ECONOMETRIC REVIEWS, 15(4), 387–400 (1996)

ESTIMATION OF SAMPLE SELECTION BIAS MODELS

Kazumitsu Nawata and Nobuko Nagase

Department of Social and International Relations
University of Tokyo
Komaba, Meguro-ku
Tokyo 153, Japan

Faculty of Economics
Toyo University
Hakusan, Bunkyo-ku
Tokyo 112, Japan

Key Words and Phrases: Sample-Selection Bias; Type II Tobit; Heckman's Two-Step Estimator; Maximum Likelihood Estimator.

ABSTRACT

Econometric models with sample selection biases are widely used in various fields of economics, such as labor economics. The Maximum Likelihood Estimator (MLE) is seldom used to estimate models because of computational difficulty, while Heckman's two-step estimator is widely used to estimate these models. However, Heckman's two-step estimator sometimes performs poorly. In this paper, methods of calculating the MLE are analysed, and finite sample properties of the MLE and Heckman's two-step estimator are compared using Monte Carlo experiments and empirical examples.

1. INTRODUCTION

Econometric models with sample selection biases (Type II Tobit models) are widely used in various fields of economics, such as labor economics. For examples and details of the models, see Gronau (1973), Heckman (1974), Maddala (1983), Amemiya (1985), and Mroz (1987). Unlike other types of econometric models, the maximum likelihood estimator (MLE) has been seldom used because of its computational difficulty. Heckman (1976, 1979) proposed a two-step estimator, which can be calculated by the probit maximum likelihood and least squares methods. Heckman's estimator has been widely used because of its computational ease. However, Heckman's estimator sometimes performs poorly (Wales and Woodland (1980), Nelson (1984), Paarsch (1984), and Nawata (1993, 1994)).

Several econometric software package programs such as LIMDEP, STATA, and TSP, which are widely used in the estimation of limited dependent variable models, have routines which calculate the MLE of the sample-selection bias models. However, these methods are incomplete. The problems are (i) the iterations do not always converge, and (ii) because of existence of local maxima, the results may not be the true MLE for some data sets (Olsen (1982)).

In this paper, the problems of Heckman's estimator are analyzed and cases where Heckman's estimator performs poorly are pointed out. In such cases, Heckman's estimator should not be used (this means that numerous studies present inaccurate results). The MLE and Heckman's two-step estimator are compared using Monte Carlo experiments and empirical examples. Since the methods employed in the various packages are incomplete, a scanning method is used to calculate the correct MLE.

2. MODELS

The model considered in this paper is

$$y_{1i} = x_i'\beta + u_i,$$

$$(2.1)\qquad y_{2i} = w_i\alpha + v_i,$$

$$d_i = 1(y_{2i} > 0), \qquad i=1,2,\ldots,N.$$

where $1(\cdot)$ is an indicator function such that $1(\cdot) = 1$ if $\cdot$ is true and 0 otherwise. y_{2i} is not observable and only the sign of y_{2i} (i.e. d_i) is observed. y_{1i} is observed when d_i=1.

(u_i, v_i) are jointly normal with mean zero, variances $(\sigma_1^2, 1)$, and covariance σ_{12}. (Since the sign of y_{2i} does not change if we multiply it by a positive constant, it can be assumed that the variance of v_i is 1 without loss of generality.)

3. HECKMAN'S TWO-STEP ESTIMATOR

The conditional expected value of y_{1i} given $d_i = 1$ is

$$(3.1)\qquad E(y_{1i}|d_i = 1) = x_i'\beta + \sigma_{12}\lambda(w_i'\alpha),$$

$$\lambda(z) = \phi(z)/\Phi(z),$$

where ϕ and Φ are the density and distribution functions of the standard normal distribution, respectively. λ is the hazard ratio or reciprocal of the Mill's ratio. It is possible to write

$$(3.2)\qquad y_{1i} = x_i'\beta + \sigma_{12}\lambda(w_i'\alpha) + \epsilon_i,$$

for i such that $d_i = 1$. Since $E(\epsilon_i|\ d_i = 1) = 0$, (3.2) can be used to estimate β.

Heckman's two-step estimator is based on (3.2). First, α is estimated by the probit maximum likelihood method, and then α is replaced by the probit MLE, $\tilde{\alpha}$. Finally, β is estimated by the ordinary least squares (OLS) method. Since Heckman's two-step estimator is easy to calculate, it is widely used to estimate models with sample selection biases, as in (2.1). The performance of the estimator depends on the properties of the hazard ratio $\lambda(z)$, which is closely approximated by a linear function of z over reasonable ranges of z. Therefore, if $w_i'\tilde{\alpha}$ and x_i are highly correlated, there is almost always a high degree of multicollinearity between $\lambda(w_i'\tilde{\alpha})$ and x_i. Hence Heckman's estimator may not perform well in these cases. For details, see Olsen (1980) and Nawata (1993).

4. MAXIMUM LIKELIHOOD ESTIMATOR

Unlike other econometric models, the MLE is seldom used in these models because of its computational difficulty. The log-likelihood function is given by

$$(4.1)\qquad \log L = \sum_{i=1}^{n}\Big[(1-d_i)\cdot\log[1-\Phi(w_i'\alpha)] + d_i\Big[\log\Phi\{[w_i'\alpha + \sigma_{12}\sigma_1^{-2}(y_{1i} - x_i'\beta)][1 - \sigma_{12}^2/\sigma_1^2]^{-1/2}\} - \log\sigma_1 + \log\phi[\sigma_1^{-1}(y_{1i} - x_i'\beta)]\Big]\Big].$$

Setting $\rho = \sigma_{12}/\sigma_1$, the log-likelihood function becomes

$$(4.2)\qquad \log L = \sum_{i=1}^{n}\Big[(1-d_i)\cdot\log[1-\Phi(w_i'\alpha)] + d_i\Big[\log\Phi\{[w_i'\alpha + \rho\sigma_1^{-1}(y_{1i} - x_i'\beta)][1 - \rho^2]^{-1/2}\} - \log\sigma_1 + \log\phi[\sigma_1^{-1}(y_{1i} - x_i'\beta)]\Big]\Big].$$

If the MLE is calculated by the standard methods, such as ones used in LIMDEP, STATA, and TSP, the following problems exist.

(i) The procedure does not always converge.

(ii) Because of the existence of local maxima (Olsen (1982)), the results may not be correct even if the procedure converges.

Olsen (1982) proved that *log L* has a unique maximum for given values of ρ. In this paper, the scanning method is modified and the following procedure is used to calculate the MLE.

i) Choose M_1 equidistant points from (-1, 1). Let δ be the distance between any two points.

ii) Let $\rho = 0$ and calculate $\hat{\alpha}_0$, $\hat{\beta}_0$, and $\hat{\sigma}_0$, which maximise the conditional likelihood function by the Newton-Raphson method. Note that this estimator is the same as the OLS estimator and the probit MLE of the first and second equations in (2.1).

iii) Let $\hat{\alpha}_j$, $\hat{\beta}_j$, and $\hat{\sigma}_j$ be the j-th estimators. Increase ρ by δ, choose the initial values of the iteration as $\hat{\alpha}_j$, $\hat{\beta}_j$, and $\hat{\sigma}_j$, and calculate the (j+1)-th estimator. Since the likelihood function is a continuous function of ρ, the previous estimators are in the neighbourhood of the maximum values.

iv) Continue (iii) and calculate estimators up to the largest values of ρ determined in (i).

v) In the same way, calculate estimators from 0 to the smallest values of ρ.

vi) Choose the value of ρ which maximises the conditional likelihood function.

vii) If necessary, choose M_2 points in the neighbourhood of the value determined in (vi), and repeat the procedure.

viii) Determine the final estimators, $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}_1$, and $\hat{\rho}$

The degree of difficulty in calculating the MLE depends on the true parameter value ρ_0. If ρ_0 is close to 0 (which means the effect of the sample-selection bias is small), the log-likelihood function behaves "well" in wide ranges of ρ, so it is easy to compute the MLE. However, if ρ_0 is close to ± 1, the behavior of the log-likelihood function is asymmetric and it is difficult to determine the MLE. If ρ_0 is close to 1, the log-likelihood function behaves well on the left side of $\hat{\rho}$. However, on the right side of $\hat{\rho}$, $|\partial \log L/\partial\rho|$ becomes large quite rapidly as ρ increases. This means that we should choose δ in (i) sufficiently small to calculate the MLE.

5. MONTE CARLO COMPARISON OF THE MLE AND HECKMAN'S ESTIMATOR

In this section, some Monte Carlo results are presented for the MLE and Heckman's two-step estimator. The basic model is given by

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + u_i, \\ d_i &= 1(\alpha_0 + \alpha_1 w_i + v_i > 0), \quad i = 1,2,3,\ldots,n. \end{aligned} \tag{5.1}$$

y_i is observable if $d_i = 1$.

The following items are considered in the Monte Carlo study.

i) The effect of the correlation of x_i and w_i.

ii) The effect of the correlation of u_i and v_i.

The values of the exogenous variables are determined as follows.

$$\begin{aligned} w_i &= \xi_{1i}, \\ x_i &= \{\pi\xi_{1i} + (1-\pi)\xi_{2i}\}/\sqrt{\pi^2 + (1-\pi)^2}. \end{aligned} \tag{5.2}$$

$\{\xi_{1i}\}$ and $\{\xi_{2i}\}$ are i.i.d. random variables distributed uniformly on (0, 20]. $\pi/\sqrt{\pi^2 + (1-\pi)^2}$ is the correlation coefficient of w_i and x_i, and $\pi = 0, 0.5, 0.8, 0.9, 0.95$, and 1.0 are considered.

$\{(u_i, v_i)\}$ are jointly normal and determined as follows.

$$\begin{aligned} v_i &= \epsilon_{1i}, \\ u_i &= 10\cdot\{\rho_0\epsilon_{1i} + (1-\rho_0)\epsilon_{2i}\}/\sqrt{\rho_0^2 + (1-\rho_0)^2}. \end{aligned} \tag{5.3}$$

$\{\epsilon_{1i}\}$ and $\{\epsilon_{2i}\}$ are independent standard normal variables. $\rho_0/\sqrt{\rho_0^2 + (1-\rho_0)^2}$ is the correlation coefficient of u_i and v_i, and $\rho_0 = 0, 0.2, 0.4, 0.6$, and 0.8 are considered.

The true parameter values are:

$$\alpha_0 = -1.0, \quad \alpha_1 = 0.1, \quad \beta_0 = -10.0, \textit{ and } \beta_1 = 1.0. \tag{5.4}$$

Since the degrees of censoring are close to 50% in many important studies (for example, see Mroz (1987)), the degree of censoring is 50%. The sample size is $n = 200$ for all cases, and the number of trials is 1000 for all cases. The MLE is calculated using the scanning methods described in Section 4.

The results of the estimates of β_1 by Heckman's estimator and the MLE are presented in Table 1. Note that the following notation is used in the table.

S.D.:	Standard Deviation,	25%:	25% Percentile,
50%:	50% Percentile (Median), and	75%:	75% Percentile.

The biases of Heckman's two-step estimator are quite small for all cases. The biases of the MLE are larger in some cases and they are as large as 0.31 for the $\rho_0 = 0.4$ and $\pi = 1.0$ case.

Table 1 Estimates of β_1 by Heckman's Estimator and the MLE

		Heckman					MLE				
		Mean	S.D.	25%	50%	75%	Mean	S.D.	25%	50%	75%
π = 0.0											
	ρ_0 = 0.0	1.01	0.17	0.89	1.01	1.13	0.99	0.18	0.86	0.98	1.10
	0.2	0.98	0.15	0.89	0.98	1.08	0.99	0.17	0.87	0.99	1.10
	0.4	1.00	0.12	0.92	1.00	1.08	1.00	0.16	0.89	1.00	1.10
	0.6	1.00	0.10	0.93	1.01	1.08	1.01	0.13	0.93	1.02	1.10
	0.8	1.00	0.10	0.93	1.00	1.06	1.01	0.10	0.95	1.01	1.07
π = 0.5											
	ρ_0 = 0.0	0.99	0.35	0.74	1.00	1.23	0.99	0.24	0.81	0.98	1.14
	0.2	1.00	0.28	0.80	1.01	1.19	0.98	0.23	0.81	0.97	1.14
	0.4	1.01	0.23	0.85	1.01	1.16	0.98	0.20	0.85	0.97	1.11
	0.6	1.00	0.19	0.86	1.01	1.14	1.00	0.16	0.90	1.00	1.10
	0.8	1.01	0.21	0.88	1.01	1.16	1.01	0.12	0.93	1.01	1.09
π = 0.8											
	ρ_0 = 0.0	1.12	0.90	0.52	1.15	1.65	0.95	0.44	0.64	0.91	1.29
	0.2	0.99	0.68	0.55	1.01	1.45	0.90	0.42	0.61	0.92	1.21
	0.4	0.99	0.54	0.60	1.00	1.34	0.84	0.39	0.55	0.92	1.10
	0.6	1.00	0.48	0.69	0.99	1.30	0.90	0.42	0.61	0.92	1.21
	0.8	0.97	0.48	0.65	0.97	1.32	1.00	0.16	0.89	1.01	1.09
π = 0.9											
	ρ_0 = 0.0	0.98	1.40	-0.05	1.00	1.89	0.94	0.48	0.58	0.91	1.26
	0.2	1.00	1.15	0.26	0.96	1.76	0.83	0.47	0.48	0.80	1.21
	0.4	0.93	0.88	0.31	0.96	1.52	0.74	0.46	0.39	0.80	1.08
	0.6	0.99	0.81	0.44	0.98	1.57	0.88	0.34	0.76	0.95	1.09
	0.8	1.00	0.74	0.49	0.98	1.52	0.99	0.17	0.88	1.00	1.11
π = 0.95											
	ρ_0 = 0.0	1.06	1.89	-0.28	1.11	2.43	0.95	0.47	0.60	0.94	1.29
	0.2	1.06	1.47	0.04	1.01	1.96	0.81	0.48	0.43	0.79	1.19
	0.4	1.20	1.20	0.23	0.98	1.74	0.70	0.47	0.31	0.73	1.08
	0.6	1.10	1.09	0.38	1.09	1.72	0.84	0.38	0.71	0.94	1.08
	0.8	0.98	1.10	0.25	0.98	1.76	0.99	0.17	0.88	1.00	1.11
π = 1.0											
	ρ_0 = 0.0	0.98	2.08	-0.32	0.92	2.23	0.92	0.47	0.58	0.93	1.25
	0.2	0.97	1.70	-0.18	0.91	2.16	0.79	0.48	0.44	0.80	1.15
	0.4	1.12	1.44	0.21	1.06	1.94	0.69	0.48	0.31	0.71	1.08
	0.6	1.00	1.14	0.13	1.05	1.71	0.84	0.38	0.73	0.94	1.10
	0.8	1.04	1.19	0.25	1.04	1.78	0.99	0.16	0.87	0.99	1.11

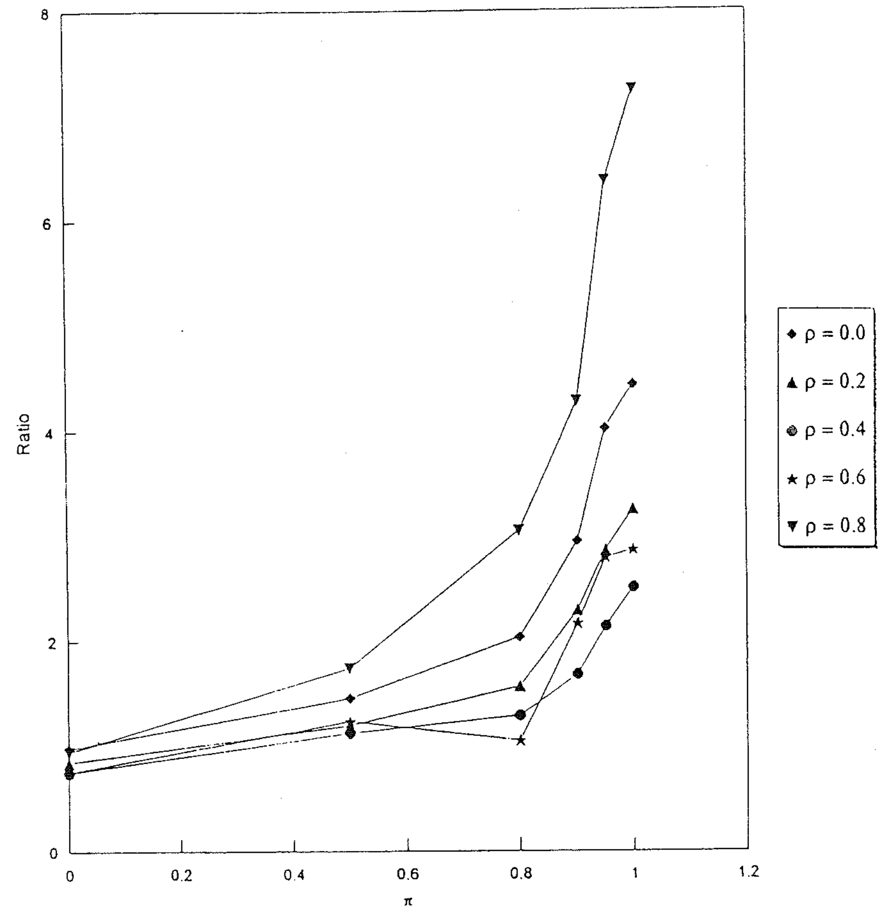

Fig. 1 Relative Efficiency of Heckman's Estimator
(SRMSE OF HECKMAN)/(SRMSE OF MLE)

Heckman's estimator performs quite well when π is small, and is even better than the MLE for $\pi = 0$. However, when π is large and close to 1, Heckman's estimator performs poorly.[1] If $\pi >$ 0.9, the standard deviations and quartile ranges of the estimates become several times as large as the MLE. The Figure 1 shows the ratios of the square root of mean squared errors (SRMSE) of the two-estimators. In terms of the mean squared error, Heckman's estimator performs reasonably well for the small values of π ; however, when π is large, it performs poorly for all values of ρ_0 and the mean squared error becomes several times as large as the MLE.

6. EMPIRICAL EXAMPLES

In this section, the empirical results of Heckman's estimator and the MLE are shown. Married-female wage and labor supply models, which are widely used in labor economics, are estimated using the 1993 survey data (Current Population Survey) of the U.S. Bureau of the Census.

The wage model consists of a participation equation and a wage equation, where the participation equation includes explanatory variables that are not included in the wage equation. As the result, the degree of muticollinearlity between $w_i'\tilde{\alpha}$ and x_i is not very high in this case.

The labor supply model consists of a participation equation and a labor supply equation. In this model, all explanatory variables of the participation equation are included in the lablor supply equation. This means that there is perfect multicollinearity between $w_i'\tilde{\alpha}$ and x_i.

6.1 DATA

The data set analyzed is chosen from the computer tape of the 1993 survey data of the U.S. Bureau of the Census. The sample of 1025 married civilian females belonging to primary families without missing information on the variables is selected from the original data set by random drawing. In the selected data sets, 626 females are working.

The definitions of the variables are given in Table 2. *WORK* is a dummy variable such that *WORK* = 1 if the female is working and 0 otherwise, *CHILD6* is the number of children under 6, *CHILD18* is the number of children under 18, *EDUCATION* and *AGE* are the educational years and the age of the female, *H_INCOME* is the income of the husband, *WAGE* is the wage rate, and *W_HOUR* is the working hours of the female. *WAGE* is observable if and only if *WORK* = 1 (i.e. *W_HOUR* > 0). The mean and the standard deviations of the variables are given in Table 3.

6.2 WAGE MODEL

The wage model consists of a participation equation and a wage equation. The participation equation is given by

$$y^* = \alpha_0 + \alpha_1 \cdot CHILD6 + \alpha_2 \cdot CHILD18 + \alpha_3 \cdot EDUCATION + \alpha_4 \cdot AGE$$

$$(6.1) \qquad + \alpha_5 \cdot AGE^2 + \alpha_6 \cdot H\_INCOME + u_1,$$

$$WORK = 1(y^* > 0).$$

When $WORK = 1$, the wage equation is observed and given by

$$(6.2) \qquad \ln(WAGE) = \beta_0 + \beta_1 \cdot EDUCATION + \beta_2 \cdot AGE + \beta_3 \cdot AGE^2 + u_2.$$

The results of the estimation by Heckman's estimator and the MLE is given by Table 4 and 5, where the standard errors are calculated by the Hessian matrix. (The standard errors are in parentheses.) Note that the package programs give the same estimates in this case.

Table 2 Definitions of the Variables

Variable	Definition
WORK	Dummy Variable; WORK = 1 if female was working in the previous year; 0 otherwise
CHILD6	Number of Children under 6
CHILD18	Number of Children under 18
EDUCATION	Educational Years (Years)
AGE	Age
H_INCOME	Husband's Income ($1000)
WAGE	Wage Rate of the Previous Year ($/Hour)
W_HOUR	Working Hours of the Previous Year (Hours)

Table 3 Mean and Standard Deviations of the Variables

Variable	Mean	S.D.
WORK	1: 626 observations;	0: 399 observations
CHILD6	0.296	0.603
CHILD18	0.882	1.133
EDUCATION	12.54	2.86
AGE	44.79	4.86
H_INCOME	22.20	23.48
WAGE (*WORK* = 1 only)	9.94	7.82
W_HOUR (*WORK* = 1 only)	1591.9	714.5

S.D.: Estimates of Standard Deviation

Table 4 Results of Estimation of the Participation Equation

Variable	Probit MLE	MLE of Selection Bias Models
Constant	-1.1282	-1.1608
	(0.5433)	(0.57375)
CHILD6	-0.3807	-0.3894
	(0.09335)	(0.09219)
CHILD18	-0.07763	-0.07349
	(0.05235)	(0.05195)
EDUCATION	0.08711	0.08988
	(0.01712)	(0.01727)
AGE	0.08729	0.08816
	(0.02501)	(0.02491)
AGE^2	-0.001508	-0.001521
	(0.0002706)	(0.0002697)
H_INCOME	-0.002213	-0.002893
	(0.002028)	(0.002062)

Table 5 Results of Estimation of the Wage Equation

Variable	Heckman	MLE of Selection Bias Model
Constant	-0.1717	-0.4582
	(0.2877)	(0.3665)
Education	0.07639	0.08406
	(0.01427)	(0.01598)
AGE	0.05886	0.06855
	(0.01223)	(0.01598)
AGE^2	-0.0006478	-0.0007935
	(0.0001459)	(0.0001947)
λ	-0.005929	
	(0.02659)	
σ_2		0.6430
		(0.00208)
ρ		0.224
		(0.148)

In this model, the differences of Heckman's estimator and the MLE are quite small, and both estimators give essentially the same results. Heckman's estimator may be good for this model. If we regress $w_i'\tilde{\alpha}$ on the explanatory variables of the wage equation (i.e. *AGE*, *EDUCATION*, and AGE^2), $R^2 = 0.7567$, which is not very high. Therefore, Heckman's estimator may escape any problem with multicollinearity suggested in this paper.

6.3 LABOR SUPPLY MODEL

The labor supply model consists of a participation equation and a labor supply equation. The participation equation is the same as (6.1). When $WORK = 1$, the labor supply of the female is observed. The labor supply equation is given by

$$W\_HOUR = \gamma_0 + \gamma_1 \cdot CHILD6 + \gamma_2 \cdot CHILD18 + \gamma_3 \cdot EDUCATION + \gamma_4 \cdot AGE + \gamma_5 \cdot AGE^2 + \gamma_6 \cdot H\_INCOME + \gamma_7 \cdot WAGE + u_3 . \quad (6.3)$$

The results of the estimation are given in Tables 6 and 7. Unlike the wage model, the procedures of the econometric package programs failed to find the correct values of the MLE due to multiple local maxima in this case. The package programs stop at $\rho = 0.17$ where the value of the log-likelihood is -5490.88; however, the correct global maximum is attained at $\rho = -0.80$ where the log-likelihood is -5489.26.

The differences of the two estimators are fairly large in the labor supply equation. For example, the signs of *CHILD6* and *EDUCATION* are opposite; *AGE* and AGE^2 are highly significant (t-values are 4.60 and -4.69) in Heckman's estimator but the t-values of the MLE are not very large (1.69 and -1.18) for the same variables. In this case, all of the explanatory variables of the participation equation are contained in the labor supply equation. Therefore, the problems of multicollinearity become serious and Heckman's estimator may not be a good method.

7. CONCLUSION

This paper analyzes Heckman's two-step estimator, which has been widely used in the estimation of sample selection bias models, by Monte Carlo technique and empirical examples. The results suggest that Heckman's estimator may not be a proper estimator when there is a high degree of multicollinearity between $w_i'\tilde{\alpha}$ and x_i. However, if x_i contains variables which are not included in w_i, and the degree of multicollinearity between $w_i'\tilde{\alpha}$ and x_i is not very high, the

Table 6 Results of Estimation of the Participation Equation

Variable	Probit MLE	MLE of Selection Bias Models
Constant	-1.1282	-1.1082
	(0.5433)	(0.5514)
CHILD6	-0.3807	-0.3349
	(0.09335)	(0.09114)
CHILD18	-0.07763	-0.05888
	(0.05235)	(0.05046)
EDUCATION	0.08711	0.08364
	(0.01712)	(0.01637)
AGE	0.08729	0.08240
	(0.02501)	(0.02422)
AGE^2	-0.001508	-0.001419
	(0.0002706)	(0.0002622)
H_INCOME	-0.002213	-0.001927
	(0.002028)	(0.002016)

Table 7 Results of Estimation of the Labor Supply Equation

Variable	Heckman	MLE of Selection Bias Model
Constant	148.08	1344.2
	(356.0)	(420.6)
CHILD6	-97.69	10.06
	(75.54)	(6.378)
CHILD18	-118.86	-87.21
	(29.56)	(34.42)
Education	6.820	-17.24
	(17.30)	(12.88)
AGE	79.06	32.90
	(17.20)	(19.47)
AGE^2	-0.9829	-0.2755
	(0.20970)	(0.2330)
H_INCOME	-1.867	-1.157
	(1.532)	(1.148)
WAGE	5.599	5.707
	(5.5269)	(3.839)
λ	11.77	
	(49.43)	
σ_2		801.0
		(376.0)
ρ		-0.801
		(0.0605)

performance of Heckman's estimator is good and it may be reasonable to use it in actual estimation. All of the empirical studies which use Heckman's method should be revised with these qualification in mind.

NOTE

1. When $\pi = 1$, the values of R^2 for an auxiliary regression of $\lambda(\tilde{\alpha}_0 + \tilde{\alpha}_1 w_i)$ on x_i are as follows. Maximum: 0.9980, 75%: 0.9934, 50%: 0.9909, 25%: 0.9877, Minimum: 0.9722, and Average: 0.9901.

ACKNOWLEDGMENTS

The authors wish to thank Michael McAleer, Janet Stotsky, the editor, and two anonymous referees for their helpful comments.

REFERENCES

Amemiya, T., 1985, *Advanced Econometrics,* Harvard University Press, Cambridge, MA.

Gronau, R., 1973, "The Effects of Children on the Housewife's Value of Time," *Journal of Political Economy,* 81, 168-199.

Heckman, J., 1974, "Shadow Prices, Market Wages and Labor Supply," *Econometrica,* 42, 679-694.

_____ , 1976, "The Common Structure of Statistical Models of Truncation, Sample Selection Bias and Limited Dependent Variables and a Simple Estimator for such Models," *Annals of Econometric and Social Measurement,* 5, 475-492.

________, 1979, "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161.

Maddala, G. S., 1983, *Limited-Dependent and Qualitative Variables in Econometrics,* Cambridge University Press.

Mroz, T. A., 1987, "The Sensitivity of Empirical Models with Sample-Selection Biases," *Econometrica,* 55, 765-799.

Nawata, K., 1993, "A Note on Estimation of Models with Sample-Selection Biases," *Economic Letters,* 42, 15-24.

_______, 1994, "Estimation of the Sample-Selection Biases Models by the Maximum Likelihood Estimator and Heckman's Two-Step Estimator," *Economic Letters,* 45,33-40.

Nelson, F. D., 1984, "Efficiency of the Two-Step Estimator for Models with Endogenous Sample Selection," *Journal of Econometrics,* 24, 181-196.

Olsen, R. J., 1980, "Approximating a Truncated Normal Regression with the Method of Moments," *Econometrica*, 48, 1099-1105.

__________, 1982, "Distributional Tests for Selectivity Bias and a More Robust Likelihood Estimator," *International Economic Review*, 23, 223-240.

Paarsch, H. J., 1984, "A Monte Carlo Comparison of Estimators for Censored Regression Models," *Journal of Econometrics*, 24, 197-213.

Wales, T. J., and A.D. Woodland, 1980, "Sample Selectivity and the Estimation of Labor Supply Functions," *International Economic Review*, 21, 437-468.

White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.