{-1}\tilde{X}(\tilde{X}'\tilde{V}_\eta^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{V}_\eta^{-1}Y$$

with asymptotic covariance matrix $V(\hat{\Theta}_{(3)}) = [\tilde{H}'\tilde{V}_\eta^{-1}\tilde{X}(\tilde{X}'\tilde{V}_\eta^{-1}\tilde{X})\tilde{X}'\tilde{V}_\eta^{-1}\tilde{H}]^{-1}$. This method is similar to a two stage generalized least squares procedure.

Method 4: Choose an instrumental variables matrix $\tilde{V}_\eta^{-1}\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{H}$. The following IV estimator can be derived,

$$\hat{\Theta}_{(4)} = [\tilde{H}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{V}_\eta^{-1}\tilde{H}]^{-1}\tilde{H}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{V}_\eta^{-1}Y .$$

The asymptotic covariance matrix of $\hat{\Theta}_{(4)}$ is

$$V(\hat{\Theta}_{(4)}) = [\tilde{H}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{V}_\eta^{-1}\tilde{H}]^{-1}\tilde{H}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{V}_\eta^{-1}\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{H}[\tilde{H}'\tilde{V}_\eta^{-1}\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{H}]^{-1}.$$

It is of interest to compare these various estimators. Since $H = \tilde{X}Z + [0, V^*, 0]$ with $Z = [J_1, \Pi^*, J_3]$ as in (4.6), it can be easily shown that $\operatorname{plim} \frac{1}{N_1}\tilde{X}'H = \operatorname{plim} \frac{1}{N_1}\tilde{X}'\tilde{X}Z$, $\operatorname{plim} \frac{1}{N_1}\tilde{X}'V_\eta^{-1}H = \operatorname{plim} \frac{1}{N_1}\tilde{X}'V_\eta^{-1}\tilde{X}Z$ where N_1 is the number of sample observations in the relevant regime. It follows that

$$\operatorname{plim} N_1 V(\hat{\Theta}_{(4)}) = \operatorname{plim} N_1 V(\hat{\Theta}_{(3)}) = \operatorname{plim} N_1[Z'\tilde{X}'V_\eta^{-1}\tilde{X}Z]^{-1} ,$$

$$\operatorname{plim} N_1 V(\hat{\Theta}_{(2)}) = \operatorname{plim} N_1[Z'\tilde{X}'\tilde{X}(\tilde{X}'V_\eta\tilde{X})^{-1}\tilde{X}'\tilde{X}Z]^{-1}$$

$$\text{and } \operatorname{plim} N_1 V(\hat{\Theta}_{(1)}) = \operatorname{plim} N_1[Z'\tilde{X}'\tilde{X}Z(Z'\tilde{X}'V_\eta\tilde{X}Z)^{-1}Z'\tilde{X}'\tilde{X}Z]^{-1} .$$

The estimators $\hat{\Theta}_{(3)}$ and $\hat{\Theta}_{(4)}$ are asymptotically equivalent. But from the computational point of view, $\hat{\Theta}_{(4)}$ is relatively simpler. Since it is obvious that[7]

---

[7] $A \geq B$ means that $A - B$ is nonnegative definite.

$$Z'\tilde{X}'V_\eta^{-1}\tilde{X}Z \geq Z'\tilde{X}'\tilde{X}(\tilde{X}'V_\eta\tilde{X})^{-1}\tilde{X}'\tilde{X}Z \geq Z'\tilde{X}'\tilde{X}Z(Z'\tilde{X}'V_\eta\tilde{X}Z)^{-1}Z'\tilde{X}'\tilde{X}Z,$$

$\hat{\Theta}_{(4)}$ and $\hat{\Theta}_{(3)}$ are asymptotically more efficient than $\hat{\Theta}_{(2)}$; and $\hat{\Theta}_{(2)}$ is more efficient than $\hat{\Theta}_{(1)}$.

The estimator $\hat{\Theta}_{(4)}$ derived from method 4 has an optimal property. It is the most efficient IV estimator in the set of IV estimators in the estimation of equation (5.5). This can be shown as follows. Let W be an arbitrary instrumental variables matrix. The corresponding IV estimator is

$$\hat{\Theta}_W = (W'H)^{-1}W'Y$$

with asymptotic covariance matrix $V(\hat{\Theta}_W) = (W'H)^{-1}W'V_\eta W(H'W)^{-1}$. Since $\text{plim}\ N_1V(\hat{\Theta}_W) = \text{plim}\ N_1[Z'\tilde{X}'W(W'V_\eta W)^{-1}W'\tilde{X}Z]^{-1}$ and $\tilde{X}'V_\eta^{-1}\tilde{X} \geq \tilde{X}'W(W'V_\eta W)^{-1}W'\tilde{X}$, $\hat{\Theta}_{(4)}$ is asymptotically more efficient than $\hat{\Theta}_W$.

Now let us consider Amemiya's principle which is applied to

$$\hat{C} = \tilde{Z}\Theta + \omega \tag{5.8}$$

where $\tilde{Z} = [J_1 \quad \hat{\Pi}^* \quad J_3]$ with $\hat{C} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y$ and $\hat{\Pi}^* = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\underline{Y}^*$. The GLS estimator derived from Amemiya's principle is

$$\hat{\Theta}_G^A = (\tilde{Z}'\tilde{\Omega}_\omega^{-1}\tilde{Z})^{-1}\tilde{Z}'\tilde{\Omega}_\omega^{-1}\hat{C} .$$

As derived in section (4.2), $\tilde{\Omega}_\omega = \tilde{X}'\tilde{X}(\tilde{X}'V_\eta\tilde{X})^{-1}\tilde{X}'\tilde{X}$. It is obvious that $\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'H = \tilde{X}\tilde{Z}$. It follows $\hat{\Theta}_G^A = \hat{\Theta}_{(2)}$, i.e., Amemiya's GLS procedure is exactly method 2. Therefore, one concludes that the estimators $\hat{\Theta}_{(3)}$ and $\hat{\Theta}_{(4)}$, are asymptotically more efficient than GLS estimator derived from Amemiya's principle. Let us now analyze the OLS estimator derived from Amemiya's principle. The OLS estimator from (5.8) is

$$\hat{\Theta}_L^A = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\hat{C}.$$

It follows

$$\hat{\Theta}_L^A = [\tilde{Z}'\tilde{X}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{X}\tilde{Z}]^{-1}\tilde{Z}'\tilde{X}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\hat{C}$$

$$= [H'\tilde{X}(\tilde{X}'\tilde{X})^{-2}\tilde{X}'H]^{-1}H'\tilde{X}(\tilde{X}'\tilde{X})^{-2}\tilde{X}'Y.$$

which is the GLS procedure with "covariance" matrix $(\tilde{X}'\tilde{X})^2$ applied to (5.6). This estimator is less efficient than $\hat{\Theta}_{(2)}$, $\hat{\Theta}_{(3)}$ and $\hat{\Theta}_{(4)}$ but $\hat{\Theta}_L^A$ and $\hat{\Theta}_{(1)}$, in general, will not dominate each other.

As a final remark, we would like to point out that $\hat{\Theta}_{(3)}$ and $\hat{\Theta}_{(4)}$ are computationally simple as the GLS estimator derived from Amemiya's principle. As demonstrated in section 2, V_η in the first regime is

$$V_\eta = V_1 + D_1(X'\Lambda X)^{-1}D_1'$$

where V_1 is a diagonal matrix. Hence the following inversion relation can be used,

$$V_\eta^{-1} = V_1^{-1} - V_1^{-1}D_1(X'\Lambda X + D_1'V_1^{-1}D_1)^{-1}D_1'V_1^{-1}$$

and we will not face the problem to invert numerically on $N_1 \times N_1$ matrix. Similarly, this is true for V_η in the second regime.

6. Conclusions

In this paper, we have derived the asymptotic coveriance matrices for several simple consistent estimators in the binary choice models with censored dependent variables. The simple consistent estimates have been demonstrated to be valuable and the asymptotic covariance matrices have also been compared with some approximations. The approximations which regard the estimated regressors as if they are the exact ones turn out to, in general, underestimate the true errors. However, the approximations used for the truncated (or censored) equations such as (2.2) and (2.3) are very close to the correct ones and the differences can almost be neglected. This holds even though the selectivity variables are highly significant. More evidence on this point is also available in Lee, Maddala and Trost (1977) in another empirical study. However, this is not the case in the estimation of the structural choice equation (2.1) by the two stage probit procedures. The approximations for the two stage probit estimates seriously underestimate the correct standard errors. Hence we conclude that the evaluation of the correct asymptotic covariance matrices is necessary to obtain reliable inferences.

We have also analysed an estimation principle of Amemiya in a general simultaneous equation model. The model consists of observable continuous endogenous variables, unobservable latent endogenous variables with dichotomous indicator, limited and censored dependent variables in a simultaneous equation framework. This general model contains

Nelson and Olson simultaneous Tobit model, Heckman simultaneous dummy endogenous variables model, censored simultaneous equation model and switching simultaneous equations models as special cases. Various consistent two stage estimation methods are generalized. Amemiya's principles are investigated in this general model with an arbitrary number of equations. Amemiya's generalized two stage estimators are compared with the other two stage estimators. It was shown that Amemiya's estimators are more efficient in all the cases. As contrary to Amemiya (1977a), his principle can also be applied to Heckman's model with structural shift. A generalized two stage estimator derived from his principle is also found to be more efficient than Heckman's approach. The proofs are general and do not depend on a case-by-case basis.

In the censored simultaneous equation models and switching simultaneous equation models, GLS and OLS estimators derived from Amemiya's principle can be identified as instrumental variables methods. Two estimation methods which give more efficient estimators than GLS estimator derived from Amemiya's principle are found. These two estimators are shown to be asymptotically equivalent and are computationally simple.

References

Amemiya, T. (1974), "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal", Econometrica, Vol. 42, No. 6, pp. 999-1012.

Amemiya, T. (1976), "The Specification and Estimation of A Multivariate Logit Model," Technical Report no. 211, Institute for Mathematical Studies in the Social Sciences, Stanford University.

Amemiya, T. (1977a), "The Estimation of a Simultaneous Equation Tobit Model", Technical Report no. 236, Institute for Mathematical Studies in the Social Sciences, Stanford University.

Amemiya, T. (1977b), "The Estimation of a Simultaneous Equation Generalized Probit Model", Technical Report no. 237, Institute for Mathematical Studies in the Social Sciences, Stanford University, forthcoming in Econometrica.

Brundy, J. M. and D. W. Jorgenson (1974), "Consistent and Efficient Estimation of Systems of Simultaneous Equations by Means of Instrumental Variables", pp. 215-244 in P. Zarembka (ed), Frontiers in Econometrics, Academic Press, New York.

Goldfeld, S. M. and R. E. Quandt (1972), Nonlinear Methods in Econometrics, Amsterdam, North-Holland Publishing Company.

Goldfeld, S. M. and R. E. Quandt (1975), "Estimation in a Disequilibrium Model and the Value of Information", Journal of Econometrics, 3, pp. 325-348.

Goldfeld, S. M. and R. E. Quandt (1976), "Estimation of Structural Change in Simultaneous Equation Models", Chapter 2, in Studies in Nonlinear Estimation, ed. by S. M. Goldfeld and R. E. Quandt, Cambridge, Massachusetts, Ballinger Publishing Co.

Heckman, J. J. (1974), "Shadow-Price Market Wages and Labor Supply", Econometrica, Vol. 42, pp. 679-694.

Heckman, J.J. (1976), "The Common Structure of a Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models", Annals of Economic and Social Measurement, Vol. 5, pp. 475-492.

Heckman, J.J. (1976), "Simultaneous Equation Models with Both Continuous and Discrete Endogenous Variables with and Without Structural Shift in the Equations", in Studies in Nonlinear Estimation, ed. by Goldfeld, S.M. and R.E. Quandt, 1976, Ballinger Publishing Co.

Heckman, J.J. (1977), "Dummy Endogenous Variable in a Simultaneous Equation System", June. Forthcoming in Econometrica.

Hendry, D. F. (1976), "The Structure of Simultaneous Equation Estimators", Journal of Econometrics, Vol. 4, No. 1, pp. 51-88.

Kiefer, N. M. (1978), "On the Value of Sample Separation Information", mimeo. University of Chicago, forthcoming in Econometrica.

Lee, L. F. (1976), "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," Discussion paper 76-77-04, University of Florida; forthcoming in International Economic Review.

Lee, L. F. (1977), "Identification and Estimation in Binary Choice Models with Limited Dependent Variables," Discussion paper no. 77-85, Center for Economic Research, University of Minnesota.

Lee, L. F., G. S. Maddala and R. P. Trost (1977), "Instrumental Variable Estimation of Simultaneous Equation System with Selectivity," Discussion paper, University of Florida.

Lewis, H. Gregg (1974), "Comments on Selectivity Biases in Wage Comparisons," Journal of Political Economy, Nov./Dec.

Maddala, G. S. and L. Lee (1976), "Recursive Models with Qualitative Endogenous Variables", Annals of Economic and Social Measurement, 1976, Fall.

Nelson, F. D. (1977), "Censored Regression Models with Unobserved, Stochastic, Censored Thresholds", Nov. 1977, Journal of Econometrics. Vol. 6, No. 3., pp. 309-327.

Nelson, F. D. and L. Olson (1977), "Specification and Estimation of a Simultaneous-Equation Model with Limited Dependent Variables", Social Science Working Paper no. 149, California Institute of Technology: forthcoming in International Economic Review.

Sargan, J. D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables, Econometrica, Vol. 26, No. 3, pp. 393-415.

Theil, H. (1971), Principles of Econometrics, John Wiley & Sons, Inc.

Appendix

A.1 Structural equation with observable continuous endogenous variable

The first G_1 equations in our model (4.1) are in this category. Y_i in (4.3) is an observable continuous variable. Let $\Theta_1' = (\delta_0', \delta_1')$, $S = [R \quad X\Pi^*]$ be a matrix with $(R_i \quad X_i\Pi^*)$ in its ith row and $\tilde{S}$ be its estimated value. OLS procedure can be applied to (4.5). The two stage estimate is thus

$$\text{(A.1.1)} \qquad \hat{\Theta}_1 = (\tilde{S}'\tilde{S})^{-1}\tilde{S}'Y \ .$$

This two stage method is similar to Theil's two stage least squares method (1971) and was used in Heckman (1976, 1977). Amemiya GLS estimate derived from (4.8) is

$$\text{(A.1.2)} \qquad \hat{\Theta}_1^A = (\tilde{Z}'\tilde{\Omega}_\xi^{-1}\tilde{Z})^{-1}\tilde{Z}'\tilde{\Omega}^{-1}\hat{c}$$

where $Z = [J_1 \quad \Pi^*]$, Ω_ξ is the variance matrix of ξ and $\tilde{Z}$ and $\tilde{\Omega}_\xi$ are their estimated values.

In the two equations Nelson-Olson and Heckman model, Amemiya derived the asymptotic variance matrices for $\hat{\Theta}_1$ and $\hat{\Theta}_1^A$. He also gave separate proofs in the two models that $\hat{\Theta}_1^A$ is more efficient than $\hat{\Theta}_1$. For our model, the asymptotic covariance matrices are quite lengthy but can be derived in a straightforward manner. The interesting thing is to compare their efficiency. This follows from the following proposition.

<u>Proposition A1</u> For equation (4.3) with observable continuous variable Y_i, the estimate $\hat{\Theta}_1^A$ in (A.1.2) is asymptotically more efficient than the estimate $\hat{\Theta}_1$ in (A.1.1)

A.2

Proof: Let $P = [X\delta_{11},\ldots,X\delta_{1M}]$, $\Theta_2' = (\Pi_1^{*\prime},\ldots,\Pi_M^{*\prime})$ where $\delta_1' = [\delta_{11},\ldots,\delta_{1M}]$ and $\Pi^* = (\Pi_1^*,\ldots,\Pi_M^*)$. Denote the asymptotic variance matrix of $\hat{\Theta}_2' = (\hat{\Pi}_1^{*\prime},\ldots,\hat{\Pi}_M^{*\prime})$ by $V_{\hat{\Theta}_2}$ and the asymptotic covariance of $\hat{\Theta}_2$ and v by E_1. From (4.5) it is obvious that the variance matrix of $\hat{\Theta}_1$ is

$$V_{\hat{\Theta}_1} = (S'S)^{-1}S'\{\sigma_v^2 I + PV_{\hat{\Theta}_2}P' - PE_1 - E_1'P'\}\, S(S'S)^{-1} .$$

The asymptotic variance matrix of $\hat{\Theta}_1^A$ is

$$V_{\hat{\Theta}_1^A} = (Z'\Omega_\xi^{-1}Z)^{-1}$$

To compare $V_{\hat{\Theta}_1}$ and $V_{\hat{\Theta}_1^A}$, one note from (4.8) that

$$\Omega_\xi = [I - P_1] \begin{bmatrix} \sigma_v^2(X'X)^{-1} & (X'X)^{-1}X'E_1' \\ * & V_{\hat{\Theta}_2} \end{bmatrix} \begin{bmatrix} I \\ -P_1' \end{bmatrix}$$

where $P_1 = \delta_1' \otimes I$, $\otimes$ the Kronecker product. Since $P = XP_1$ and $S = XZ$,

$$\begin{aligned} V_{\hat{\Theta}_1} &= (Z'X'XZ)^{-1}Z'X'\{\sigma_v^2 I + XP_1V_{\hat{\Theta}_2}P_1'X' - XP_1E_1 - E_1'P_1'X'\}XZ(Z'X'XZ)^{-1} \\ &= (Z'X'XZ)^{-1}Z'X'X\Omega_\xi X'XZ(Z'X'XZ)^{-1} . \end{aligned}$$

It follows $V_{\hat{\Theta}_1} - V_{\hat{\Theta}_1^A}$ is nonnegative definite and $\hat{\Theta}_1^A$ is more efficient.

Q.E.D.

A.3

It is interesting to note that the proposition holds no matter how $\hat{\Theta}_2$ is derived so far as the asymptotic variance $\hat{\Theta}_2 - \Theta_2$ exists and $\hat{\Theta}_2$ is consistent.

A.2 Structural equation with Tobit structure

This is the case in which Y_i in (4.3) is limited dependent. Equations $G_1+1,\ldots,G_2$ in the general model (4.1) are in this category. A two stage estimate $\hat{\Theta}_1$ was proposed in Nelson and Olson (1977). This estimation method can be generalized to our model. The two stage estimates are derived from maximizing the following function.

$$\text{(A.2.1)}\quad \ln L(\Theta_1) = \Sigma_{i=1}^{N}\{-\tfrac{1}{2}I_i \ln\sigma_v^2 - \frac{1}{2\sigma^2} I_i(Y_i - R_i\delta_0 - (X_i\hat{\Pi}^*)\delta_1)^2$$

$$+ (1-I_i)\ln(1-\Phi(R_i\delta_0 + (X_i\hat{\Pi}^*)\delta_1))\}$$

Define

$$a_{11i} = (\frac{s_i\Theta_1}{\sigma_v}\phi_i - \frac{\phi_i^2}{1-\Phi_i} - \Phi_i)/\sigma_v^2$$

$$a_{22i} = \frac{1}{4\sigma_v^4}(\phi_i(\frac{s_i\theta_1}{\sigma_v})^3 + \frac{s_i\Theta_1}{\sigma_v}\phi_i - (\frac{s_i\Theta_1}{\sigma_v})^2\frac{\phi_i^2}{1-\Phi_i} - 2\Phi_i)$$

$$a_{12i} = \frac{-\phi_i}{2\sigma_v^3}((\frac{s_i\theta_1}{\sigma_v^2})^2 + 1 - \frac{s_i\Theta_1}{\sigma_v}\frac{\phi_i}{1-\Phi_i}) \;;$$

let A_{ij} be the $N\times N$ diagonal matrix whose k-th diagonal element is a_{ijk}, $i,j = 1,2$. and

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix}.$$

It is easy to show that

$$E(\frac{\partial^2 \ln L(\Theta_1)}{\partial\Theta_1\partial\Theta_1'}) = Z^{*\prime}X^{*\prime}AX^*Z^*$$

A.5

$$E\left(\frac{\partial^2 \ell n L(\Theta_1,\Theta_2)}{\partial\Theta_1\ \partial\Theta_2'}\right) = -Z^{*\prime}X^{*\prime}AP^{*}$$

and $$\frac{\partial n L(\Theta_1)}{\partial\Theta_1} = Z^{*\prime}X^{*}u$$

where $X^* = \begin{bmatrix} X & 0 \\ 0 & \ell \end{bmatrix}$, $Z^* = \begin{bmatrix} Z & 0 \\ 0 & 1 \end{bmatrix}$, $P^* = \begin{bmatrix} P \\ 0 \end{bmatrix}$,

ℓ is $N\times 1$ vector with unity in all the components, 0 is the appropriate zero matrix or vector and u is a $2N\times 1$ vector,

$$u' = [\frac{1}{\sigma_v}\frac{\phi_1}{1-\Phi_1}(1-I_1) + \frac{1}{\sigma_v^2} I_1v_1,\ldots,\frac{1}{\sigma_v}\frac{\phi_N}{1-\Phi_N}(1-I_N) + \frac{1}{\sigma_v^2} I_Nv_N ,$$

$$\frac{s_1\Theta_1}{2\sigma_v^3}\frac{\phi_1}{1-\Phi_1}(1-I_1) - \frac{1}{2\sigma_v^2} I_1 + \frac{1}{2\sigma_v^4} I_1v_1^2,\ldots,$$

$$\frac{s_N\Theta_N}{2\sigma_v^3}\frac{\phi_N}{1-\Phi_N}(1-I_N) - \frac{1}{2\sigma_v^2} I_N + \frac{1}{2\sigma_v^4} I_Nv_N^2]$$

It follows from Taylor series expansion,

$$\text{(A.2.2)} \quad \hat{\Theta}_1 - \Theta_1 \triangleq L_1[Z^{*\prime}X^{*\prime}AX^{*}Z^{*}]^{-1}(Z^{*\prime}X^{*\prime}u - Z^{*\prime}X^{*\prime}AP^{*}(\hat{\Theta}_2 - \Theta_2)) .$$

where $L_1 = [I, 0]$ is an identity matrix augmented with a column of zeroes. The detailed expression for $V_{\hat{\Theta}_1}$ is quite lengthy but can be derived in a straightforward manner as in Amemiya (1977a).

With consistent estimates $\hat{\Theta}_2$ from the first stage and $\hat{c}$ from the Tobit maximum likelihood, one can compare the above two stage estimate with Amemiya's generalized least squares estimate $\hat{\Theta}_1^A$.

Proposition A.2 For equation (4.3) with limited dependent variable Y_i, the two stage estimate $\hat{\Theta}_1$ derived from maximizing equation (A.2.1) is asymptotically less efficient than $\hat{\Theta}_1^A$.

Proof: From (A.2.2), the asymptotic variance of $\hat{\Theta}_1$ is

$$V_{\hat{\Theta}_1} = L_1(Z^{*\prime}X^{*\prime}AX^{*}Z^{*})^{-1}Z^{*\prime}X^{*\prime}AX^{*}\{(X^{*\prime}AX^{*})^{-1} + P_1^{*}V_{\hat{\Theta}_2}P_1^{*\prime} - (X^{*\prime}AX^{*})^{-1}X^{*\prime}EP_1^{*} - P_1^{*}E'X^{*}(X^{*\prime}AX^{*})^{-1}\}\, X^{*\prime}AX^{*}Z^{*} (Z^{*\prime}X^{*\prime}AX^{*}Z^{*})^{-1}L_1'$$

where

$P_1^{*} = \begin{bmatrix} P_1 \\ 0 \end{bmatrix}$ and E is the asymptotic covariance matrix of u and $(\hat{\Theta}_2 - \Theta_2)'$. To compare the asymptotic variance of $\hat{\Theta}_1^A$, one notes that $\hat{c}$ is a Tobit maximum likelihood estimate and hence

$$\hat{c} - c \overset{\Delta}{=} -L_2(X^{*\prime}AX^{*})^{-1}X^{*\prime}u$$

where L_2 is an appropriate identity matrix augmented with a zero column.

It follows that the asymptotic variance of $\hat{\Theta}_1^A$ is

$$V_{\hat{\Theta}_1^A} = (Z'\Omega_\xi^{-1}Z)^{-1}$$

where

$$\Omega_\xi = L_2(X^{*\prime}AX^{*})^{-1}L_2' + P_1V_{\hat{\Theta}_2}P_1' - P_1E'X^{*}(X^{*\prime}AX^{*})^{-1}L_2' - L_2(X^{*\prime}AX^{*})^{-1}X^{*\prime}EP_1' .$$

A.7

To compare $V_{\hat{\Theta}_1}$ with $V_{\hat{\Theta}_1 A}$, one has to evaluate $L_1(Z^{*\prime}X^{*\prime}AX^{*}Z^{*})^{-1}$ and $L_2(X^{*\prime}AX^{*})^{-1}L'_2$. This can be done with the well-known formulae to find the inverse of a partitioned matrix. Let

$$B = A_{11} - A_{12}\ell(\ell' A_{22}\ell)^{-1}\ell' A_{21} .$$

It is easy to check that the following equalities hold:

$$\text{(i)}\quad L_1(Z^{*\prime}X^{*\prime}AX^{*}Z^{*})^{-1} = (Z'X'BXZ)^{-1}[I - Z'X'A_{12}\ell(\ell' A_{22}\ell)^{-1}]$$

$$\text{(ii)}\quad L_2(X^{*\prime}AX^{*})^{-1} = (X'BX)^{-1}[I - Z'X'A_{12}\ell(\ell' A_{22}\ell)^{-1}]$$

$$\text{(iii)}\quad L_2(X^{*\prime}AX^{*})^{-1}X^{*\prime} = (X'BX)^{-1}[X' - Z'X'A_{12}\ell(\ell' A_{22}\ell)^{-1}\ell']$$

$$\text{(iv)}\quad [I - Z'X'A_{12}\ell(\ell' A_{22}\ell)^{-1}]Z^{*\prime}X^{*\prime}AX^{*\prime}P_1^{*} = Z'X'BXP_1$$

$$\text{(v)}\quad [I - Z'X'A_{12}\ell(\ell' A_{22}\ell)^{-1}]Z^{*\prime}X^{*\prime} = Z'X'BXL_2(X^{*\prime}AX^{*})^{-1}X^{*\prime} .$$

It follows from (i) - (v)

$$V_{\hat{\Theta}_1} = (Z'X'BXZ)^{-1}Z'X'BX\Omega_{\xi}X'BXZ(Z'X'BXZ)^{-1}$$

and hence $V_{\hat{\Theta}_1} - V_{\hat{\Theta}_1 A}$ is nonnegative definite. Q.E.D.