IDENTIFICATION AND ESTIMATION IN BINARY CHOICE MODELS WITH LIMITED DEPENDENT VARIABLES

by

Lung-Fei Lee*

Discussion Paper No. 77-85, June 1977

---

*This paper is based partly on my Ph.D. Thesis submitted to the University of Rochester. I would like to express my gratitude to G. S. Maddala for his supervision and encouragements. Thanks also to S. Rosen and C. A. Sims for many valuable discussions. I am indebted to Mr. Robert B. Litterman for correcting my English. Any errors remaining are solely my own.

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

IDENTIFICATION AND ESTIMATION IN BINARY CHOICE MODELS WITH LIMITED DEPENDENT VARIABLES

By

Lung-Fei Lee

1. Introduction

In handling the statistical problems of choices among finite discrete outcomes, many models have been suggested and studied. Among these, the most well known are the linear probability model, Probit analysis [6] and Logit analysis [3]. In these models, probability functions which assign probabilities to various discrete outcomes are estimated. The probabilities vary over subjects under study. They depend on the attributes of outcomes and the characteristics of subjects. These methods have long histories and are used extensively in the literature of bioassay. In economics, they became popular in recent years. In an excellent paper, McFadden [24] gives rigorous theoretical interpretations based on consumer behavioral analysis for those techniques. McFadden summarized his findings and empirical analysis in a recent book on urban travel demand with Domencich [4]. Most of these models, however, do not involve simultaneous structures. In certain cases, decisions are based on the possible outcomes under alternative choices and observed outcomes are final outcomes of the decision process. So decisions and outcomes are interrelated. While there are many possible ways to formulate these relationships our main concern is in normal probability models.

In this paper, we will specify probit models with continuous endogeneous variables or limited dependent variables. While some of those specifications are not new (see Maddala and Nelson [22], Westin [28]), our concerns are in the identification and estimation in those models which have not been explored.

The paper is organized as follows. In Section 2, a binary choice model with limited dependent variables is discussed. We discuss the identification problems involved in this model and suggest two stage estimation methods to get consistent estimates. With the consistent estimates available, simpler maximum likelihood procedures are then developed. In Section 3, we prove the consistency of a two stage probit estimator. In Section 4, we extend the binary choice model to cases with multivariate limited dependent variables and switching simultaneous equations. In Section 5, we point out some special cases in our models and some empirical applications.

2. Binary Choice Model with Limited Dependent Variables.

In many binary choices, possible outcomes will influence decisions which are realized as the choices are made. So those outcomes are not exogeneous but are endogeneously determined. While there are several possible specifications (see Westin [28], Maddala and Nelson [22]) available, we are interested in the following model.

$$Y_{1t} = X_{1t}\beta_1 + \varepsilon_{1t} \quad \text{iff} \quad Z_t\gamma + Y_{1t}\zeta_1 + Y_{2t}\zeta_2 \geq v_t$$

$$Y_{2t} = X_{2t}\beta_2 + \varepsilon_{2t} \quad \text{iff} \quad Z_t\gamma + Y_{1t}\zeta_1 + Y_{2t}\zeta_2 < v_t \ .$$

In this model, the error terms are serially independent, normally distributed with zero mean and covariance matrix Σ ,

$$\Sigma = \text{cov} \begin{vmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ v_t \end{vmatrix} = \begin{vmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1v} \\ \sigma_{21} & \sigma_2^2 & \sigma_{2v} \\ \sigma_{1v} & \sigma_{2v} & \sigma_v^2 \end{vmatrix}$$

Also we assume that the binary outcome can be observed, i.e. sample separation is available. But the endogeneous variables Y_t will be observed only if the choice is made.[1]

Since the endogeneous variables Y_{1t} and Y_{2t} are involved in the decision process and are outcomes of the final choice, observed values of Y_{1t} and Y_{2t} are limited dependent. The limited dependent notion is similar to Tobin [27]. Given the exogeneous variables X_t , the population distributions of Y_{it} $(i = 1, 2)$ are normally distributed, but the observable distribution is truncated.

Since the endogeneous variables are included in the decision process, the system as a whole is a simultaneous equation system. It is different from the usual simultaneous equation system in econometrics, however, as one of the dependent variable is dichotomous. As in usual simultaneous equation model, we can expect the model will not be estimable without more restrictions on parameters. In our model we will show that there are more serious identification problems.

The above model can be written in the switching model as follows,

$$Y_{1t} = X_{1t}\beta_1 + \varepsilon_{1t} \quad \text{iff} \quad Z_t\frac{\gamma}{\sigma^*} + X_{1t}\frac{\zeta_1\beta_1}{\sigma^*} + X_{2t}\frac{\zeta_2\beta_2}{\sigma^*} \geq \varepsilon_{ot}$$

$$Y_{2t} = X_{2t}\beta_2 + \varepsilon_{2t} \quad \text{iff} \quad Z_t\frac{\gamma}{\sigma^*} + X_{1t}\frac{\zeta_1\beta_1}{\sigma^*} + X_{2t}\frac{\zeta_2\beta_2}{\sigma^*} < \varepsilon_{ot}$$

where $\varepsilon_{ot} = \frac{\gamma}{\sigma^*}(v_t - \zeta_1\varepsilon_{1t} - \zeta_2\varepsilon_{2t})$ and $\sigma^{*2} = E(v_t - \zeta_1\varepsilon_{1t} - \zeta_2\varepsilon_{2t})^2$.

This switching model can be regarded as reduced form of the original system. Since sample separation is available, there is no difficulty in identifying the parameters of the reduced form and hence the parameters β_1 and β_2.[2] The identification problem will occur only in the structural decision function and the parameters of disturbances in the system. A more detailed study of this switching model is reported in Lee [15] and Lee and Trost [18].

The decision function in reduced form is a probit model

$$I_t^* = Z_t \frac{\gamma}{\sigma^*} + X_{1t} \frac{\zeta_1 \beta_1}{\sigma^*} + X_{2t} \frac{\zeta_2 \beta_2}{\sigma^*} - \varepsilon_{ot} \quad .$$

Coefficients γ, ζ_1 and ζ_2 can be identified only up to a positive proportion. As the reduced form suggests, what we try to identify is the parameters $\frac{\gamma}{\sigma^*}$, $\frac{\zeta_1}{\sigma^*}$ $\frac{\zeta_2}{\sigma^*}$ instead of γ, ζ_1, ζ_2. However, even these parameters can not easily be identified without further restrictions. To simplify the expressions, let us consider zero order type restrictions only. For general linear restrictions, similar analysis can be applied. The parameter $\frac{\gamma}{\sigma^*}$ $\frac{\zeta_1}{\sigma^*}$ and $\frac{\zeta_2}{\sigma^*}$ cannot be identified if the vector of exogeneous variables Z_t contains all the exogeneous variables in X_{1t} and X_{2t}. To achieve identification, some exogeneous variables in Z_t have to be excluded in the decision function.

To simplify the notation, let us denote

$$X_{1t} = [Z_t, W_t] \quad \text{and} \quad X_{2t} = [Z_t, W_t] \; .$$

Thus the equations Y_{1t} and Y_{2t} can be written as

$$Y_{1t} = Z_t \beta_{10} + W_t \beta_{11} + \varepsilon_{1t}$$

$$Y_{2t} = Z_t \beta_{20} + W_t \beta_{21} + \varepsilon_{2t}$$

where $\beta_1 = (\beta_{10}, \beta_{11})$ and $\beta_2 = (\beta_{20}, \beta_{21})$. The vector W_t is a vector of exogeneous variables whose elements are included in either Y_{1t} or Y_{2t} but not in the decision function. Combining common terms in the decision function, we have

$$I_t^* = Z_t\left(\frac{\gamma}{\sigma^*} + \frac{\zeta_1}{\sigma^*}\beta_{10} + \frac{\zeta_2}{\sigma^*}\beta_{20}\right) + W_t\left(\frac{\zeta_1}{\sigma^*}\beta_{11} + \frac{\zeta_2}{\sigma^*}\beta_{21}\right) - \varepsilon_{ot}$$

Denote $C_1 = \frac{\gamma}{\sigma^*} + \frac{\zeta_1}{\sigma^*}\beta_{10} + \frac{\zeta_2}{\sigma^*}\beta_{20}$ and $C_2 = \frac{\zeta_1}{\sigma^*}\beta_{11} + \frac{\zeta_2}{\sigma^*}\beta_{21}$.

From this probit model, C_1 and C_2 are identifiable and since β_1 and β_2 are also identifiable, we can investigate identification of the parameters $\frac{\gamma}{\sigma^*}$, $\frac{\zeta_1}{\sigma^*}$ and $\frac{\zeta_2}{\sigma^*}$ under these equations;

$$\frac{\gamma}{\sigma^*} = [C_1, \beta_{10}, \beta_{20}] \begin{vmatrix} 1 \\ \frac{-\zeta_1}{\sigma^*} \\ \frac{-\zeta_2}{\sigma^*} \end{vmatrix}$$

and

$$[\beta_{11}, \beta_{21}] \begin{vmatrix} \frac{\zeta_1}{\sigma^*} \\ \frac{\zeta_2}{\sigma^*} \end{vmatrix} = C_2$$

From these relations, the parameters ζ_1/σ^* , ζ_2/σ^* , γ/σ^* are identifiable if and only if $[\beta_{11}, \beta_{21}]$ has full column rank, i.e., rank equals 2. A necessary condition is that at least two exogeneous variables which appear in the X_{1t} or X_{2t} are excluded from Z_t . These conditions are thus similar to the rank condition and order condition for usual simultaneous equations models.

Finally it remains to consider identification of parameters in the residuals. From the reduced form the parameters $\sigma_1^2 = \mathrm{var}(\varepsilon_{1t})$, $\sigma_2^2 = \mathrm{var}(\varepsilon_{2t})$, $\sigma_{1\varepsilon_o} = \mathrm{cov}(\varepsilon_{1t}\, \varepsilon_{ot})$ and $\sigma_{2\varepsilon_o} = \mathrm{cov}(\varepsilon_{2t}\varepsilon_{ot})$ are identifiable. However, as pointed out in Lee and Trost [18],

σ_{12} will not be identifiable from the reduced form. As all the parameters in Σ are involved in the expressions for σ_1^2, σ_2^2, $\sigma_{1\varepsilon_o}$ and $\sigma_{2\varepsilon_o}$, identification of these parameters can be investigated.

The explicit relations between these parameters are as follows,

$$\sigma_{1\varepsilon_o} = E(\frac{1}{\sigma^*}(\varepsilon_{1t}v_t - \zeta_1\varepsilon_{1t}^{\ 2} - \zeta_2\varepsilon_{1t}\varepsilon_{2t}))$$

$$= \frac{\sigma_{1v}}{\sigma^*} - (\frac{\zeta_1}{\sigma^*})\sigma_I^2 - (\frac{\zeta_2}{\sigma^*})\sigma_{12}$$

$$\sigma_{2\varepsilon_o} = E(\frac{1}{\sigma^*}(\varepsilon_{2t}v_t - \zeta_1\varepsilon_{1t}\varepsilon_{2t} - \zeta_2\varepsilon_{2t}^{\ 2}))$$

$$= \frac{\sigma_{2v}}{\sigma^*} - (\frac{\zeta_1}{\sigma^*})\sigma_{12} - (\frac{\zeta_2}{\sigma^*})\sigma_2^2$$

and $1 = \operatorname{var}(\varepsilon_o) = E(\frac{1}{\sigma^*}(v_t - \zeta_1\varepsilon_{1t} - \zeta_2\varepsilon_{2t}))^2$

$$= (\frac{\sigma_v}{\sigma^*})^2 + (\frac{\zeta_1}{\sigma^*})^2\sigma_1^2 + (\frac{\zeta_2}{\sigma^*})^2\sigma_2^2 - 2(\frac{\zeta_1}{\sigma^*})(\frac{\sigma_{1v}}{\sigma^*})$$

$$- 2(\frac{\zeta_2}{\sigma^*})(\frac{\sigma_{2v}}{\sigma^*}) + 2(\frac{\zeta_1}{\sigma^*})(\frac{\zeta_2}{\sigma^*})\sigma_{12}$$

From these three equations, we can not identify four unknown parameters σ_{12}, σ_{1v}/σ^*, σ_{2v}/σ^*, σ_v/σ^*.

Under additional assumptions that $\sigma_{12} = 0$, the parameters $\frac{\sigma_{1v}}{\sigma^*}$, $\frac{\sigma_{2v}}{\sigma^*}$, and $\frac{\sigma_v}{\sigma^*}$ will be identifiable. When $\sigma_{12} = 0$, we have

$$\frac{\sigma_{1v}}{\sigma^*} = \sigma_{1\varepsilon_0} + (\frac{\zeta_1}{\sigma^*})\sigma_1^2$$

$$\frac{\sigma_{2v}}{\sigma^*} = \sigma_{2\varepsilon_0} + (\frac{\zeta_2}{\sigma^*})\sigma_2^2$$

and $(\frac{\sigma_v}{\sigma^*})^2 = 1 - (\frac{\zeta_1}{\sigma^*})^2 \sigma_1^2 - (\frac{\zeta_2}{\sigma^*})^2 \sigma_2^2 + 2(\frac{\zeta_1}{\sigma^*})\sigma_{1\varepsilon_o} + 2(\frac{\zeta_2}{\sigma^*})\sigma_{2\varepsilon_o}$.

Under the alternative assumption that v_t is independent of ε_{1t} and ε_{2t}, parameters σ_{12} and σ_v/σ^* will be identifiable whenever $\zeta_1 \neq 0$ or $\zeta_2 \neq 0$. If $\zeta_1 = 0$ and $\zeta_2 = 0$, the system will be exactly the reduced form and σ_{12} will not be identifiable.

Let us now consider the estimation of this model. If Y_{1t} and Y_{2t} are always observable, we have in fact a recursive system involving continuous and dichotomous endogeneous variables.[3] If ε_{1t}, ε_{2t} and v_t are mutually independent, we have a full recursive system and the estimation procedure will be straightforward. Coefficients of β can be estimated by ordinary least squares and γ, ζ can be estimated by the probit maximum likelihood method. This limited information maximum likelihood procedure is also a full information maximum likelihood procedure. However when Y_{1t} and Y_{2t} are limited dependent, disturbances ε_{1t}, ε_{2t} are truncated and straightforward least squares applied to observed subsamples will not give consistent estimates. The inconsistency is similar to Tobin's model [27]. An alternative procedure which will give consistent and asymptotically efficient estimates is the maximum likelihood method. However, as our model is highly non-linear, maximum likelihood methods which depend on numerical iterative procedures will not be easily accomplished without good initial estimates. As shown by Amemiya [1] in Tobin's model, if we can start with consistent estimates, the maximum likelihood procedure will be greatly simplified; each Newton-Raphson iteration (or Modified Newton-Raphson [2]) will give consistent and asymptotically efficient estimates.

As suggested in Lee [15], we can estimate the parameters in the switching regression model by simple two stage methods.[4] The two stage methods utilize modified least squares in the first stage and probit maximum likelihood in the second stage. More specifically, denote $\psi_t = Z_t C_1 + W_t C_2$.

The reduced form becomes

$$Y_{1t} = X_{1t}\beta_1 + \varepsilon_{1t} \quad \text{iff} \quad \psi_t \geq \varepsilon_{ot}$$

$$Y_{2t} = X_{2t}\beta_2 + \varepsilon_{2t} \quad \text{iff} \quad \psi_t < \varepsilon_{ot}$$

The underlying conditions define a probit model as

$$I_t = 1 \quad \text{iff} \quad \psi_t \geq \varepsilon_{ot}$$

$$I_t = 0 \quad \text{iff} \quad \psi_t < \varepsilon_{ot}$$

Hence the parameters C_1 and C_2 can be estimated consistently by the probit analysis. To estimate the β's, we notice that

$$E(\varepsilon_{1t}|I_t = 1) = -\sigma_{1\varepsilon_o} \frac{f(\psi_t)}{F(\psi_t)}$$

$$E(\varepsilon_{2t}|I_t = 0) = \sigma_{2\varepsilon_o} f(\psi_t)/(1-F(\psi_t))$$ where f and F are standard normal density and distribution functions respectively. Hence

$$E(Y_{1t}|I_t = 1) = X_{1t}\beta_1 - \sigma_{1\varepsilon_o} f(\psi_t)/F(\psi_t)$$

$$E(Y_{2t}|I_t = 0) = X_{2t}\beta_2 + \sigma_{2\varepsilon_o} f(\psi_t)/(1-F(\psi_t)).$$

which can be rewritten as,

$$Y_{1t} = X_{1t}\beta_1 - \sigma_{1\varepsilon_o} f(\psi_t)/F(\psi_t) + \eta_{1t}$$

$$Y_{2t} = X_{2t}\beta_2 + \sigma_{2\varepsilon_o} f(\psi_t)/(1-F(\psi_t)) + \eta_{2t}$$

where

$$E(\eta_{1t}|I_t = 1) = 0 \quad \text{and} \quad E(\eta_{2t}|I_t = 0) = 0$$

In the second stage, the probit estimates of C_1 and C_2 are used to get an estimate $\hat{\psi}_t$ of ψ_t. With subsamples corresponding to $I_t = 1$, ordinary least squares estimate β from the equation

$$Y_{1t} = X_{1t}\beta_1 - \sigma_{1\varepsilon_o} f(\hat{\psi}_t)/F(\hat{\psi}_t) + \eta_{1t}$$

It has been shown in Lee and Trost [18] that these estimates are consistent under general conditions. Similarly we can consistently estimate β_2. If some coefficients are equal a priori in the two equations Y_{1t} and Y_{2t}, it is also possible to incorporate them in the two stage procedure. The two equations can be combined with the D method proposed by Goldfeld and Quandt [7] into a single equation. Two stage method is then applied to this combined equation. A more detail discussion on this method can be found in Lee [15].

With estimates $\hat{C}_1$ and $\hat{C}_2$ from the probit analysis and two stage estimates $\hat{\beta}_1$ and $\hat{\beta}_2$, consistent estimates of ζ_1/σ^*, ζ_2/σ^* and γ/σ^* can be derived via the equations:

$$\hat{C}_1 = \frac{\gamma}{\sigma*} + \frac{\zeta_1}{\sigma*}\hat{\beta}_{10} + \frac{\zeta_2}{\sigma*}\hat{\beta}_{20}$$

and $$\hat{C}_2 = \frac{\zeta_1}{\sigma*}\hat{\beta}_{11} + \frac{\zeta_2}{\sigma*}\hat{\beta}_{21} .$$

If the model is exactly identified, we will have a unique solution and the estimation procedure corresponds to the usual indirect least square procedure. However this is not the case if it is over-identified. To overcome the ambiguity, a two stage procedure is available. We can modify the decision function as

$$I* = Z_t(\frac{\gamma}{\sigma*}) + (X_{1t}\hat{\beta}_1)\ (\frac{\zeta_1}{\sigma*}) + (X_{2t}\hat{\beta}_2)\ (\frac{\zeta_2}{\sigma*}) - \tilde{\varepsilon}_o$$

where $\tilde{\varepsilon}_o$ is a resultant disturbance which is asymptotically standard normal. The probit maximum likelihood procedure is then applied to estimate $\gamma/\sigma*$, $\zeta_1/\sigma*$ and $\zeta_2/\sigma*$.

To give it a name, we will call it a <u>two stage probit analysis</u>. Under general conditions, this two stage probit estimates are consistent. The proof is presented in the next section. Estimates of parameters in Σ can be derived from reduced form parameters σ_1^2 , σ_2^2 , $\sigma_{1\varepsilon_o}$ and $\sigma_{2\varepsilon_o}$ if additional restrictions are available. Otherwise, they can not be identified and hence are not estimable. To estimate σ_1^2 , σ_2^2 , $\sigma_{1\varepsilon_o}$ and $\sigma_{2\varepsilon_o}$, the estimated residuals can be used.

As shown in Lee and Trost [18] or Johnson and Katz [14],

$$E(\varepsilon_{1t}^2 | I_t = 1) = \sigma_1^2 - \sigma_{1\varepsilon_o}^2 \psi_t \frac{f(\psi_t)}{F(\psi_t)}$$

and

$$E(\varepsilon_{2t}^2 | I_t = 0) = \sigma_2^2 + \sigma_{2\varepsilon_o}^2 \psi_t f(\psi_t)/(1-F(\psi_t))$$

which gives

$$\varepsilon_{1t}^2 = \sigma_1^2 - \sigma_{1\varepsilon_o}^2 \psi_t {}^{f(\psi_t)}/_{F(\psi_t)} + \xi_{1t}$$

$$\varepsilon_{2t}^2 = \sigma_2^2 + \sigma_{2\varepsilon_o}^2 \psi_t {}^{f(\psi_t)}/_{(1-F(\psi_t))} + \xi_{2t}$$

To estimate σ_1^2 and $\sigma_{1\varepsilon_o}^2$, ordinary least squares can be applied to

$$\hat{\varepsilon}_{1t}^2 = \sigma_1^2 - \sigma_{1\varepsilon_o}^2 \hat{\psi}_t f(\hat{\psi}_t)/F(\hat{\psi}_t) + \tilde{\xi}_{1t}$$

where $\hat{\varepsilon}_{1t} = Y_t - \hat{\beta}_1' X_{1t}$ are the estimated residuals. Similarly we can estimate σ_2^2 and $\sigma_{2\varepsilon_o}^2$. All these estimates are consistent as shown in Lee and Trost [18]. As we recall we also have consistent estimates on $\sigma_{1\varepsilon_o}$ and $\sigma_{2\varepsilon_o}$ from the two stage procedure, we may simplify the procedure to estimate σ_1^2 and σ_2^2 only. Those estimates from the above procedures can be shown to be asymptotically normally distributed. However the asymptotic variance matrices are quite complicated.

With all the parameters consistently estimated, asymptotically efficient estimates can be derived with two step maximum likelihood procedures and their asymptotic variances can easily be derived. The likelihood function for this model is

$$L = \prod_{t-1}^{T} \left(\int_{-\infty}^{\psi_t} f_1(Y_{1t} - Z_t\beta_{10} - W_t\beta_{11}, \varepsilon_{ot})d\varepsilon_{ot}\right)^{I_t}$$

$$\left(\int_{\psi_t}^{\infty} f_2(Y_{2t} - Z_t\beta_{20} - W_t\beta_{21}, \varepsilon_{ot})d\varepsilon_{ot}\right)^{1-I_t}$$

where f_1 and f_2 are the jointly normal density functions for ε_{1t}, ε_{ot} and ε_{2t}, ε_{ot} respectively. The logarithmic likelihood function is

$$\ell nL = \sum_{t=1}^{T} \{I_t[\ell n \frac{1}{\sigma_1} \phi\left(\frac{Y_{1t}-Z_t\beta_{10}-W_t\beta_{11}}{\sigma_1}\right) + \ell n\ \Phi\ (\eta_{1t})]$$

$$+ (1-I_t)\ [\ell n \frac{1}{\sigma_2} \phi\left(\frac{Y_{2t}-Z_t\beta_{20}-W_t\beta_{21}}{\sigma_2}\right) + \ell n\ \Phi\ (\eta_{2t})]\}$$

where $\eta_{1t} = [\psi_t - \frac{\rho_1}{\sigma_1}(Y_{1t} - Z_t\beta_{10} - W_t\beta_{11})] / \sqrt{1-\rho_1^2}$

$$\eta_{2t} = [\psi_t - \frac{\rho_2}{\sigma_2}(Y_{2t} - Z_t\beta_{20} - W_t\beta_{21})] / \sqrt{1-\rho_2^2}$$

with ρ_1, ρ_2 the correlation coefficients of ε_{1t}, ε_{ot} and ε_{2t}, ε_{ot} respectively and ϕ and Φ are standard normal density and distribution functions.

The two step maximum likelihood estimates (2SML) Θ^* with consistent estimators $\tilde{\Theta}$ are defined as

$$\Theta^* = \tilde{\Theta} - \left[\frac{\partial^2 \ln L(\tilde{\Theta})}{\partial\Theta\partial\Theta'}\right]^{-1} \frac{\partial \ln L(\tilde{\Theta})}{\partial\Theta}$$

The square roots of the diagonal elements in

$$-\left[\frac{\partial^2 \ln L(\tilde{\Theta})}{\partial\Theta\partial\Theta'}\right]^{-1}$$

will consistently estimate the asymptotic standard errors of these estimates.

The analytical first and second derivatives of the logarithmic likelihood function can be derived in our model. However, the expressions for the second derivatives are complicated. Instead of using second derivatives of the logarithmic likelihood function, we can use the covariance matrix of the gradient

$$\left[\sum_{t=1}^{T} \frac{\partial \ln L_t(\tilde{\Theta})}{\partial\Theta} \cdot \frac{\partial \ln L_t(\tilde{\Theta})}{\partial\Theta'}\right]$$

where

$$L_t = I_t\left[\ln \frac{1}{\sigma_1}\phi\left(\frac{Y_{1t}-Z_t\beta_{10}-W_t\beta_{11}}{\sigma_1}\right) + \ln\Phi(\eta_{1t})\right]$$

$$+ (1-I_t)\left[\ln \frac{1}{\sigma_2}\phi\left(\frac{Y_{2t}-Z_t\beta_{20}-W_t\beta_{21}}{\sigma_2}\right) + \ln\Phi(\eta_{2t})\right]$$

The 2SML is then defined as

$$\Theta^* = \tilde{\Theta} + \left[\sum_{t=1}^{T} \frac{\partial \ln L_t(\tilde{\Theta})}{\partial\Theta} \frac{\partial \ln L_t(\tilde{\Theta})}{\partial\Theta'}\right]^{-1} \frac{\partial \ln L(\tilde{\Theta})}{\partial\Theta}$$

and the asymptotic covariance matrix of the 2SML can be consistently estimated by

$$[\Sigma_{t=1}^{T} \frac{\partial \ell n L_t(\tilde{\Theta})}{\partial \Theta} \frac{\partial \ell n L_t(\tilde{\Theta})}{\partial \Theta'}]^{-1} \quad .$$

The first derivatives in our model are as follows. To simplify notations, subscripts t will be dropped out and we denote $\gamma^* = \gamma/\sigma^*$, $\zeta_1^* = \zeta_1/\sigma^*$, $\zeta_2^* = \zeta_2/\sigma^*$.

$$\frac{\partial \ell n L}{\partial \beta_{10}} = \Sigma_{t=1}^{T} \{I[\frac{Y_1 - Z\beta_{10} - W\beta_{11}}{\sigma_1^2} + \frac{\phi(\eta_1)}{\Phi(\eta_1)} \cdot (\frac{\rho_1}{\sigma_1} + \zeta_1{}^*)/ \sqrt{1-\rho_1^2}$$

$$+ (1-I) [- \frac{\zeta_1^*}{\sqrt{1-\rho_2^2}} \cdot \frac{\phi(\eta_2)}{1-\Phi(\eta_2)}]\} Z'$$

$$\frac{\partial \ell n L}{\partial \beta_{11}} = \Sigma_{t=1}^{T} \{I [\frac{Y_1 - Z\beta_{10} - W\beta_{11}}{\sigma_1^2} + \frac{\phi(\eta_1)}{\Phi(\eta_1)} \cdot (\frac{\zeta_1}{\sigma_1} + \zeta_1^*)/ \sqrt{1-\rho_1^2}]$$

$$+ (1-I) [- \frac{\zeta_1^*}{\sqrt{1-\rho_2^2}} \cdot \frac{\phi(\eta_2)}{1-\Phi(\eta_2)}]\} W'$$

$$\frac{\partial \ell n L}{\partial \beta_{20}} = \Sigma_{t=1}^{T} \{I(\frac{\zeta_2{}^*}{\sqrt{1-\rho_1^2}} \cdot \frac{\phi(\eta_1)}{\Phi(\eta_1)}) + (1-I)[\frac{Y_2 - Z\beta_{20} - W\cdot\beta_{21}}{\sigma_2^2} -$$

$$\frac{\phi(\eta_2)}{1-\Phi(\eta_2)} \cdot (\frac{\rho_2}{\sigma_2} + \zeta_2^*) /\sqrt{1-\rho_2^2}]\} Z'$$

$$\frac{\partial \ell nL}{\partial \beta_{21}} = \Sigma_{t=1}^{T} \{I \left(\frac{\zeta_2^*}{\sqrt{1-\rho_1^2}} \cdot \frac{\phi(\eta_1)}{\Phi(\eta_1)}\right) + (1-I)\left[\frac{Y_2 - Z\beta_{20} - W\beta_{21}}{\sigma_2^2} - \right.$$

$$\left.\frac{\phi(\eta_2)}{1-\Phi(\eta_2)} \cdot \left(\frac{\rho_2}{\sigma_2} + \zeta^*\right) / \sqrt{1-\rho_2^2}\right] \} W'$$

$$\frac{\partial \ell nL}{\partial \gamma^*} = \Sigma_{t=1}^{T} \{I \frac{\phi(\eta_1)}{\Phi(\eta_1)} \cdot \frac{1}{\sqrt{1-\rho_1^2}} - (1-I) \frac{\phi(\eta_2)}{1-\Phi(\eta_2)} \cdot \frac{1}{\sqrt{1-\rho_2^2}} \} Z'$$

$$\frac{\partial \ell nL}{\partial \zeta_1^*} = \Sigma_{t=1}^{T} \{I \frac{\phi(\eta_1)}{\Phi(\eta_1)} \cdot \frac{1}{\sqrt{1-\rho_1^2}} - (1-I) \frac{\phi(\eta_2)}{1-\Phi(\eta_2)} \cdot \frac{1}{\sqrt{1-\rho_2^2}} \}(Z\beta_{10} + W\beta_{11})$$

$$\frac{\partial \ell nL}{\partial \zeta^*_2} = \Sigma_{t=1}^{T} \{I \frac{\phi(\eta_1)}{\Phi(\eta_1)} \cdot \frac{1}{\sqrt{1-\rho_1^2}} - (1-I) \frac{\phi(\eta_2)}{1-\Phi(\eta_2)} \cdot \frac{1}{\sqrt{1-\rho_2^2}} \}(Z\beta_{20} + W\beta_{21})$$

$$\frac{\partial \ell nL}{\partial \sigma_1} = \Sigma_{t=1}^{T} I\left[-\frac{1}{\sigma_1} + \frac{(Y_1 - Z\beta_{10} - W\beta_{11})^2}{\sigma_1^3} + \frac{\phi(\eta_1)}{\Phi(\eta_1)} \cdot \frac{\rho_1}{\sqrt{1-\rho_1^2}} \cdot \frac{(Y_1 - Z\beta_{10} - W\beta_{11})}{\sigma_1^2}\right]$$

$$\frac{\partial \ell nL}{\partial \sigma_2} = \Sigma_{t=1}^{T} (1-I)\left[-\frac{1}{\sigma_2} + \frac{(Y_2 - Z\beta_{20} - W\beta_{21})^2}{\sigma_2^3} - \frac{\phi(\eta_2)}{1-\Phi(\eta_2)} \cdot \frac{\rho_2}{\sqrt{1-\rho_2^2}} \cdot \frac{(Y_2 - Z\beta_{20} - W\beta_{21})}{\sigma_2^2}\right]$$

$$\frac{\partial \ell nL}{\partial \rho_1} = \Sigma_{t=1}^{T} I \frac{\phi(\eta_1)}{\Phi(\eta_1)} \cdot \left(-\frac{Y_1 - Z\beta_{10} - W\beta_{11}}{\sigma_1 (1-\rho_1^2)^{3/2}}\right)$$

$$\frac{\partial \ell nL}{\partial \rho_2} = \Sigma_{t=1}^{T}(1-I)\frac{\phi(\eta_2)}{1-\Phi(\eta_2)} \cdot \frac{(Y_2-Z\beta_{20}-W\beta_{21})}{\sigma_2(1-\rho_2^2)^{3/2}}$$

3. Consistency of Two Stage Probit Estimator

In this section, we would like to show that the two stage probit estimator is strongly consistent. To prove this property, we need some lemmas.

Lemma 1: Let $f_m(\omega,\phi)$, $m=1,\dots,\infty$ be a sequence of measurable functions on a measurable space Ω and for each $\omega\varepsilon\Omega$, a continuous function for $\phi\varepsilon\Phi$, Φ being compact. Then there exists a sequence of measurable functions $\hat{\phi}_m(\omega)$, $m=1,\dots,\infty$ such that $f_m(\omega,\hat{\phi}_m(\omega)) = \sup_{\phi\varepsilon\Phi} f_m(\omega,\phi)$ for all $\omega\varepsilon\Omega$ and $m=1,\dots,\infty$. Furthermore, if for almost every $\omega\varepsilon\Omega$, $f_m(\omega,\phi)$ converges to $f(\phi)$ uniformly for all $\phi\varepsilon\Phi$ and if $f(\phi)$ has a unique global maximum at $\phi^*\varepsilon\Phi$, then $\hat{\phi}_m$ converges to ϕ^* for almost every $\omega\varepsilon\Omega$.

Lemma 2: Let μ be a probability measure over a Euclidean space S, let Φ be a compact subset of a Euclidean space and let $g(s,\phi)$ be a continuous function of ϕ for each $s\varepsilon S$ and a measurable function of s for each $\phi\varepsilon\Phi$. Assume also that $|g(s,\phi)|\leq\alpha$ for all s and ϕ and some finite α. For any sequence $\omega = s_1, s_2,\dots,$ let $f_M(\omega,\phi) = \Sigma_{m=1}^{M} g(s_m,\phi)/M$ and let Ω be the set of all sequences ω. If sequence ω are drawn as random samples from S, then for almost every realized such sequence, as $M\to\infty$

$$f_M(\omega,\phi) \longrightarrow E(g(s,\phi))$$

uniformly for all $\phi\varepsilon\Phi$.

Lemma 1 is in Amemiya [1]. Lemma 2 is a law of large numbers in Jennrich [13]. These lemmas have been used in many other studies in the literature, see e.g., Manski and Lermann [23].

Lemma 3: Let $f_T(\Theta_1,\Theta_2)$ be a sequence of continuous function on a compact set $H_1 \times H_2$. Suppose $\hat{\Theta}_{2T}$ is a strongly consistent estimator of Θ_2° which is an interior point in H_2. f_T converges to f uniformly on $H_1 \times H_2$. Then $f_T(\Theta_1, \hat{\Theta}_{2T})$ converges to $f(\Theta_1,\Theta_2^\circ)$ a.e. uniformly on H_1.

Proof: $\forall\ \varepsilon > 0$, there exists $T_o > 0$ such that

$$|f_T(\Theta_1,\Theta_2) - f(\Theta_1,\Theta_2)| \leq \varepsilon/2, \quad \text{for all } T \geq T_o \text{ and}$$

$$\text{for all } (\Theta_1, \Theta_2) \varepsilon\, H_1 \times H_2 .$$

As a uniform limit function of a sequence of functions, f is continuous. Since $H_1 \times H_2$ is compact, f is also uniformly continuous. Thus there exists $T_1 > 0$ such that

$$|f(\Theta_1,\hat{\Theta}_{2T}) - f(\Theta_1,\Theta_2^\circ)| \leq \varepsilon/2\ , \quad \text{for all } T \geq T_1$$

$$\text{and } \Theta_1\ \varepsilon\ H_1\ , \text{ a.e.}$$

It follows then $\forall\ T \geq T_o\ ,\ T^* \geq T_1$,

$$|f_T(\Theta_1,\hat{\Theta}_{2T^*}) - f(\Theta_1,\Theta_2^\circ)| \leq |\ f_T(\Theta_1,\hat{\Theta}_{2T^*}) - f(\Theta_1,\hat{\Theta}_{2T^*})| + |f(\Theta_1,\hat{\Theta}_{2T^*})$$
$$- f(\Theta_1,\Theta_2^\circ)| \leq \varepsilon$$

Therefore $\forall\ T \geq T_2 = \max\ (T_o, T_1)$,

$$|f_T(\Theta_1,\hat{\Theta}_{2T}) - f(\Theta_1,\Theta_2^\circ)| \leq \varepsilon, \quad \text{for all } \Theta_1\ \varepsilon\ H_1, \text{ a.e.}$$

Q.E.D.

With these lemmas, we can prove our theorem. To simplify the notations, let $X_t = (Z_t, X_{1t}, X_{2t})$, $S_t = (I_t, X_t)$, $\omega = \{S_t\}$ $\zeta = (\gamma, \zeta_1, \zeta_2)$ and $\beta = (\beta_1, \beta_2)$.

Theorem: Assume that the following conditions are satisfied:

1. $\{\varepsilon_{ot}\}$ are independently identically distributed standard normal variables.
2. X_t is a random sample drawn from a compact measurable space X with bounded density function $g(X_t)$ and X_t is independent with ε_{ot} for all t .
3. The parameter space $H_1 \times H_2$ of (ζ, β) are compact and the true parameter $(\zeta°, \beta°)$ is an interior point in $H_1 \times H_2$.
4. The rank condition for equation I* is satisfied.
5. $\hat{\beta}_T$ is a strongly consistent estimator of β

Then the two stage probit estimator $\hat{\zeta}_T$ is strongly consistent.

Proof: Let us denote

$$P(I_t, X_t, \zeta, \beta) = F(Z_t\gamma + (X_{1t}\beta_1)\zeta_1 + (X_{2t}\beta_2)\zeta_2)^{I_t}$$

$$(1 - F(Z_t\gamma + (X_{1t}\beta_1)\zeta_1 + (X_2 + \beta_2)\zeta_2)^{1-I_t} \ .$$

The log likelihood function divided by sample size is

$$L_T^*(\omega, \zeta, \beta) = \frac{1}{T} \Sigma_{t=1}^{T} (\log P(I_t, X_t, \zeta, \beta) + \log g(X_t)) \ .$$

Let $\phi(\zeta, \beta) = E(\log P(I_t, X_t, \zeta, \beta) + \log g(X_t))$. By lemma 2, $L_T^*(\omega, \zeta, \beta)$ converges to $\phi(\zeta, \beta)$ uniformly on $H_1 \times H_2$.

By lemma 3, it follows $L_T^*(\omega,\zeta,\hat{\beta}_T)$ converges to $\phi(\zeta,\beta^\circ)$ uniformly on H_1. This theorem will follow from lemma 1 if $\phi(\zeta,\beta^\circ)$ has a unique maximum at $\zeta^\circ \epsilon H_1$. So it remains to prove that $\phi(\zeta,\beta^\circ)$ has a unique maximum at ζ°. It is known that a probit likelihood function is concave on the parameter space (see Haberman [10]). Hence, $L_T^*(\omega,\zeta,\hat{\beta}_T)$ is concave on H_1. As a limit function of $L_T^*(\omega,\zeta,\hat{\beta}_T)$, $\phi(\zeta,\beta^\circ)$ must be concave. Since ζ° is an interior point in H_1, $\phi(\zeta,\beta^\circ)$ must be strictly concave if $\phi(\zeta,\beta^\circ)$ has a local strict It follows $\phi(\zeta,\beta^\circ)$ has a unique maximum at ζ° if $\phi(\zeta,\beta^\circ)$ has a local strict maximum at ζ°.

Consider

$$\frac{\partial^2}{\partial\zeta\partial\zeta'}\phi(\zeta,\beta^\circ) .$$

Denote $\Psi_t = Z_t\gamma + (X_{1t}\beta_1^\circ)\zeta_1 + (X_{2t}\beta_2^\circ)\zeta_2$. It can easily be shown that

$$\frac{\partial}{\partial\zeta}\phi(\zeta,\beta^\circ) = E\left(\frac{f(\Psi_t)}{F(\Psi_t)(1-F(\Psi_t))}\begin{bmatrix} Z_t' \\ X_{1t}\beta_1^\circ \\ X_{2t}\beta_2^\circ \end{bmatrix}(I_t - F(\psi_t))\right)$$

where f and F are the standard normal density and distribution function respectively. It follows that $\frac{\partial}{\partial\zeta}\phi(\zeta^\circ,\beta^\circ) = 0$.

$$\frac{\partial^2}{\partial\zeta\partial\zeta'}\phi(\zeta^\circ,\beta^\circ) = -\int \frac{f^2(\Psi_t^\circ)}{F(\Psi_t^\circ)(1-F(\Psi_t^\circ))}\begin{vmatrix} Z_t' \\ X_{1t}\beta_1^\circ \\ X_{2t}\beta_2^\circ \end{vmatrix}[Z_t \;\; X_{1t}\beta_1^\circ \;\; X_{2t}\beta_2^\circ]g(X_t)dX_t$$

where $\Psi_t^\circ = Z_t\gamma^\circ + (X_{1t}\beta_1^\circ)\zeta_1^\circ + (X_{2t}\beta_2^\circ)\zeta_2^\circ$.

Since H_1 and X are compact, there exists a constant $k > 0$ such that

$$\frac{f^2(\Psi_t)}{F(\Psi_t)(1-F(\Psi_t))} \geqq k \quad \text{on } H_1 \times X$$

Hence $$\frac{\partial}{\partial\zeta\partial\zeta'}\phi(\zeta,\beta^\circ) \leqq -k \int_X \begin{vmatrix} Z_t' \\ X_{1t}\beta_1^\circ \\ X_{2t}\beta_2^\circ \end{vmatrix} [Z_t \; X_{1t}\beta_1^\circ \; X_{2t}\beta_2^\circ] g(X_t)dX_t \; .$$

As $X_{1t}\beta_1^\circ = Z_t\beta_{10}^\circ + W_t\beta_{11}^\circ$; $X_{2t}\beta_2^\circ = Z_t\beta_{20}^\circ + W_t\beta_{21}^\circ$,

$$[Z_t \; X_{1t}\beta_1^\circ \; X_{2t}\beta_2^\circ] = [Z_t \; W_t]\begin{bmatrix} I & \beta_{10}^\circ & \beta_{20}^\circ \\ 0 & \beta_{11}^\circ & \beta_{21}^\circ \end{bmatrix}$$

It implies $\dfrac{\partial^2}{\partial\zeta\partial\zeta'}\phi(\zeta^\circ,\beta^\circ) \leqq A$.

where $A = -k \begin{bmatrix} I & \beta_{10}^\circ & \beta_{20}^\circ \\ 0 & \beta_{11}^\circ & \beta_{21}^\circ \end{bmatrix}' \Sigma_{ZW} \begin{bmatrix} I & \beta_{10}^\circ & \beta_{20}^\circ \\ 0 & \beta_{11}^\circ & \beta_{21}^\circ \end{bmatrix}$;

Σ_{ZW} is the variance-covariance of (Z_t, W_t) .

From the rank condition, $\begin{bmatrix} I & \beta_{10}^\circ & \beta_{20}^\circ \\ 0 & \beta_{11}^\circ & \beta_{21}^\circ \end{bmatrix}$ has full column rank.

A is therefore a negative definite matrix and so is $\dfrac{\partial^2}{\partial\zeta\partial\zeta'}\phi(\zeta^\circ,\beta^\circ)$.

The theorem follows as $\phi(\zeta^\circ,\beta^\circ)$ has a local strict maximum at ζ° .

Q.E.D.

4. Binary Choice Models with Multivariate Limited Dependent Variables

The binary choice models are ready to be extended to the multivariate and switching simultaneous equation systems. In those models, many limited dependent variables will be involved in the decision process. First, let us consider the multivariate case.

$$Y_{1t} = Z_t\beta_{10} + W_t\beta_{11} + \varepsilon_{1t}$$

$$Y_{2t} = Z_t\beta_{20} + W_t\beta_{21} + \varepsilon_{2t}$$

$$\vdots$$

$$Y_{kt} = Z_t\beta_{ko} + W_t\beta_{k1} + \varepsilon_{kt}$$

$$I_t^* = Z_t\gamma + Y_{1t}\zeta_1 + Y_{2t}\zeta_2 + \dots + Y_{k_o t}\zeta_{k_o} - \varepsilon_t$$

where Z_t and W_t are vectors of exogeneous variables. In this model, the disturbances are assumed to be jointly normal with zero mean; independently and identically distributed for each observation. The endogeneous variables Y_t's are limited dependent. The k equations of limited dependent variables are divided into two regimes and the sample separation is assumed to be available. In the decision function I_t^*, the $k_o(\leqq k)$ limited dependent variables involved come from either one of regimes.

In this system, the β's are always identifiable. It remains to investigate identification conditions for the coefficients in the decision function. The reduced form of the decision function is

$$I_t^* = Z_t\left(\frac{\gamma}{\sigma^*} + \frac{\zeta_1}{\sigma^*}\beta_{10} + \cdots + \frac{\zeta_{k_o}}{\sigma^*}\beta_{k_o0}\right) + W_t\left(\frac{\zeta_1}{\sigma^*}\beta_{11} + \frac{\zeta_1}{\sigma^*}\beta_{21} + \cdots + \frac{\zeta_{k_o}}{\sigma^*}\beta_{k_o1}\right) - \varepsilon_{ot}$$

where $\operatorname{var}(\varepsilon_{ot}) = 1$. Denote

$$C_1 = \frac{\gamma}{\sigma^*} + \frac{\zeta_1}{\sigma^*}\beta_{10} + \cdots + \frac{\zeta_{k_o}}{\sigma^*}\beta_{k_o0}$$

$$C_2 = \frac{\zeta_1}{\sigma^*}\beta_{11} + \frac{\zeta_2}{\sigma^*}\beta_{21} + \cdots + \frac{\zeta_{k_o}}{\sigma^*}\beta_{k_o1}$$

Obviously C_1 and C_2 are identifiable. Thus the identification condition for the coefficients in the decision function is that $[\beta_{11},\ldots\ldots,\beta_{k_o1}]$ has full column rank, i.e., rank equals k_o. A necessary condition is that the number of excluded exogeneous variables in the decision function is at least k_o which is the number of limited dependent variables involved in the decision function.

Now let us consider a switching simultaneous equation model. In this model, there are two subsystems of usual simultaneous equations and a decision function.

$$B_1Y_{1t} + \Gamma_{10}Z_t + \Gamma_{11}W_t = \varepsilon_{1t}$$

$$B_2Y_{2t} + \Gamma_{20}Z_t + \Gamma_{21}W_t = \varepsilon_{2t}$$

$$I_t^* = \gamma Z_t + \zeta_1 Y_{1t} + \zeta_2 Y_{2t} - \varepsilon_t$$

where Y_{1t}, Y_{2t} are G_1 and G_2 vectors of endogeneous variables. In this system, sample separation is assumed to be available. Y_{1t} and Y_{2t} are limited dependent and they are observable only when the relevant choice is made. As in the previous models, disturbances are jointly normal and are independent for different observations.

It is easy to see that the simultaneous equations in each regime will be identifiable under usual rank conditions for each regime. To identify the decision function, we can proceed from the reduced form. The reduced form for the whole system is

$$Y_{1t} = \Pi_{10}Z_t + \Pi_{11}W_t + v_{1t}$$

$$Y_{2t} = \Pi_{20}Z_t + \Pi_{21}W_t + v_{2t}$$

$$I^* = \left(\frac{\gamma}{\sigma^*} + \frac{\zeta_1}{\sigma^*}\Pi_{10} + \frac{\zeta_2}{\sigma^*}\Pi_{20}\right) Z_t + \left(\frac{\zeta_1}{\sigma^*}\Pi_{11} + \frac{\zeta_2}{\sigma^*}\Pi_{21}\right)W_t - \varepsilon_{ot}$$

where $E(\varepsilon_{ot}) = 0$. Denote

$$C_1 = \frac{\gamma}{\sigma^*} + \frac{\zeta_1}{\sigma^*}\Pi_{10} + \frac{\zeta_2}{\sigma^*}\Pi_{20}$$

$$C_2 = \frac{\zeta_1}{\sigma^*}\Pi_{11} + \frac{\zeta_2}{\sigma^*}\Pi_{21}$$

Thus

$$\frac{\gamma}{\sigma^*} = C_1 - \frac{\zeta_1}{\sigma^*}\Pi_{10} - \frac{\zeta_2}{\sigma^*}\Pi_{20}$$

$$\left[\frac{\zeta_1}{\sigma^*}, \frac{\zeta_2}{\sigma^*}\right] \begin{bmatrix} \Pi_{11} \\ \Pi_{21} \end{bmatrix} = C_2 \qquad (**)$$

So all the parameters Π and C_1 , C_2 are identifiable, parameters $\frac{\gamma}{\sigma^*}$, $\frac{\zeta_1}{\sigma^*}$, $\frac{\zeta_2}{\sigma^*}$ will be identifiable under certain conditions on $[\Pi_{11}', \Pi_{21}']$.

Assume $\frac{\zeta_1}{\sigma^*} = [\frac{\zeta_{11}}{\sigma^*}, 0]$, $\frac{\zeta_{21}}{\sigma^*} = [\frac{\zeta_{21}}{\sigma^*}, 0]$ where all the components of ζ_{11} and ζ_{21} are nonzero. The equation (**) is

$$[\frac{\zeta_{11}}{\sigma^*}, 0, \frac{\zeta_{21}}{\sigma^*}, 0] \begin{vmatrix} \Pi_{11}^* \\ \Pi_{11}^{**} \\ \Pi_{21}^* \\ \Pi_{21}^{**} \end{vmatrix} = C_2$$

The necessary and sufficient condition for the identification of the coefficients in the decision function is that

$$[\Pi_{11}^{*'}, \Pi_{21}^{*'}] \text{ has full column rank.}$$

An equivalent condition in terms of coefficients in the structural equation is as follows. To simplify the expression, let us rewrite the original system as

$$[B_{10}\ B_{11}] \begin{vmatrix} Y_{1t}^* \\ Y_{1t}^{**} \end{vmatrix} + \Gamma_{10} Z_t + \Gamma_{11} W_t = \varepsilon_{1t}$$

$$[B_{20} B_{21}] \begin{vmatrix} Y_{2t}^* \\ Y_{2t}^{**} \end{vmatrix} + \Gamma_{20} Z_t + \Gamma_{21} W_t = \varepsilon_{2t}$$

$$I^*_t = \gamma Z_t + \zeta_1 Y_{1t}^* + \zeta_2 Y_{2t}^* - \varepsilon_t$$

where $Y_{1t}' = (Y_{1t}^{*'}, Y_{1t}^{**'})$ and $Y_{2t}' = (Y_{2t}^{*'}, Y_{2t}^{**'})$.

Premultiply the matrix $\begin{bmatrix} B_{10} & B_{11} & 0 & 0 & \Gamma_{10} & \Gamma_{11} \\ 0 & 0 & B_{20} & B_{21} & \Gamma_{20} & \Gamma_{21} \end{bmatrix}$

by $\begin{bmatrix} B_1^{-1} & 0 \\ 0 & B_2^{-1} \end{bmatrix}$; it is

$$\begin{bmatrix} B_1^{-1} & 0 \\ 0 & B_2^{-1} \end{bmatrix} \begin{bmatrix} B_{10} & B_{11} & 0 & 0 & \Gamma_{10} & \Gamma_{11} \\ 0 & 0 & B_{20} & B_{21} & \Gamma_{20} & \Gamma_{21} \end{bmatrix} = \begin{bmatrix} I & 0 & -\Pi_{10} & -\Pi_{11} \\ 0 & I & -\Pi_{20} & -\Pi_{21} \end{bmatrix}$$

Hence

$$\begin{bmatrix} B_1^{-1} & 0 \\ 0 & B_2^{-1} \end{bmatrix} \begin{bmatrix} B_{11} & 0 & \Gamma_{11} \\ 0 & B_{21} & \Gamma_{21} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -\Pi_{1\bar{1}}^{*} \\ I & 0 & -\Pi_{1\bar{1}}^{**} \\ 0 & 0 & -\Pi_{2\bar{1}}^{*} \\ 0 & I & -\Pi_{2\bar{1}}^{**} \end{bmatrix}$$

Thus $\begin{bmatrix} \Pi_{1\bar{1}}^{*} \\ \Pi_{2\bar{1}}^{*} \end{bmatrix}$ has full row rank is equivalent to that

$\begin{bmatrix} B_{11} & 0 & \Gamma_{11} \\ 0 & B_{21} & \Gamma_{21} \end{bmatrix}$ has full row rank. This identification condition is similar to the rank condition for usual simultaneous equation models with zero restrictions.

As pointed out in previous sections, parameters of the disturbances cannot be identifiable without strong assumptions. This is also the case for the multivariate and simultaneous equation system. Under the assumption that disturbances in different regimes are independent, all parameters will be identifiable. Anyway, it should be noted that var (ε_{1t}) , var (ε_{2t}) , cov $(\varepsilon_{1t}\varepsilon_{ot})$ and cov $(\varepsilon_{2t}, \varepsilon_{ot})$ will also be identifiable if the structural coefficients are identified. This is so, since var (v_{1t}) , var (v_{2t}) , cov $(v_{1t}\ \varepsilon_{ot})$ and cov $(v_{2t}\varepsilon_{ot})$ are always identifiable. As

$$\text{var}(v_{1t}) = B_1^{-1}\text{var}(\varepsilon_{1t})B_1'^{-1}$$

$$\text{var}(v_{2t}) = B_2^{-1}\text{var}(\varepsilon_{2t})B_2'^{-1}$$

$$\text{cov}(v_{1t}\varepsilon_{ot}) = B_1^{-1}\ \text{cov}(\varepsilon_{1t}\varepsilon'_{ot})$$

$$\text{and } \text{cov}(v_{2t}\varepsilon_{ot}) = B_2^{-1}\ \text{cov}(\varepsilon_{2t}\varepsilon'_{ot}) ,$$

it follows that

$$\text{var}(\varepsilon_{1t}) = B_1\ \text{var}(v_{1t})B_1'$$

$$\text{var}(\varepsilon_{2t}) = B_2\ \text{var}(v_{2t})B_2'$$

$$\text{cov}(\varepsilon_{1t}\varepsilon'_{ot}) = B_1\text{cov}(v_{1t}\varepsilon_{ot})$$

$$\text{cov}(\varepsilon_{2t}\varepsilon'_{ot}) = B_2\text{cov}(v_{2t}\varepsilon_{ot}) .$$

Now let us consider the estimations for these models. As the multivariate system can be regarded as the reduced form of the simultaneous equation system, it is sufficient to consider the estimation procedure for the simultaneous equation system. The two stage estimation procedures discussed are ready to be applied to each equation in the reduced form and the decision function. So the reduced form parameters Π and $\frac{\gamma}{\sigma^*}, \frac{\zeta_1}{\sigma^*}, \frac{\zeta_2}{\sigma^*}$ can be consistently estimated. If the simultaneous equations in each regime are exactly identified, the structural coefficients can be derived uniquely from them. This is similar to indirect least squares procedure. If equations are overidentified, it would be more appropriated to follow other procedures. One of the procedures that can be used is again a two stage procedure.

Without loss of generality, let us consider the first structural equation in regime 1. The first structural equation can be rewritten as

$$Y_{11t} = Y_{12t}\beta_{112} + Y_{13t}\beta_{113} + \cdots + Y_{1G_1t}\beta_{11G_1} + \alpha_{110} +$$

$$X_{1t}\alpha_{111} + \cdots + X_{kt}\alpha_{11k} + \varepsilon_{11t}$$

Denote the reduced form of equations in regime 1 as

$$Y_{1t} = \tilde{\Pi}_1 X_t + v_{1t}\ , \qquad \tilde{\Pi}_1 = [\tilde{\Pi}_{11},\ \tilde{\Pi}_{12},\ldots,\tilde{\Pi}_{1G_1}]$$

where X_t's are exogeneous variables. Based on the subsamples corresponding to $I_t = 1$, we have

$$Y_{11t} = (\tilde{\Pi}_{12}X_t)\beta_{112} + \cdots + (\tilde{\Pi}_{1G_1}X_t)\beta_{11G_1} + \alpha_{110} +$$

$$X_{1t}\alpha_{111} + \cdots + X_{kt}\alpha_{11k} - \sigma_{11\varepsilon}f(\Psi_t)/F(\Psi_t) + v_{11t}$$

where

$$\Psi_t = \left(\frac{\gamma}{\sigma^*} + \frac{\zeta_1}{\sigma^*}\Pi_{10} + \frac{\zeta_2}{\sigma^*}\Pi_{20}\right)Z_t + \left(\frac{\zeta_1}{\sigma^*}\Pi_{11} + \frac{\zeta_2}{\sigma^*}\Pi_{21}\right)W_t$$

and

$$E(v_{11t} \mid I_t = 1) = 0.$$

After the estimated Π and the estimated Ψ_t are substituted into the above equation, ordinary least squares can then be applied to estimate β_{11} and α_{11} consistently. Similarly, we can estimate all the other structural coefficients. If there are equality restrictions on coefficients in corresponding equations in different regimes, the equations can be combined as pointed out in the previous model by the D method and two stage method is then applied.

It remains now to estimate the identifiable covariance parameters of the disturbances. It is noted that

$$E(v_{1it}v_{1jt}|I_t = 1) = \sigma_{v_{1i}v_{1j}} - \sigma_{v_{1i}\varepsilon_o}\sigma_{v_{1j}\varepsilon_o}\Psi_t \frac{\phi(\Psi_t)}{\Phi(\Psi_t)}$$

$$\forall \quad i,j = 1,\ldots,G_1$$

where $\sigma_{v_{1i}v_{1j}} = \mathrm{cov}(v_{1it}, v_{1jt})$, $\sigma_{v_{1i}\varepsilon_o} = \mathrm{cov}(v_{1it}, \varepsilon_{ot})$

Similarly,

$$E(v_{2it}v_{2jt}|I_t = 0) = \sigma_{v_{2i}v_{2j}} + \sigma_{v_{2i}\varepsilon_o}\sigma_{v_{2j}\varepsilon_o}\Psi_t \frac{\phi(\Psi_t)}{1-\Phi(\Psi_t)}$$

$$\forall \quad i,j = 1,\ldots,G_2$$

With these equations and the estimated residuals, $\mathrm{var}(v_{1t})$, $\mathrm{var}(v_{2t})$ $\mathrm{cov}(v_{1t}\varepsilon_{ot})$ and $\mathrm{cov}(v_{2t}\varepsilon_{ot})$ can be consistently estimated by least squares. It follows $\mathrm{var}(\varepsilon_{1t})$, $\mathrm{var}(\varepsilon_{2t})$, $\mathrm{cov}(\varepsilon_{1t}\varepsilon_{ot})$ and $\mathrm{cov}(\varepsilon_{2t}\varepsilon_{ot})$ can be estimated.

With all the identifiable parameters consistently estimated, the two step maximum likelihood procedure is then applicable. For this simultaneous equation model, the likelihood function is

$$L(B_1,B_2,\Gamma_{10},\Gamma_{11},\Gamma_{20},\Gamma_{21},\ \Omega_1,\ \Omega_2,\ \Omega_{1\varepsilon_o},\Omega_{2\varepsilon_o}|\ Y,Z,W)$$

$$= \prod_{t=1}^{T} [\frac{|B_1|}{(2\Pi)^{G_1/2}\,|\Omega_1|^{1/2}}\ \exp\{-\tfrac{1}{2}(B_1Y_{1t} + \Gamma_{10}Z_t + \Gamma_{11}W_t)'\Omega_1^{-1}$$

$$(B_1Y_{1t} + \Gamma_{10}Z_t + \Gamma_{11}W_t)\}$$

$$\cdot \int_{-\infty}^{\Psi_t} \frac{1}{\sqrt{2\Pi}|1-\Omega_{1\varepsilon_o}\Omega_1^{-1}\Omega'_{1\varepsilon_o}|^{1/2}}\exp\{-\frac{[\varepsilon_{o_t}-\Omega_{1\varepsilon_o}\Omega_1^{-1}(B_1Y_{1t}+\Gamma_{10}Z_t+\Gamma_{11}W_t)]^2}{2(1-\Omega_{1\varepsilon_o}\Omega_1^{-1}\Omega'_{1\varepsilon_o})^{1/2}}\}$$

$$d\varepsilon_{ot}]^{I_t}\cdot[\frac{|B_2|}{(2\Pi)^{G_2/2}\,|\Omega_2|^{1/2}}\ \exp\{-\tfrac{1}{2}(B_2Y_{2t}+\Gamma_{20}Z_t+\Gamma_{21}W_t)'\Omega_2^{-1}\ (B_2Y_{2t}$$

$$+\Gamma_{20}Z_t+\Gamma_{21}W_t)\}\int_{\Psi_t}^{\infty}\frac{1}{\sqrt{2\Pi}|1-\Omega_{2\varepsilon_o}\Omega_2^{-1}\Omega'_{2\varepsilon_o}|^{1/2}}\cdot$$

$$\exp\{-\frac{[\varepsilon_{ot}-\Omega_{2\varepsilon_o}\Omega_2^{-1}(B_2Y_{2t}+\Gamma_{20}Z_t+\Gamma_{21}W_t)]^2}{2(1-\Omega_{2\varepsilon_o}\Omega_2^{-1}\Omega'_{2\varepsilon_o})}\}\ d\varepsilon_{ot}]^{1-I_t}$$

where

$\Omega_1 = \mathrm{var}(\varepsilon_{1t})$, $\Omega_2=\mathrm{var}(\varepsilon_{2t})$, $\Omega_{1\varepsilon_o}=\mathrm{cov}(\varepsilon_{1t}',\varepsilon_{ot})$ and $\Omega_{2\varepsilon_o}=\mathrm{cov}(\varepsilon_{2t}',\varepsilon_{ot})$.

5. Empirical Applications:

The models we have discussed are quite general. Many limited dependent variables in the literature can be regarded as special cases of these models. Among those, Tobin's model [27], Heckman's female labor supply model [11], Nelson's censored regression models [25], disequilibrium market models with sample separation such as models in Fair and Jaffee [5], Maddala and Nelson [21] and Goldfeld and Quandt [8], can be analysed and estimated by our procedures. We have also applied our models and procedures to other areas of empirical studies.

In Lee [16], we have studied the simultaneous effect of unionism on wage rates and workers decision to join labor unions. In that study, a three equations model with limited dependent variables and dichotomous endogeneous variables has been estimated with operatives data from SEO surveys. We found significance effects on both directions. An estimated average union nonunion wage differentials of about 15% was obtained.

In Lee and Trost [18], we have studied a housing expenditure model. The model differs from previous studies in that it takes into account the simultaneous determination of how much to spend and the housing purchasing decision. The model is estimated with survey data from A Panel Study of Income Dynamics. By using a maximum likelihood ratio test, we found evidence that simultaneity does exist.

In Lee [17], a modal choice of travel to work model with incomplete data has been estimated. In that model, we develope a five equations models -- two costs equations, two time equations and a decision equation. The data we used are again from A Panel Study of Income Dynamics. In that data set, the cost and time data for the chosen mode are available but not the alternatives. The model is for the whole population in U.S. economy rather than re-regional models. In this problem, we are interested in workers' choices of driving his own car or using public carrier to workplaces. We have found that besides the adequacy of public transportation, costs and time, personal characteristics such as sex and race also play an important role in choosing the travel modes. The value of time was found to be 21 percent of the average wage rate in our sample which is compatible with previous studies.

Footnotes

1. If Y_{1t} and Y_{2t} are always observable, the identification problem will be the same but estimation will be easier. In this case, equations Y_{1t} and Y_{2t} can be estimated by least squares and the decision function can be estimated by two stage probit method as discussed below.

2. Parameters β_1, β_2, σ_1^2, σ_2^2, $\sigma_{1\varepsilon_o}$ and $\sigma_{2\varepsilon_o}$ can be estimated consistently and hence they must be identifiable.

3. Discussions on various different recursive models with qualitative variables can be found in Maddala and Lee [20].

4. Similar procedures have also been discussed in the context of labor supply models in Gronau [9], Lewis [19], and Heckman [12].

References

1. Amemiya, T. "Regression Analysis When the Dependent Variable is Truncated Normal", Econometrica, Nov. 1973, pp. 997-1016.
2. Berndt, E.K., Hall, B. H., Hall, R. E., and Hausman, J.A., "Estimation and Inference in Nonlinear Structural Models", Annals of Economic and Social Measurement, 1974, pp. 653-665.
3. Cox, D. Analysis of Binary Data, Methuen, London, 1970.
4. Domencich, T. and D. McFadden, Urban Travel Demand: A Behavioral Analysis, North-Holland Publishing Company, Amsterdam, 1975.
5. Fair, R. C. and D. M. Jaffee, "Methods of Estimation for Markets in Disequilibrium", Econometrica, 1972, pp. 497-514.
6. Finney, D. Probit Analysis, 3rd Edition, Cambridge University Press, 1971.
7. Goldfeld, S. M. and R. E. Quandt, Nonlinear Methods in Econometrics, North-Holland, 1972.
8. Goldfeld, S. M. and R. E. Quandt, "Estimation in a Disequilibrium Model and the Value of Information", Journal of Econometrics 3, 1975, pp. 325-348.
9. Gronau, Reuben, "Wage Comparisons - A Selectivity Bias", Journal of Political Economy, November/December, 1974.
10. Haberman, Shelly, The Analysis of Frequency Data, University of Chicago Press, 1974.
11. Heckman, J. "Shadow Prices, Market Wages and Labor Supply", Econometrica, Vol. 42 (No. 4), July 1974, pp. 679-694.

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

13. Jennrich, R. "Asymptotic Properties of Non-Linear Least Squares Estimates", Annals of Mathematical Statistics, 1969, pp. 633-643.

14. Johnson, N. and S. Kotz, Distributions in Statistics: Continuous Multivariate Distributions. Wiley, New York, 1972.

15. Lee, L. F. "Estimation of Some Limited Dependent Variable Models by Two-Stage Methods", Discussion paper, University of Rochester, September, 1975.

16. Lee, L. F. "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables", forthcoming, International Economic Review.

17. Lee, L. F. "Estimations of A Modal Choice Model for the Work Journey with Incomplete Observations", Manuscript, 1977.

18. Lee, L. F. and R. P. Trost, "Estimation of Some Limited Dependent Variable Models With Application to Housing Demand", Manuscript, 1977.

19. Lewis, H. Gregg, "Comments on Selectivity Biases in Wage Comparisons", Journal of Political Economy, November/December, 1974.

20. Maddala, G. S. and L. F. Lee, "Recursive Models with Qualitative Endogeneous Variables", Annals of Economic and Social Measurement, 1976, Vol. 5, pp. 525-545.

21. Maddala, G. S. and F. D. Nelson, "Maximum Likelihood Methods for the Estimation of Models of Markets in Disequilibrium", Econometrica, 42, 1974, pp. 1013-1030.

22. Maddala, G. S. and F. D. Nelson, "Switching Regression Models with Exogenous and Endogeneous Switching" -- Proceedings of the Business and Economics Statistics Section, American Statistical Association, 1975, pp. 423-426.

23. Manski, C. F. and S. R. Lerman, "The Estimation of Choice Probabilities from Choice Based Samples", forthcoming, Econometrica.

24. McFadden, D.,"Conditional Logit Analysis of Qualitative Choice Behavior", in P. Zarembka, ed., Frontiers in Econometrics, Academic Press, New York.

25. Nelson, F.D., "Censored Regression Models with Unobserved, Stochastic, Censored Thresholds" -- Social Science Working Paper No. 115, March 1976, California Institute of Technology.

26. Rao, C. R., Linear Statistical Inference and Its Applications 2nd Edition, Wiley, 1973.

27. Tobin, J., "Estimation of Relationships for Limited Dependent Variables", Econometrica, 26 (1958), 24-36.

28. Westin, R. B., "Statistical Models for Interrelated Discrete and Continuous Choices", Paper presented at the Econometric Society, World Congress 3 at University of Toronto, Toronto.