⟨Research Paper⟩

Journal of the Korean
Statistical Society
Vol. 14, No. 1, 1985

On the Bivariate Dichotomous Choice Model

Hea-Jung Kim*

ABSTRACT

Data set generated by the bivariate dichotomous choice made by individuals often occurs in practice. This paper presents general model of how such data set is generated as well as methods of estimation. The M.L.E. is examined and found to be computationally burdensome. A simpler estimator, the bivariate dichotomous two-stage estimator, is suggested as an alternative. The two-stage estimator is found to be as efficient as the M.L.E.

1. Introduction

Lee (1970) and Duncan (1982), among others, formulated mixed, continuous/discrete dependent variable models with normal distribution. But they considered the case when the dependent variable is limited by binary or polytomous choice. Amemiya (1974) and Morimune (1979) proposed estimation methods for the bivariate discrete dependent variable model considered by Ashford and Sowden (1970).

In this paper we try to bridge the gap existing between the two models. We develop a method of estimating parameters of a mixed, continuous/discrete dependent variable model when the dependent variable is limited by the bivariate dichotomous choice made by individuals. The following model shows the reason for the appreciation. In Section 4 we give an example of the model.

$$\begin{aligned} Y_{1t} &= \beta'_1 X_{1t} + e_{1t} \quad \text{iff } I^t_{00}=1 \\ Y_{2t} &= \beta'_2 X_{2t} + e_{2t} \quad \text{iff } I^t_{10}=1 \\ Y_{3t} &= \beta'_3 X_{3t} + e_{3t} \quad \text{iff } I^t_{01}=1 \end{aligned} \tag{1.1}$$

* Department of Statistics, Dongguk University, Seoul 100, Korea.

$$Y_{4t}=\beta_4' X_{4t}+e_{4t} \qquad \text{iff } I_{11}^t=1$$
$$t=1,2,\cdots\cdots, \quad T,$$

where

$$I_{00}^t=\begin{cases}1 & \text{if } a'U_t \leqq e_{ta} \text{ and } b'V_t \leqq e_{tb} \\ 0 & \text{o.w.}\end{cases}$$
$$I_{10}^t=\begin{cases}1 & \text{if } a'U_t > e_{ta} \text{ and } b'V_t \leqq e_{tb} \\ 0 & \text{o.w.}\end{cases}$$
$$I_{01}^t=\begin{cases}1 & \text{if } a'U_t \leqq e_{ta} \text{ and } b'V_t > e_{tb} \\ 0 & \text{o.w.}\end{cases}$$
$$I_{11}^t=\begin{cases}1 & \text{if } a'U_t > e_{ta} \text{ and } b'V_t > e_{tb} \\ 0 & \text{o.w.}\end{cases}$$

$a'U_t$ and $b'V_t$ are individual bivariate dichotomous choice functions, e_{ta} and e_{tb} are random threshold levels of the choice, X_{it} 's $(i=1,2,3,4)$ are vectors of exogeneous explanatory variables with or without overlapping elements, β_1', β_2', β_3', β_4', a' and b' are unknown parameters, and $e'=(e_{1t},\ e_{2t},\ e_{3t},\ e_{4t},\ e_{ta},\ e_{tb})$ is a 6×1 unobservable correlated $N(0, \Sigma)$ random vectors, where

$$\Sigma=\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{1a} & \sigma_{1b} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{2a} & \sigma_{2b} \\ & & \sigma_3^2 & \sigma_{34} & \sigma_{3a} & \sigma_{3b} \\ \text{symm.} & & & \sigma_4^2 & \sigma_{4a} & \sigma_{4b} \\ & & & & 1 & \rho \\ & & & & & 1 \end{bmatrix}. \tag{1.2}$$

We desire a consistent and asymptotically normal estimate of (1.1). Nelson and Hahn (1972) pointed out that ordinary least squares produce inconsistent estimates of regression parameters if the dependent variables (Y_{it}'s) are censored or truncated. We refer to this inconsistency as the selectivity bias.

2. Maximum Likelihood Estimation Method

Let $e'=(e_{1t},\ e_{2t},\ e_{3t},\ e_{4t},\ e_{ta},\ e_{tb})$ have multinormal distribution given in (1.2), and $\{Y_t|I_{ij}^t=(Y_{1t}|I_{00}^t,\ Y_{2t}|I_{10}^t,\ Y_{3t}|I_{01}^t,\ Y_{4t}|I_{11}^t) : t=1,2,\ \cdots,\ T\}$ be a random sample of size T. Then the likelihood function for the model (1.1) is given by

$$L=\prod_{t=1}^{T}\left[\int_{b'V_t}^{\infty}\int_{a'U_t}^{\infty} f(Y_{1t}-\beta_1' X_{1t},\ e_{ta},\ e_{tb})\,de_{ta}de_{tb}\right]^{I_{00}^{t}}$$
$$\cdot\left[\int_{b'Vt}^{\infty}\int_{-\infty}^{a'U_t} g(Y_{2t}-\beta_2' X_{2t},\ e_{ta},\ e_{tB})\,de_{ta}de_{t}\right]^{I_{10}^{t}} \tag{2.1}$$
$$\cdot\left[\int_{-\infty}^{b'V_t}\int_{a'U_t}^{\infty} h(Y_{3t}-\beta_3' X_{3t},\ e_{ta},\ e_{tb})\,de_{ta}de_{tb}\right]^{I_{01}^{t}}$$
$$\cdot\left[\int_{-\infty}^{b'V_t}\int_{-\infty}^{a'U_t} l(Y_{4t}-\beta_4' X_{4t},\ e_{ta},\ e_{tb})\,de_{ta}de_{tb}\right]^{I_{11}^{t}}$$

where $f(\cdot)$, $g(\cdot)$, $h(\cdot)$ and $l(\cdot)$ are joint densities of random variables (e_{1t}, e_{ta}, e_{tb}), (e_{2t}, e_{ta}, e_{tb}), (e_{3t}, e_{ta}, e_{tb}) and (e_{4t}, e_{ta}, e_{tb}), respectively.

The maximum likelihood estimates of likelihood function (2.1) are obtained by maximizing (2.1) with respect to all parameters. But this procedure may be cumbersome in practice, because the likelihood function involves intractable double integrals of trivariate normal densities.

3. Bivariate Dichotomous Two-stage Estimation Method

A class of computationally simple estimators, called two-step or two-stage estimators gained popularity in recent years, particularly in the mixed continuous/discrete dependent variable case. The class of estimators alluded to is the class of M-estimators which satisfy Huber's condition (1981), and hence the two-stage estimators are consistent and asymptotically normal estimators. This assertion has been justified by the following theorem.

Theorem 1. (Duncan, 1982) Let $\hat{\theta}_1$ be a root of $\sum\psi_1(Z_i;\ \hat{\theta}_1)/N=0$ satisfying the conditions of Huber's Theorems 2.4 and 3.1 (Huber, 1981, p.131~133), and let $\hat{\beta}(\theta)$ be a root of $\sum\ \psi_2(Z_i;\theta,\hat{\beta})/N=0$ also satisfying the conditions of Huber's Theorems 2.4 and 3.1 for every θ. If $\psi_2(Z_i;\theta,\hat{\beta})$ is boundedly differentiable in θ, the two-stage estimator $(\hat{\theta}_1,\hat{\beta}(\hat{\theta}_1))$ satisfies the same conditions and hence is a consistent and asymptotically normal estimator.

3.1. Correct model eliminating selectivity bias

Lemma. (Anderson, 1958) For $X: p\times 1$, let $X\sim N\ (\theta,\Sigma)$, and X be partitioned as $X'=(Y',\ Z')$. Then the conditional distribution of Y given $(Z=z)$ is also normal with mean vector a linear function of z, and covariance matrix independent of z; that is, the conditional distribution of Y, given that $(Z=z)$ is

$$(Y|Z=z)\sim N(\theta_Y+\Sigma_{12}\Sigma_{22}^{-1}(z-\theta_Z),\ \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

where $\underset{1\times p}{\theta'}=(\underset{1\times q}{\theta_Y'},\ \underset{1\times r}{\theta_Z'})$, $\Sigma=\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$, Σ_{11} ; $q\times q$, Σ_{22} ; $r\times r$, Σ_{12} ; $q\times r$ and $p=q+r$.

Definition 1. A random variable X_1 and X_2 are said to have a truncated standard bivariate normal distribution, if the joint probability density function $g(x_1,\ x_2)$ is given by

$$g(x_1,\ x_2)=(2\pi\sqrt{1-\rho^2})^{-1}\cdot\exp\{-(x_1{}^2-2\rho x_1x_2+x_2{}^2)/2(1-\rho^2)\}/F(A,B,\rho),$$

where A and B are the truncation points, for X_1 and X_2, respectively and $F(A,B,\rho)$ is the normalizing constant.

Theorem 2. In the standardized form, the first moments about the origin of the X_1 are

$$F_1(A,B,\rho)EX_1=-\phi(A)\Phi(l^*)-\rho\phi(B)\Phi(u^*)$$
$$F_2(A,B,\rho)EX_1=\phi(A)\Phi(l^*)-\rho\phi(B)\{1-\Phi(u^*)\}$$
$$F_3(A,B,\rho)EX_1=-\phi(A)\{1-\Phi(l^*)\}+\rho\phi(B)\Phi(u^*)$$
$$F_4(A,B,\rho)EX_1=\phi(A)\{1-\Phi(l^*)\}+\rho\phi(B)\{1-\Phi(u^*)\},$$

where

$$l^*=(B-\rho A)/\sqrt{1-\rho^2},\quad u^*=(A-\rho B)/\sqrt{1-\rho^2},$$
$$F_1(A,B,\rho)=P(X_1<A,\ X_2<B)=\int_{-\infty}^{A}\int_{-\infty}^{B}f(x_1,\ x_2)\,dx_1dx_2,$$
$$F_2(A,B,\rho)=P(X_1>A,\ X_2<B)=\int_{A}^{\infty}\int_{-\infty}^{B}f(x_1,\ x_2)\,dx_1dx_2,$$

similarly, $F_3(A,B,\rho)=P(X_1<A,\ X_2>B)$, $F_4(A,B,\rho)=P(X_1>A,\ X_2>B)$, $f(x_1,\ x_2)$ is a p.d.f. of bivariate standard normal distribution, and ϕ and Φ are p.d.f. and c.d.f. of standard normal distribution, respectively. We can easily calculate EX_2 by using symmetry of bivariate normal distribution. Proof of Theorem 2 is done by strightforward intergration and is therefore omitted.

We may reparameterize the model (1.1) by using Lemma and Theorem 2 such that the reparameterized model is

$$Y_{1t}=\beta_1'X_{1t}-\sigma_{1a}\phi(a'U_t)\Phi(l^*)/F_1(a'U_t,\ b'V_t;\ \rho)$$
$$-\sigma_{1b}\phi(b'V)\Phi(U^*)/F_1(a'U_t,\ b'V_t;\ \rho)+\eta_{1t}\qquad \text{for } I_{11}^t=1$$
$$Y_{2t}=\beta_2'X_{2t}-\sigma_{2a}\phi(a'U_t)\{1-\Phi(l^*)\}/F_2(a'U_t,\ b'V_t;\ \rho)$$
$$+\sigma_{2b}\phi(b'V_t)\Phi(u^*)/F_2(a'U_t,\ b'V_t;\ \rho)+\eta_{2t}\qquad \text{for } I_{10}^t=1 \quad (3.1)$$
$$Y_{3t}=\beta_3'X_{3t}+\sigma_{3a}\phi(a'U_t)\Phi(l^*)/F_3(a'U_t,\ b'V_t;\ \rho)$$
$$-\sigma_{3b}\phi(b'V_t)\{1-\Phi(u^*)\}/F_3(a'U_t,\ b'V_t;\ \rho)+\eta_{3t}\quad \text{for } I_{01}^t=1$$
$$Y_{4t}=\beta_4'X_{4t}+\sigma_{4a}\phi(a'U_t)\{1-\Phi(l^*)\}/F_4(a'U_t,\ b'V_t;\ \rho)$$

$$+\sigma_{4b}\phi(b'V_t)\{1-\Phi(u^*)\}/F_4(a'U_t,\ b'V_t;\rho)+\eta_{4t} \qquad \text{for } I_{00}^t=1$$

where $E[\eta_{it}]=0,\ i=1,2,3,4,\ t=1,2,\cdots,T,\ u^*=(a'U_t-b'V_t\cdot\rho)/\sqrt{1-\rho^2}$, and $l^*=(b'V_t-a'U_t\cdot\rho)/\sqrt{1-\rho^2}$.

3.2. First stage of the estimation (Estimation of a, b, ρ or c)

The probability model dealt with in this section is the modified model from bivariate dichotomous case of normal models proposed by Ashford and Sowden (1970) and Morimune (1979). Using the same symbols as the model (1.1), the modified model is as follows:

$$\begin{aligned} &P_{11}(t)=F(a'U_t,\ b'V_t,\ c'W_t),\\ &P_{1\cdot}(t)=P_{11}(t)+P_{10}(t)=\Phi(a'U_t),\\ &P_{\cdot 1}(t)=P_{11}(t)+P_{01}(t)=\Phi(b'V_t),\\ &|c'W_t|\leqq 1, \end{aligned} \tag{3.2}$$

where Φ is the distribution function of a standard normal variable, and F is the cumulative distribution function of the bivariate standard normal variable with correlation coefficient $c'W_t$. If the chi-square test, proposed by Amemiya (1974), accepts the null hypothesis that ρ is constant over t, we may replace $c'W_t$ by constant ρ as in Ashford-Sowden model.

Theorem 3. Modified Full Information Minimum Chi-square (FIMC) Probit estimators $\hat{a}$, $\hat{b}$ and c are obtained by

$$\hat{\gamma}=\begin{pmatrix}\hat{a}\\ \hat{b}\\ \hat{c}\end{pmatrix}=\left(\sum_{t=1}^{T}X_t^{*\prime}\hat{\Sigma}_t^{-1}X_t^{*}\right)^{-1}\sum_{t=1}^{T}X_t^{*\prime}\hat{\Sigma}_t^{-1}Y_t^{*}, \tag{3.3}$$

Theorem 4. Asymptotic distribution of $\hat{\gamma}$ is as follows.

$$\mathscr{L}\{\sqrt{T}(\hat{\gamma}-\gamma)\} \overset{\text{asym.}}{\rightsquigarrow} N\left(0,\ \left[\sum_{t=1}^{T}X_t^{*\prime}\hat{\Sigma}_t^{-1}X_t^{*}/T\right]^{-1}\right).$$

Proofs. The proof of Theorem 4 is essentially the same as that of Amemiya (1974) and is omitted. To prove Theorem 3 we modify the derivation of FIMC (Full Information Minimum Chi-square) probit estimator proposed by Amemiya (1974).

Define $r_{10}(t)=\sum_{t=1}^{nt}I_{10}^{tk},\ r_{01}(t)=\sum_{t=1}^{nt}I_{01}^{tk},\ r_{11}(t)=\sum_{t=1}^{nt}I_{11}^{tk},$

$$\hat{p}_{10}(t)=r_{10}(t)/n_t,\ \hat{p}_{01}(t)=r_{01}(t)/n_t,\ \hat{p}_{11}(t)=r_{11}(t)/n_t,$$

$$\hat{p}_{1\cdot}(t)=\hat{p}_{10}(t)+\hat{p}_{11}(t) \text{ and } \hat{p}_{\cdot 1}(t)=\hat{p}_{01}(t)+\hat{p}_{11}(t).$$

By a Taylor expansion of $\Phi^{-1}\{p_{1\cdot}(t)\}$ around $p_{1\cdot}(t)$ we have approximately

$$\Phi^{-1}\{\hat{p}_{1\cdot}(t)\}\simeq a'U_t+[\hat{p}_{1\cdot}(t)-p_{1\cdot}(t)]/\phi(a'U_t), \tag{3.4}$$

where ϕ is the density function of a standard normal variable. Similarly, we have

$$\Phi^{-1}\{\hat{p}_{\cdot 1}(t)\}\simeq b'V_t+[\hat{p}_{\cdot 1}(t)-p_{\cdot 1}(t)]/\phi(b'V_t). \tag{3.5}$$

Next, we solve the model (3.2) for $c'W_t$ as a function of $p_{11}(t)$, $p_{1\cdot}(t)$ and $p_{\cdot 1}(t)$, and call that function G. Thus we have

$$c'W_t=G\{p_{11}(t),\ p_{1\cdot}(t),\ p_{\cdot 1}(t)\}.$$

By a Taylor expansion of $G\{\hat{p}_{11}(t),\ \hat{p}_{1\cdot}(t),\ \hat{p}_{\cdot 1}(t)\}$ around $G\{p_{11}(t),\ p_{1\cdot}(t),\ p_{\cdot 1}(t)\}$ we have

$$\begin{aligned}G\{p_{11}(t),\ \hat{p}_{1\cdot}(t),\ \hat{p}_{\cdot 1}(t)\}\simeq c'W_t&+G_{1t}\{\hat{p}_{11}(t)-p_{11}(t)\}\\ &+G_{2t}\{\hat{p}_{1\cdot}(t)-p_{1\cdot}(t)\}+G_{3t}\{\hat{p}_{\cdot 1}(t)-p_{\cdot 1}(t)\},\end{aligned} \tag{3.6}$$

where $G_{1t}=1/f_t$, $G_{2t}=-\Phi\{(b'V_t-c'W_t\cdot a'U_t)/\sqrt{1-(c'W_t)^2}\}/f_t$, $G_{3t}=-\Phi\{(a'U_t-b'V_t\cdot c'W_t)/\sqrt{1-(c'W_t)^2}\}/f_t$, and $f_t=f(a'U_t,\ b'V_t,\ c'W_t)$ is the standard bivariate normal density function with correlation coefficient $c'W_t$.

We rewrite equations (3.4), (3.5) and (3.6) as

$$\begin{aligned}&Q_t=a'U_t+e_t^1\\ &S_t=b'V_t+e_t^2\\ &P_t=c'W_t+e_t^3,\quad t=1,2,\ \cdots,\ T,\end{aligned} \tag{3.7}$$

where the definitions of the new symbols are obvious.

The weighted least squares method for multivariate regression (3.7) leads to

$$\begin{pmatrix}\hat{a}\\ \hat{b}\\ \hat{c}\end{pmatrix}=\left(\sum_{t=1}^{T}X_t^{*\prime}\hat{\Sigma}_t^{-1}X_t^*\right)^{-1}\sum_{t=1}^{T}X_t^{*\prime}\hat{\Sigma}_t^{-1}Y_t^*,$$

where $X_t^*=\begin{bmatrix}U_t & & 0\\ & V_t & \\ 0 & & W_t\end{bmatrix}$ $Y_t^*=(Q_t,\ S_t,\ P_t)'$, $\Sigma_t=\mathrm{MDM}'$,

$$M=\begin{bmatrix}1/(a'U_t) & 0 & 0\\ 0 & 1/(b'V_t) & 0\\ G_{2t} & G_{3t} & G_{1t}\end{bmatrix},\quad D=1/T\begin{bmatrix}b_{11} & b_{12} & b_{13}\\ & b_{22} & b_{23}\\ \text{symm.} & & b_{33}\end{bmatrix},$$

$$b_{11}=\{p_{11}(t)+p_{10}(t)\}\ \{p_{01}(t)+p_{00}(t)\},$$

$$b_{12}=p_{11}(t)-\{p_{11}(t)+p_{10}(t)\}\ \{p_{11}(t)+p_{01}(t)\},$$

$$b_{22}=\{p_{11}(t)+p_{01}(t)\}\ \{(p_{10}(t)+p_{00}(t)\},$$

$$b_{13}=p_{11}(t)\ \{p_{01}(t)+p_{00}(t)\},\quad b_{23}=p_{11}(t)\ \{p_{10}(t)+p_{00}(t)\},$$

$$b_{33}=p_{11}(t)\{1-p_{11}(t)\},$$

and $\hat{\Sigma}_t^{-1}$ is obtained by replacing the population parameters in Σ_t^{-1} with their sample estimates.

3.3. Second-stage of the estimation method (Estimation of β_i's, $i=1,2,3,4$)

Construct the equations by substituting the modified FIMC probit estimators $\hat{a}$, $\hat{b}$ and $\hat{c}$ for a, b and c in equations (3.1), such that

$$\begin{aligned}
Y_{1t}&=\beta_1'X_{1t}-\sigma_{1a}\phi(\hat{a}'U_t)\Phi(l^*/F_1(\hat{a}'U_t,\ \hat{b}'V_t,\ \hat{c}'W_t)\\
&\quad-\sigma_{1b}\phi(\hat{b}'V_t)\Phi(\hat{u}^*)/F_1(\hat{a}'U_t,\ \hat{b}'V_t,\ \hat{c}'W_t)+\eta_{1t}\\
Y_{2t}&=\beta_2'X_{2t}-\sigma_{2a}\phi(\hat{a}'U_t)\{1-\Phi(l^*)\}/F_2(\hat{a}'U_t,\ \hat{b}'V_t,\ \hat{c}'W_t)\\
&\quad+\sigma_{2b}\phi(\hat{b}'V_t)\Phi(u^*)/F_2(\hat{a}'U_t,\ \hat{b}'V_t,\ \hat{c}'W_t)+\eta_{2t}\\
Y_{3t}&=\beta_3'X_{3t}-\sigma_{3a}\phi(\hat{a}'U_t)\Phi(l^*)/F_3(\hat{a}'U_t,\ \hat{b}'V_t,\ \hat{c}'W_t)\\
&\quad+\sigma_{3b}\phi(\hat{b}'V_t)\{1-\Phi(\hat{u}^*)\}/F_3(\hat{a}'U_t,\ \hat{b}'V_t,\ \hat{c}'W_t)+\eta_{3t}\\
Y_{4t}&=\beta_4'X_{4t}+\sigma_{4a}\phi(\hat{a}'U_t)\{1-\Phi(l^*)\}/F_4(\hat{a}'U_t,\ \hat{b}'V_t,\ \hat{c}'W_t)\\
&\quad+\sigma_{4b}\phi(\hat{b}'V_t)\{1-\Phi(\hat{u}^*)\}/F_4(\hat{a}'U_t,\ \hat{b}'V_t,\ \hat{c}'W_t)+\eta_{4t},
\end{aligned}\tag{3.8}$$

where $E[n_{it}]=0,\ i=1,2,3,4,\ t=1,2,\ \cdots,\ T,$

$$\hat{u}^*=(\hat{a}'U_t-\hat{b}'V_t\cdot\hat{c}'W_t)/\sqrt{1-(\hat{c}'W_t)^2},$$

$$l^*=(\hat{b}'V_t-\hat{a}'U_t\cdot\hat{c}'W_t)/\sqrt{1-(\hat{c}'W_t)^2},$$

$$\begin{aligned}
Var(\eta_{1t})&=E[e_{1t}^2|e_{ta}\leqq\hat{a}'U_t,\ e_{tb}\leqq\hat{b}'V_t]-\{E[e_{1t}|e_{ta}\leqq\hat{a}'U_t,\ e_{tb}\leqq\hat{b}'V_t]\}^2,\\
Var(\eta_{2t})&=(E[e_{2t}^2|e_{ta}\leqq\hat{a}'U_t,\ e_{tb}>\hat{b}'V_t]-\{E[e_{2t}|e_{ta}\leqq\hat{a}'U_t,\ e_{tb}>\hat{b}'V_t]\}^2,\\
Var(\eta_{3t})&=E[e_{3t}^2|e_{ta}>\hat{a}'U_t,\ e_{tb}\leqq\hat{b}'V_t]-\{E[e_{3t}|e_{ta}>\hat{a}'U_t,\ e_{tb}\leqq\hat{b}'V_t]\}^2,\\
\text{and}\quad Var(\eta_{4t})&=E[e_{4t}^2|e_{ta}>\hat{a}'U_t,\ e_{tb}>\hat{b}'V_t]-\{E[e_{4t}|e_{ta}>\hat{a}'U_t,\ e_{tb}>\hat{b}'V_t]\}^2.
\end{aligned}\tag{3.9}$$

Finally, we get consistent estimates $\hat{\beta}_i$, $\hat{\sigma}_{ia}$, $\hat{\sigma}_{ib}^2$ and $\hat{\sigma}_i^2$ $(i=1,2,3,4)$ by applying O.L.S method for each equations (3.8). Here, we take M.S.E. of each equations as $\hat{\sigma}_i^2$.

To get more efficient estimators, we may proceed a numerical iterative maximum likelihood procedure to the likelihood function (2.1), taking these consistent estimates as a set of initial points. But computationally simpler Aitken estimators (Generalized Least Squares) are readily obtained using estimates of $Var(\eta_{it})$'s. We may obtain consistent estimates of $Var(\eta_{it})$'s by replacing consistent estimates $\hat{\sigma}_{ia}$, $\hat{\sigma}_{ib}$ and $\hat{\sigma}_i^2$ into the equations (3.9) such that

$$\hat{Var}(\eta_{it})=Var(\eta_{it})\,|\,(\hat{\sigma}_{ia},\ \hat{\sigma}_{ib},\ \hat{\sigma}_i^2),\ i=1,2,3,4.$$

See Kim (1984) for explicit parameteric form of the equations (3.9).

Theorem 5. The bivariate dichotomous two-stage estimator ($\hat{\gamma}$, $\hat{\beta}_i(\hat{\gamma})$) is a consistent and asymptotically normal estimator.

Proof For this proof let $\sum\psi_1(I^t, X_t : \hat{\gamma})/T=0$ be the normal equation of (3.3), and let $\sum\psi_2(Y_t, X_t : \hat{\beta}_i(\hat{\gamma}))/T$ be one of the normal equation of the equations (3.8).

Theorem 4 says $\hat{\gamma}$ is a consistent and asymptotically normal, and $\hat{\beta}_i$ (γ) is the Aitken's estimator which is also consistent and asymptotically normal for every γ. Hence, $\hat{\gamma}$ and $\hat{\beta}_i$ (γ) satisfy the conditions of Huber's Theorem 2.4 and 3.1. Consequently, the result of Theorem 1 immediately completes the proof of the Theorem 5. See Duncan (1982) for the detailed proof.

4. A Data Application

In this section we apply an economic data set to the model (1.1). We take the data from "Panel Study of Income Dynamics" (Survey Research Center, University of Michigan). For the adoption of the model, we define, the bivariate dichotomous discrete variable, $I_{11}^t=1$ if the i-th family in the t-th income group owns a house with more than five rooms and 0 if not; $I_{10}^t=1$ if the same family owns a house with less than five rooms, and so on. We use a constant term and a weighted mean of income in each group for the elements of U_t and V_t, a weighted mean of housing expense in each group for the dependent variable (Y_t), and the constant term and the weighted mean of income in each group for the elements of X_t. Our purpose of data analysis is to get an efficient estimate of the marginal propensity of housing expense (β_i) for each group, $i=1, 2, 3, 4$.

Table 1 includes estimated values by the bivariate dichotomous probit model (3.2). Table 2 tabulates all the estimated coefficients, standard errors, and t values for the probit model. In Table 3, the estimated values of parameters of the model (1.1) are tabulated and compared to the estimated values of parameters under the model without consideration of selectivity bias. To show that selectivity biases against β 's are an appreciable disadvantage of the usual least square method under the linear model with no consideration of selectivity bias, we apply the working rule discussed in Cochran (1977). The working rule says that the effect of bias on the accuracy of an estimate is negligible if the bias less than one tenth of the standard deviation of the estimate.

Table 1 Numbers of Households Responding for Economic Attributes in Terms of Annual Income

Houseowner / Number of rooms / Income	Yes		No		Total
	Yes (>5)	No (≤5)	Yes (>5)	No (≤5)	
0—5000 (3500)	41 (48) [420]	58 (72) [401]	36 (35) [527]	123 (103) [463]	258
5000—6500 (6000)	63 (60) [675]	69 (68) [702]	29 (31) [876]	75 (77) [877]	236
6500—8000 (7350)	85 (86) [812]	93 (80) [799]	29 (35) [1021]	70 (77) [1023]	277
8000—9500 (8950)	87 (79) [923]	64 (62) [987]	24 (26) [1336]	45 (52) [1136]	220
9500—11500 (10600)	101 (98) [1218]	64 (64) [1236]	29 (25) [6180]	41 (47) [1328]	235
11500—13500 (12750)	99 (105) [1596]	57 (58) [1527]	24 (22) [2036]	43 (38) [1720]	236
13500—16000 (14750)	151 (143) [1840]	61 (65) [1763]	25 (22) [2326]	29 (36) [1936]	266
16000—20000 (17550)	161 (157) [2036]	55 (53) [1983]	15 (16) [2663]	19 (24) [2168]	250
20000—30000 (23000)	185 (196) [2789]	38 (36) [2513]	9 (8) [3268]	20 (12) [2880]	252
Total	973	559	220	465	2217

$\chi^2_{(21)}=26.9$ p-Value≒.15

Note: Numbers in parentheses are estimated cell frequencies by the Bivariate Dichotomous Probit model.
Numbers in brackets are housing expenditures.

Table 2 Estimated Coefficients

	Coeff.	S.D	t-value
a_1 (income)	0.7681	0.0535	14.4
a_2 (constant)	−0.3564	0.0651	5.5
b_1 (income)	0.0684	0.0274	25
b_2 (constant)	−0.7005	0.0630	11
c_1 (income)	0.1651	0.015	11
c_2 (constant)	0.1923	0.073	2.7

Note: Coefficients and standard errors of income variables, a_1, b_1 and c_1 are multiplied by 10^4. This table is an excerpt from Table *V* of Morimune (1979).

Tabble 3 Estimated Coefficients and Their *t*-value

	Estimated values of:							
	β_{01}	β_1	β_{02}	β_2	β_{03}	β_3	β_{04}	β_4
With consideration of self-selectivity	94	0.12114 [.024]*	294.1	0.08203 [.01896]*	—	0.12368 [.01882]*	234.6	0.14618 [.02074]*
	(0.13)	(5.05)	(0.70)	(4.33)	(—)	(6.57)	(1.82)	(7.05)
Without consideration of self-selectivity (L.S. estimates)	−64.6	0.12389	35.18	0.111004	—	0.14981	93.96	0.121439
	(−1.9)	(29.5)	(0.78)	(31.99)	(—)	(59.19)	(2.14)	(35.83)

Note: Numbers in the parentheses are t-values. Numbers in the brackets are standard deviations. * denotes significance of selectivity bias ($R>.1$).

5. Concluding Remarks

We considered a data set generated by the bivariate dichotomous choice by individuals. This paper develops a model of how such data are generated, and derives a simple estimation procedure. Suggested model unifies censored regression models and discrete choice models. We have shown that the bivariate dichotomous two-stage estimators can be obtained even in the cases where the corresponding maximum likelihood estimators are computationally burdensome. Asymptotic properties of the two-stage estimators are proved to be consistent and asymototically normal.

6. Acknowledgement

The detailed and constructive comments of referees were very useful in making the paper more readable and in correcting some errors. Their efforts are greatly appreciated.

References

(1) Amemiya, T. (1974). Bivariate Probit Analysis, *J. Amer. Statist. Assoc.*, Vol. 69, 940—944.

(2) Anderson, T.W. (1958). *An Introduction io Multivariate Statistical Analysis*, Wiley, New York.

(3) Ashford, J.R. and Sowden, R.R. (1970). Multivariate Probit Analysis, *Biometrika*, Vol. 26, 535—546.

(4) Cochran, W.G. (1977). *Sampling Techniques*, Wiley, New York.

(5) Duncan, G.M. (1982). Asymptotic Properties of Two-Step Estimators with Applications to Sample Selectivity Corrections, Working Paper No. 1182—2, Washington State University.

(6) Huber, P. (1981). *Robust Statistics*, Wiley, New York.

(7) Kim, H.J. (1984). Bivariate Dichotomous Selectivity Models: Formulation and Estimation Procedures, Ph. D. Dissertation, University of California, Riverside.

(8) Lee, L.F. (1979). Identification and Estimation in Binary Choice Models with Limited (Censored) Denpendent Variables, *Econometrica*, Vol. 47, 977—996.

(9) Morimune, K. (1979). Comparison of normal and logistic models in the bivariate dichotomus analysis, Econometrica, Vol. 47, 957∼975.

(10) Nelson, W and Hahn, G.J. (1972). Linear Estimation of a Regressoin Relationship from Censored Data, *Technometrics*, Vol. 14. 247—269.