The Korean Economic Review, Vol. 4, No. 1

THE SIMPLIFIED ESTIMATION OF A NONLINEAR SIMULTANEOUS EQUATIONS MODEL WITH SELECTIVITY

HOE KYUNG LEE*

I. INTRODUCTION

Research by Amemiya (1974) and others (for instance, Heckman (1974), Nelson and Olson (1978)) provided a basis for the alternative methods to the estimation of censored linear simultaneous equations models. The methods suggested are two-step procedures in which the reduced forms are explicitly specified. It has been shown that the resulting estimators have desirable properties.

When the simultaneous equations censored-data model is nonlinear in endogenous variables, those methods are not applicable since the reduced form, in general, cannot be expressed in closed form due to the nonlinearity of the variables. Theoretically, maximum likelihood (ML) estimation can be applied to the structural model, but in practice it may cause serious computational difficulties since the likelihood function has multiple integrals which, in general, have to be calculated numerically. In this study an alternative procedure to ML is proposed which alleviates the computational burden.

The remainder of this paper is organized as follows. Section two discusses the problems arising from ML with a simplified two-equations nonlinear model, and formulates an alternative two-step procedure. Throughout the steps, the quasi-ML estimation is used. White (1982) and Domowitz and White (1982) showed that the quasi-ML estimators under suitable conditions provide the consistent estimates for the parameters which minimizes the Kullback-Leibler (1951) information. The Monte Carlo experiments are presented in section three. We have concentrated on the performance of the alternative procedure in doing the experiments because we know the large sample properties of the ML estimator and the experiments with ML take a substantial amount of computational cost. The results of the Monte Carlo study comparing different parameter values and different sets of exogenous variables in the approximation are discussed. Some concluding remarks are included in section four.

* Korea Institute of Technology

Ⅱ. THE MODEL and ESTIMATORS

The most general form of the censored nonlinear simultaneous equations system may be specified as follows along with the appropriate selection rules :

(1) $k_i(Y_t^*,\ X_t,\ r_i) = u_{it} \quad i=1, \cdots, G$ and $t=1, \cdots, T$

where Y_t is a vector of endogenous variables, X_t is a vector of exogenous variables, r_i is a vector of unknown parameters, and u_{it} is a scalar independent identically distributed random variable. Nonlinear simultaneous models without censoring or truncation have been studied extensively (for instance, Kelejian (1971), Goldfeld and Quandt (1968), Amemiya (1974, 1975, 1977)). In this section, we set up a simplified nonlinear censored model and examine what causes the diffculties in applying the maximum likelihood method and propose an alternative to ML.

The simplified version of (1) which will be considered here has the following features : (i) it has two endogenous variables y_1, y_2 ; and (ii) both endogenous variables are censored or truncated under the single selection rule ; and (iii) the model is nonlinear only in variables. Within the group of models specified by (i)-(iii) is (2) below.

$$
(2)\quad
\begin{aligned}
y_{1t}^* &= \alpha_1 y_{2t}^* + X_{1t}\beta_1 + u_{1t} \\
\log\ y_{2t}^* &= \alpha_2 y_{1t}^* + X_{2t}\beta_2 + u_{2t} \\
y_{1t} &= y_{1t}^* && \text{if } y_{1t}^* > 0 \\
&= 0 && \text{otherwise} \\
y_{2t} &= y_{2t}^* && \text{if } y_{1t}^* > 0 \\
&= \text{unobserved} && \text{otherwise}
\end{aligned}
$$

where X_{it} is a $(1\times K_i)$ vector of exogenous variables, β_i is a $(K_i\times 1)$ vector of parameters, and α_i is a scalar parameter. In (2) y_{1t}^*, y_{2t}^* are the unobservable latent variables and y_{1t}, y_{2t} are their observed counterparts respectively. We assume that the distribution of u_{1t} and u_{2t} is bivariate normal with mean zero and variance-covariance Ω. The immediate application of the model of this type would be a simultaneous system for the labor demand and supply relationships where y_{1t} is the hours of work and y_{2t} is the market wage rate. Then, the first equation of (2) can be regarded as the labor force participation equation and the second one can be the market wage determination equation.

We also assume that $\alpha_1\alpha_2 < 0$ which guarantees the unique solution for y_{1t}^*, y_{2t}^*. In other words, $\alpha_1\alpha_2 < 0$ is a sufficient condition for the one-to-one transformation between (u_{1t}, u_{2t}) and (y_{1t}^*, y_{2t}^*) so that the Jacobian of the transformation can be used in specifying the distribution function of (y_{1t}^*, y_{2t}^*). We will assume specifically $\alpha_1 > 0$ and $\alpha_2 < 0$ in the remainder of this paper. The restrictions on the parameters of the endogenous variables may be referred to the internal or logical consistency condition for the nonlinear censored model (for the linear case, see Schmidt (1981)). However, unlike the linear case it does not seem to have the general condition for the internal consistency due to the functional form which defines the nonlinearity.

It can be easily shown that the model (2) is identified by the theorems on identifying nonlinear models provided by Fisher (1966).

1. Maximum Likelihood Estimation

The likelihood function for (2) can be written as

$$(3)\quad L = \prod_{+} f_{v_{1t}v_{2t}}(y_{1t}, y_{2t}) \cdot \prod_{0} P(y^*_{1t} \leqq 0)$$

$$= \prod_{+} f_{u_{1t}u_{2t}}(y_{1t} - \alpha_1 y_{2t} - X_{1t}\beta_1, \log y_{2t} - \alpha_2 y_{1t} - X_{2t}\beta_2) \cdot |1/y_{2t} - \alpha_1\alpha_2|$$

$$\cdot \prod_{0} \int_{-\infty}^{0}\int_{0}^{\infty} f_{u_{1t}u_{2t}}(y_{1t} - \alpha_1 y_{2t} - X_{1t}\beta_1, \log y_{2t} - \alpha_2 y_{1t} - X_{2t}\beta_2)$$

$$\cdot |1/y_{2t} - \alpha_1\alpha_2|\, dy_{2t}\, dy_{1t}$$

Since the integration cannot be done analytically, it must be numerically at each function evaluation, and it is computationally burdensome. If there were additional censored variables y_{jt}, $j=3, 4, \cdots$, then these multiple integrals could cause serious computational diffculties. These problems lead us to consider other procedures which could alleviate the computational burden.

Because the distribution function for (y^*_{1t}, y^*_{2t}) involves the infinite intervals of integration, it should be transformed to a manageable distribution function in order to do the numerical integration. Because of the nonlinearity there does not seem to be a general way to find a manageable distribution function of (y^*_{1t}, y^*_{2t}) which can be handled relatively easily on the computer. However, for our specific model (2), it is possible to transform the distribution function into the one which has a single integral so that a proper method of numerical integration can be applied with much less error bounds.

2. An Alternative Procedure to the ML Estimator

A natural step to reduce the computational difficulties caused by the two multiple integrals in structural ML would be to use the concept of the two stage estimators of various kinds by thansforming the structural model to the reduced form equations. The obvious problem is then in general the reduced form cannot be expressed in closed form due to the nonlinearity in endogenous variables. A quasi-reduced form, however, may be defined by taking conditional expectations of functions of endogenous variables on the r. h. s. of each equation with respect to the exogenous variables. The conditioned variables, then, either be the observed y_{it} or be the unobserved latent variables y^*_{it}. It would be more appropriate to use y^*_{it} then y_{it} as conditioned variables, because we know X_t even when $y^*_{1t} \leqq 0$.

Let us take the conditional expectation of the functions of the latent variables. Since y^*_{1t} and y^*_{2t} appear on the r. h. s. of each equation we have

$$(4)\quad E(y^*_{1t}\mid X_t) = h_{1t}(X_t)$$
$$E(y^*_{2t}\mid X_t) = h_{2t}(X_t)$$

provided the expectations exist. Thus the quasi-reduced form may be written as

$$(5)\quad \begin{aligned} y^*_{1t} &= h_{1t}(X_t) + \varepsilon_{1t} \\ y^*_{2t} &= h_{2t}(X_t) + \varepsilon_{2t} \\ y_{1t} &= y^*_{1t} && \text{if } y^*_{1t} > 0 \\ &= 0 && \text{otherwise} \\ y_{2t} &= y^*_{2t} && \text{if } y^*_{1t} > 0 \\ &= \text{unobserved} && \text{otherwise} \end{aligned}$$

The equations $h_{it}(X_t)$ and the error structure of ε_{it} are unknown. However, we have the realized y_{it} for each y^*_{it} which induces us to use the ML technique to get a good predictor of y_{it} in the Kullback-Leibler information context.

Let us assume that $h_{it}(X_t)$ are approximated by $X_t R_i$ which are linear functions in quasi-reduced form parameters. Then

$$(6)\quad \begin{aligned} y^*_{1t} &= X_t R_1 + \nu_{1t} \\ y^*_{2t} &= X_t R_2 + \nu_{2t} \\ y_{1t} &= y^*_{1t} && \text{if } y^*_{1t} > 0 \\ &= 0 && \text{otherwise} \\ y_{2t} &= y^*_{2t} && \text{if } y^*_{1t} > 0 \\ &= \text{unobserved} && \text{otherwise} \end{aligned}$$

and the quasi-likelihood function is

$$(7)\quad L = \prod_{+} f_{\nu_{1t}\nu_{2t}}(y_{1t} - X_t R_1,\ y_{2t} - X_t R_2) \cdot \prod_{0} P(y^*_{1t} \leq 0)$$

The use of the quasi—ML estimator as a way to get a good predictor of y_{it} can be explained by the following theorem which is a modification of White (1982), Domowitz and White (1982)

Assumption 1. The random vectors $v_t = (v_{1t}, v_{2t})$ have the joint distribution function $G_t(v_t)$ on a measurable Euclidean space $\tilde{\Omega}$, and have measurable joint density function $g_t(v_t)$.

Assumption 2. Assumed joint distribution function $F_t(v_t, \theta)$ have the joint density function $f_t(v_t, \theta)$ which are continuous in θ for each v in $\tilde{\Omega}$ and measurable in v for each θ in Θ.

Let θ_T be the quasi-ML estimator which is the solution to

$$(8)\quad \text{Min } S_t(v, \theta) = 1/T \sum \log L$$

where

$$L = \prod_{+} \phi_t(y_{1t} - X_t R_1,\ y_{2t} - X_t R_2)$$
$$\cdot \prod_{0} \int_{-\infty}^{0}\int_{0}^{\infty} \phi_t(y_{1t} - X_t R_1,\ y_{2t} - X_t R_2)\, dy_{2t}\, dy_{1t}$$

The measurability of θ_T is proved by Jenrich (1969). Also let θ^* be the parameter vector which solves

(9) Min $E[1/T \Sigma I_t (g_t; f_t, \theta_T)]$

where $I_t (g_t; f_t, \theta_T)$ is the Kullback-Leibler information defined as

(10)
$$I_t(g_t; f_t, \theta) = E\left[\log \frac{g_t(v_t)}{f_t(v_t, \theta)}\right]$$

$$= \int \log \frac{g_t(v_t)}{f_t(v_t, \theta)} \cdot g_t(v_t)\, dv_t$$

Assumption 3. $E[\log g_t(v_t)]$ exists and $|\log f_t(v_t, \theta)| \leqq K(v_t)$ for all θ in Θ, where $K(v_t)$ is integrable with respent to the probability distribution function F_t. Then we have the following theorem

Theorem 2.1 Under the assumptions 1, 2 and 3, the quasi-ML estimator $\hat{\theta}_t$ converges stongly to θ^*.

Proof follows immediatly from Theorems (2) and (5) of Domowitz and White (1982) ; their theorem (2) ensures

$|S_T(v, \theta) - E[1/T \Sigma I_t (g_t; f_t, \theta)]| \cdots\cdots\cdots\cdots> 0$

uniformly for all θ in Θ and stong consistency of θ_T to θ^* is provided by their theorem (5)

If the random vectors v_t are i. i. d, then theorem (2) of White (1982) is a special case of theorem (3.1). For simplicity, we will adopt the i. i. d assumption of v_t, and use quasi-ML estimation for the model (2).

Let the quasi-ML estimator be $\hat{R}_i$, $i=1,2$. Then at the second step, for positive y_{it}.

(11)
$$\begin{aligned} y_{1t} &= a_1 y_{2t} + X_{1t}\beta_1 + u_{1t} \\ &= \alpha_1 (X_t \hat{R}_2 + \hat{v}_{2t}) + X_{1t}\beta_1 + u_{1t} \\ &= \alpha_1 X_t \hat{R}_2 + X_{1t}\beta_1 + \eta_{1t} \end{aligned}$$

where

$$\eta_{1t} = \alpha_1 \hat{v}_{2t} + u_{1t}.$$

Similarly, for y_{2t},

(12)
$$\begin{aligned} y_{2t} &= \alpha_2 y_{1t} + X_{2t}\beta_2 + u_{2t} \\ &= \alpha_2 X_t e\, \hat{R}_1 + X_{2t}\beta_2 + \eta_{2t} \end{aligned}$$

where

$$\eta_{2t} = \alpha_1 \hat{v}_{2t} + u_{2t}.$$

Following the same line of reasoning in the first step, let the quasi-ML estimators $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_1$, $\hat{\beta}_2$ be the solutions of the quasi-likelihood function

$$L = \prod_{+} f_{\eta_1\eta_2}(y_{1t} - \alpha_1 X_t \hat{R}_2 - X_{1t}\beta_1, \log y_{2t} - \alpha_2 X_t \hat{R}_1 - X_{2t}\beta_2)$$

$$\cdot \prod_{0} P(y_{1t}^* \leq 0).$$

Then $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_1$, and $\hat{\beta}_2$ converge strongly to α_1^*, α_2^*, β_1^*, and β_2^* respectively which minimizes the Kullback-Leibler information of the suitable form. It should be noted, however, that $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_1$, and $\hat{\beta}_2$ will not necessarily be consistent estimates of the structural parameters. Because of the approximations at both steps, it is very hard or may not be possible to get analytic asymptotic properties of the alternative procedure. It may, then, be worth investigating the quality of the procedure by the Monte Carlo experiments.

Ⅲ. MONTE CARLO EXPERIMENTS

1. The Model and Estimators

In this section, the simplified nonlinear censored simultaneous equations model (13) is investigated.

$$(13)\quad \begin{aligned} y_{1t}^* &= \alpha_1 y_{2t}^* + \beta_{11} + \beta_{12} x_{2t} + u_{1t} \\ \log y_{2t}^* &= \alpha_2 y_{1t}^* + \beta_{22} x_{2t} + u_{2t} \\ y_{1t} &= y_{1t}^* && \text{if } y_{1t}^* > 0 \\ &= 0 && \text{otherwise} \\ y_{2t} &= y_{2t}^* && \text{if } y_{1t}^* > 0 \\ &= \text{unobserved} && \text{otherwise} \end{aligned}$$

where x_{1t}, x_{2t} are scalar exogenous variables, (u_{1t}, u_{2t}) is bivariate normal with mean zero and variance-covariance $\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$. For the purpose of reducing the computer time, the constant term of the second equation is excluded and the covariance σ_{12} between u_{1t} and u_{2t} is assumed to be zero.

Two estimators are examined in this study ; the maximum likelihood estimator applied to the stuctural model and the alternative procedure. We have concentrated, however, on the behavior of the alternative procedure, and the ML technique was tried only to get the standard errors in the ML estimates. The main reason not to focus on ML estimation for the structural model is that the asymptotic properties of ML are already known and the computational cost of ML was too expensive to do several sampling experiments.

2. Design of Experiments and Generation of Data

The sampling experiments reported in this study were performed in order to explore the effect of three main characteristics on the performance of the alternative procedure. They are the degree of censoring, the variation in the disturbances which corresponds to the R^2, and the distributional assumption of the data generating mechanism. The degree of censoring plays an important role in examining the quality of estimators in the tobit-type models. Throughout the study the constant term β_{11} was chosen to be the parameter determining the desired degree of censoring. We take two different degrees of censoring, fifty and twenty percent for the main twelve experimets, and additional five different ratios of censoring especially for the normal distribution with zero mean and unit variance. To examine the effect for the variation in the disturbances, two types of variance of u_1 and u_2 are employed which correspond to the R^2 of about 5 and 3 respectively in the regression equations before censoring. For the normal distriution, R^2 of 1 was also tried. These R^2's were picked since the real life censored data usually yield low R^2's.

It is often criticized that in the tobit model ML estimators under the normality assumption do not yield consistent estimates in general when the true distribution is not normal. (See Arabmazar and Schmidt(1982)). The proposed alternative procedure suffers the same problem, and it may be interesting to see the sensitivity of this estimator under the different distributional assumptions. In this study, normal, uniform, and Laplace distributions with the same mean and variance are used to see the robustness of the alternative estimator.

x_{it}'s were drawn from the unit normal distribution, and u_{it}'s were selected from the normal, uniform, and Laplace distributions. For all three distributions means were set to zero and the variances to 1 and 1.96 which correspond to the R^2's of .5 and .3 before censoring respectively. In deciding the true parameter values, we have tried to retain about equal variation of each r. h. s. variables. For given x_{it} and u_{it}, y^*_{it} can now be obtained, and in our model it must be done numerically since there is no closed reduced form. A group of uncensored y^*_{1t} and y^*_{2t} can then be generated by successive drawings on u_{1t}, u_{2t} with x_{it} fixed. If y^*_{1t} which is obtained from the first equation of (13) after inserting calculated y^*_{2t} into that equation is positive, both y^*_{1t} and y^*_{2t} are recorded as y_{1t} and y_{1t} respectively. If y^*_{1t} turn out to be nonpositive, zeros are assigned to y_{1t} and "unobserved" to y_{2t}

Because the alternative estimator need be consistent in general and it dose not have the analytic asymptotic distribution, we chose relatively large sample size of 1200 to examine the qualiy of the estimator empirically in a large sample context. The number of repetitions for each experiment was chosen to be two hundred and fifty. This number seems to be enough to trace out the empirical distributions for the alternative procedure which does not have desirable asymptotic properties.

For ML estimation, the probability that y^*_{1t} is nonpositive must be calculated, and in many cases it must be done numerically. As noted in section two, direct application of a numerical method to the multiple integrals which have infinite intervals may well be prohibitively costly. It could also produce inaccurate results due to the cumulated errors so that the numerical quality of estimators may be unreliable. However, for the specific model (13), it is possible to transform the double integration problem into single integration which is numerically tractable and also provides more accurate results. The cost of ML from this transformed likelihood function turned out to be almost thirty times more expensive than that of the proposed alternative procedure.

3. Results of the Experiments

The results from the twelve main experiments are summarized in Table 1-4. The last three columns of Tables 1-4 give the mean values of the estimated parameters from two hundred and fifty repetitions under the different distributional assumptions. Throughout the tables, the standard errors from the empirical distributions are reported in parentheses below the mean values of the estimates. Table 1 and 2 show that under fifty percent censoring the alternative procedure yields biased estimates, as expected, for all parameters except β_{12}. These biases, however, shrink quickly when the degree of censoring decreases to twenty percent as shown in Tables 3-4.

In examining the effect of the degree of censoring, several more experiments were done with the unit normal distribution and the results are presented in Tables 5. These tables show that as the degree of censoring declines the asymptotic biases approach to zero uniformly for all estimates except β_{12} : β_{12} shows stable estimated results throughout the experiments. Table 5 also shows that the standard errors of α_1 tend to increase as the degree of censoring decreases below thirty to forty percent. α_1 is the parameter of the r. h. s. endogenous variables y^*_{2i} of the first equation in (13), and this phenomenon may result from the decreasing variation in y_2 because the straight line of the first equation shifts to the right and cuts the relatively flat portion of the curve corresponding to the second equation as the degree of censoring declines. In other words, after some point of the degree of censoring, the increasing trend of the standard errors of α_1 due to the smaller variation in y_2 overpowers the decreasing trend resulting from taking more information due to lower censoring.

The alternative procedure appears to be robust to a variety of distributional assumptions of the true error structure. There is no indication that the alternative procedure produces better estimates under normality than any other two distributions. With fifty percent censoring, the estimates except β_{12} are rather improved under the Laplace distribution, and with twenty percent even the uniform distribution yields better results for some parameters than the other two. It is not surprising that the alternative procedure shows robustness to the three assumed distributions since the underlying true distribution for quasi ML estimation at both steps were unknown and may not belong to any of the assumed distributions.

It is also observed from Table 6 that at fifty percent censoring the asymptotic biases of the estimates become greater as the variation in the disturbances increases, that is, as the R^2 decreases. This has been expected since the explanatory power of the r. h. s. variables diminishes or more variables are excluded from the r. h. s. as the variance of the disturbances increases. However, Tables 1-4 indicate that the differences in estimates arising from different variations decrease when the degree of censoring becomes twenty percent.

The quality of approximations of the unknown functions at the first step was also investigated by employing linear, quadratic, and cubic functions separately as approximating functions. The results are reported in Table 7 and they show that the estimates are not uniformly improved as higher order terms of x_{it} are used in the approximation. It indicates that the better approximation of the quasi reduced form equations of the nonlinear model investigated in our study may be found with some other functions than the second or third order polynomial approximation.

Ⅳ. CONCLUSION

We have investigated the nonlinear censored simultaneous equations model which is directly related to the labor supply model. We have shown how the application of maximum likelihood estimation to the structural model causes difficulties in practice. The main reason may be summarized as (i) the computational cost, (ii) the difficulty of locating the optimum, which may result from the cumulated errors in calculating the probability, $P(y_{1t}^{*}\leqq 0)$, numerically, and (iii) the difficulty of specifying the manageable distribution function for calculating $P(y_{1t}^{*}\leqq 0)$ in the likelihood function. The proposed alternative procedure employs quasi-ML estimation at both first and second steps : reduced form parameters are estimated after specifying the quasi-reduced form at the first step, and the fitted values of the functions of r. h. s. endogenous variables are substituted into the structural equations to estimate the structural parameters at the second step. Because obtaining the analytic sampling distribution of the alternative estimator appears to be intractable due to the difficulty arising from the successive approximations of the unknown functions at both steps, Monte Carlo experiments were done to examine the performance of the alternative procedure. It has been shown that the asymptotic biases of the estimates from the alternative procedure approach zero as the degree of censoring decreases. Monte Carlo study also shows that the alternative procedure may serve at least obtaining the starting points of ML estimation, if available, of the structural model. Since the alternative estimator was explored for the specific model directly related to the labor supply equations, a further research would be the development of a more general treatment of the estimation problem in the general nonlinear censored simultaneous equations models.

REFERENCES

[1] AMEMIYA, T (1974), "Multivariate Regression and Simultaneous Equation Models When the Dependent Varibles are Truncated Normal," *Econometrica* 42, 999-1012.

[2] _____ (1975), "The Nonlinear Limited-Information Maximum-Likelihood Estimator and the Modified Nonlinear Two-Stage Least-Squares Estimator," *Journal of Econometrics* 3, 375-386.

[3] _____ (1977), "The Maximum Likelihood and the Nonlinear Three-stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," *Econometrica* 45, 955-968

[4] ARABMAZAR, A. and SCHMIDT, P. (1982), "An Investigation of the Robustness of the Tobit Estimator to Non-Normality," *Econometrica* 50, 1055-1063.

[5] DOMOWITZ, I. and WHITE H. (1982), "Misspecified Models with Dependent Observations," *Journal of Econometrics* 20, 35-58.

[6] FISHER, F. M. (1976), *The Identification Problem in Econometrics*, N. Y : Robert E. Krieger Publishing Company.

[7] GOLDFELD, S. M. and QUANDT, R. E. (1968), "Nonlinear Simultaneous Equations : Estimation and Prediction", *International Economic Review* 9, 113-136.

[8] HECKMAN, J. (1974), "Shadow Prices, Market Wages, and Labor Supply," *Econometrica* 42, 679-694.

[9] JENNRICH, R. I. (1969), "Asymptotic Properties of Non-Linear Least Squares Estimators," *The Annals of Mathematical Statistics* 40, 633-643.

[10] KELEJIAN, H. H. (1971) "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in Endogenous Variables," *Jounal of the American Statistical Assoctation* 66, 373-374.

[11] KULLBACK, S. and LEIBLER, R. A. (1951) "On Information and Sufficiency," *The Annals of Mathematical Statistics* 22, 79-86.

[12] NELSON F. and OLSON L. (1978), "Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables, " *International Economic Review* 19, 695-709.

[13] SCHMIDT, P. (1981). "Constraints on the Parameters in Simultaneous Tobit and Probit Models," in C. F. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, Mass ; M. I. T. Press.

[14] WHITE, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50, 1-25.

[Table 1] Average Values of Estimates
Mean 0, Variance 1
Fifty Percent Censoring

	Truth	Normal	Uniform	Laplace
α_1	1	0.880	0.852	0.930
		(0.069)	(0.064)	(0.081)
β_{11}	-1.5	-1.293	-1.238	-1.391
		(0.136)	(0.128)	(0.157)
β_{12}	-1	-0.987	-0.976	-1.017
		(0.065)	(0.059)	(0.079)
α_2	-1	-1.129	-1.157	-1.096
		(0.085)	(0.085)	(0.090)
β_{22}	1	1.132	1.150	1.111
		(0.059)	(0.058)	(0.061)

[Table 2] Average Values of Estimates
Mean 0, Variance 1.96, Fifty Percent Censoring

	Truth	Normal $\beta_{11}=-1.7$	Uniform $\beta_{11}=-1.8$	Laplace $\beta_{11}=-1.6$
α_1	1	0.835	0.785	0.915
		(0.085)	(0.077)	(0.102)
β_{11}		-1.377	-1.357	-1.464
		(0.179)	(0.170)	(0.206)
β_{12}	-1	-0.986	-0.968	-1.037
		(0.081)	(0.073)	(0.101)
α_2	-1	-1.200	-1.278	-1.125
		(0.125)	(0.136)	(0.123)
β_{22}	1	1.202	1.240	1.163
		(0.088)	(0.091)	(0.087)

[Table 3] Average Values of Estimates
Mean 0, Variance 1, Twenty Percent Censoring

	Truth	Normal	Uniform	Laplace
α_1	1	0.957	0.947	0.971
		(0.082)	(0.081)	(0.082)
β_{11}	.1	0.137	0.136	0.138
		(0.074)	(0.076)	(0.069)
β_{12}	-1	-0.992	-0.981	-1.006
		(0.043)	(0.043)	(0.047)
α_2	-1	-0.998	-0.992	-1.007
		(0.027)	(0.028)	(0.027)
β_{22}	1	1.030	1.028	1.035
		(0.036)	(0.036)	(0.035)

[Table 4] Average Values of Estimates
Mean 0, Variance 1.96, Twenty Percent Censoring

	Truth	Normal	Uniform	Laplace
		$\beta_{11}=.3$	$\beta_{11}=.4$	$\beta_{11}=.2$
α_1	1	0.947	0.955	0.959
		(0.116)	(0.126)	(0.113)
β_{11}		0.341	0.400	0.265
		(0.106)	(0.118)	(0.098)
β_{12}	-1	-0.990	-0.981	-1.009
		(0.059)	(0.062)	(0.064)
α_2	-1	-0.982	-0.963	-1.003
		(0.033)	(0.033)	(0.035)
β_{22}	1	1.033	1.019	1.047
		(0.049)	(0.048)	(0.049)

[Table 5] Changes in the Degree of Censoring (70-10%)
N (0, 1)

	Truth	70% $\beta_{11}=-2.8$	60% $\beta_{11}=-2.1$	50% $\beta_{11}=-1.5$	40% $\beta_{11}=-1.0$	30% $\beta_{11}=-.4$	20% $\beta_{11}=.1$	10% $\beta_{11}=.9$
α_1	1	0.859	0.861	0.880	0.901	0.931	0.957	0.986
		(0.087)	(0.070)	(0.069)	(0.069)	(0.073)	(0.082)	(0.106)
β_{11}		-2.387	-1.785	-1.293	-0.863	-0.328	0.137	0.910
		(0.280)	(0.177)	(0.136)	(0.105)	(0.082)	(0.074)	(0.065)
β_{12}	-1	-0.978	-0.978	-0.987	-0.990	-0.992	-0.992	-0.992
		(0.096)	(0.076)	(0.065)	(0.053)	(0.047)	(0.043)	(0.041)
α_2	-1	-1.101	-1.170	-1.129	-1.066	-1.015	-0.998	-0.993
		(0.130)	(0.119)	(0.085)	(0.058)	(0.039)	(0.027)	(0.018)
β_{22}	1	1.085	1.140	1.132	1.098	1.056	1.030	1.008
		(0.070)	(0.069)	(0.059)	(0.050)	(0.041)	(0.036)	(0.033)

[Table 6] Changes in R^2 Fifty Percent Censoring

	Truth	$R^2=.5$ N (0, 1) $\beta_{11}=-1.5$	$R^2=.3$ N (0, 1.96) $\beta_{11}=-1.7$	$R^2=.1$ N (0, 9) $\beta_{11}=-2.7$
α_1	1	0.880	0.835	0.762
		(0.069)	(0.085)	(0.166)
β_{11}		-1.293	-1.377	-1.890
		(0.136)	(0.179)	(0.473)
β_{12}	-1	-0.987	-0.986	-0.990
		(0.065)	(0.081)	(0.162)
α_2	-1	-1.129	-1.200	-1.359
		(0.085)	(0.125)	(0.311)
β_{22}	1	1.132	1.202	1.363
		(0.059)	(0.088)	(0.222)

[Table 7] Higher Order Terms in X_{it}
N (0, 1)
Fifty Percent Censoring

	Truth	Linear	Quadratic	Cubic
α_1	1	0. 880	0. 865	0. 838
		(0. 069)	(0. 071)	(0. 070)
β_{11}	-1. 5	-1. 293	-1. 280	-1. 244
		(0. 136)	(0. 141)	(0. 137)
β_{12}	-1	-0. 987	-0. 992	-0. 987
		(0. 065)	(0. 070)	(0. 069)
α_2	-1	-1. 129	-1. 125	-1. 115
		(0. 085)	(0. 083)	(0. 081)
β_{22}	1	1. 132	1. 133	1. 128
		(0. 059)	(0. 063)	(0. 061)