IV SIMULTANEOUS EQUATIONS MODELS WITH DISCRETE ENDOGENOUS VARIABLES

9 Simultaneous Equations Models with Discrete and Censored Dependent Variables

Lung-Fei Lee

9.1 Introduction

Recently a class of econometric models involving dichotomous, limited, and censored dependent variables was introduced by Amemiya (1974), Heckman (1976a, 1976b, 1977), Lee (1976, 1977), Nelson and Olsen (1977), and others in econometrics literature. In this chapter we will investigate the estimation principle posed by Amemiya, using a unified general simultaneous equation model. The simultaneous equation model includes censored simultaneous equation models, switching simultaneous equation models, and Nelson-Olson and Heckman models without structural change as special cases. Estimation methods that are computationally simple and consistent are also proposed. Alternative estimates can be derived from Amemiya's principle (Amemiya 1977a, 1977b). The Amemiya principle is a general principle used to derive structural parameter estimates from estimated reduced form parameters. Amemiya proved that his principle can lead to simple estimators more efficient than the estimators derived in Nelson-Olsen (1977) and Heckman (1977).[1] Since Amemiya only demonstrated the efficiency of his approach in a case-by-case basis, one may wonder whether it still holds in more general and complicated models.

9.2 Two-Stage Methods and Amemiya's Principle

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

This chapter is a shortened version of a paper prepared for the NBER-NSF Conference on Decision Rules and Uncertainty, Carnegie-Mellon University, March 29–April 1, 1978, with a slightly different title: "On the Estimation of Probit Choice Model with Censored Dependent Variables and Amemiya's Principle." I would like to thank Jerry Hausman, James Heckman, G.S. Maddala, Charles F. Manski, Daniel McFadden, Randall Olsen, Richard Westin, and a referee for their comments. Any errors remaining are solely my own.

1. In the Nelson-Olson model he used one limited dependent and one continuous endogenous variable (Amemiya 1977a) and in Heckman's model a continuous and dichotomous dependent variable (Amemiya 1977b).

$$\mathbf{Y}_i = \mathbf{Y}_i\mathbf{B} + \mathbf{X}_i\mathbf{\Gamma} + \boldsymbol{\varepsilon}_i, \tag{9.1}$$

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

1. the first G_1 variables $Y_{1i}, \ldots, Y_{G_1 i}$ are observable continuous variables,

2. the next $G_2 - G_1$ variables $Y_{G_1+1i}, \ldots, Y_{G_2 i}$ are limited dependent variables, that is, one can observe it only when $Y_{ji} > 0$,

3. the next $G_3 - G_2$ variables $Y_{G_2+1i}, \ldots, Y_{G_3 i}$ are unobservable latent variables. However binary indicators I_{ji} are observable and are determined by the latent variable Y_{ji} as follows;

$I_{ji} = 1$ iff $Y_{ji} > 0$,
$I_{ji} = 0$ otherwise,
$j = G_2 + 1, \ldots, G_3$.

4. the last $G - G_3$ variables are censored dependent variables. The variables $Y_{G_3+1i}, \ldots, Y_{Gi}$ are censored by a subset of latent variables in the preceding assumption. Specifically the index set $\{G_3 + 1, \ldots, G\}$ can be partitioned into finite mutually exclusive and exhausted nonempty subsets $\mathbf{S}_k$, that is, $\{G_3 + 1, \ldots, G\} = U_{l=1}^{L}\mathbf{S}_l$ where $L \leq G_3 - G_2$. For each $1 \leq l \leq L$ there is a unique latent variable Y_{G_2+li} activating it; $Y_{j_l i}$ is observed if $Y_{G_2+li} > 0$, for all $j_l \in \mathbf{S}_l^*$, and $Y_{j_l i}$ is observed if $Y_{G_2+li} < 0$, for $j_l \in \mathbf{S}_l/\mathbf{S}_l^*$, where $\mathbf{S}_l^*$ is a subset of $\mathbf{S}_l$ which may be empty or equal $\mathbf{S}_l$.

The model in (9.1) is well defined and contains models developed by Lee (1976, 1977) as well as Heckman's models without structural shifts (1976, 1977) and Nelson and Olson (1977) models as special cases.

The $\mathbf{B}$ and $\mathbf{\Gamma}$ can be identified under rank conditions and suitable normalization rules. However, in general only certain nonlinear transformation of $\mathbf{\Sigma}$ will be identifiable when $G_3 < G$ and $\phi \neq \mathbf{S}_l^* \subsetneqq \mathbf{S}_l$ for some l. The analysis of identification conditions will be similar to that in Lee (1977), but the details will be omitted here.

To estimate model (9.1), maximum likelihood methods are too complicated to be useful. However, consistent methods proposed by Heckman (1976, 1977), Lee (1977), Maddala and Lee (1976), and Nelson and Olson (1977) can be easily extended. Alternative estimates can also be derived from Amemiya's principle (Amemiya 1977a, 1977b). All those methods require estimation of reduced form parameters in the first stage. For the model in (9.1) reduced form equations always exist:

$$\mathbf{Y}_i = \mathbf{X}_i\mathbf{\Pi} + \mathbf{u}_i, \tag{9.2}$$

where $\mathbf{\Pi} = \mathbf{\Gamma}(\mathbf{I} - \mathbf{B})^{-1}$ and $\mathbf{u}_i = \boldsymbol{\varepsilon}_i(\mathbf{I} - \mathbf{B})^{-1}$. Equation (9.2) can be estimated by a single equation method such as probit or tobit maximum likelihood methods, depending on the nature of the dependent variable. The variables are defined in terms of the four specified categories.

The second stage is to estimate the structural parameters. To simplify notations, each single equation is specified as

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + \mathbf{Y}_i^*\boldsymbol{\delta}_1 + \varepsilon_i, \tag{9.3}$$

where $\mathbf{Y}_i^*$ is a subvector of endogenous variables other than Y_i in $\mathbf{Y}_i$. Equation (9.3) can be modified to

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + (\mathbf{X}_i\mathbf{\Pi}^*)\boldsymbol{\delta}_1 + v_i, \tag{9.4}$$

where $\mathbf{Y}_i^* = \mathbf{X}_i\mathbf{\Pi}^* + \mathbf{u}_i^*$. With consistent estimates $\hat{\mathbf{\Pi}}^*$ derived in the first stage, the second stage in the methods proposed by Lee (1976, 1977), Maddala and Lee (1976) and Nelson and Olson (1977) is to estimate $(\boldsymbol{\delta}_0, \boldsymbol{\delta}_1)$ from

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + (\mathbf{X}_i\hat{\mathbf{\Pi}}^*)\boldsymbol{\delta}_1 + w_i, \tag{9.5}$$

where $w_i = v_i + \mathbf{X}_i(\mathbf{\Pi}^* - \hat{\mathbf{\Pi}}^*)\boldsymbol{\delta}_1$. Equation (9.5) is estimated by probit, tobit, and so on, depending on the nature of Y_i in (9.5).

Instead of estimating (9.5), Amemiya suggests one should solve by regression methods the structural parameters from the estimated reduced form parameters. Based on this principle, one can derive alternative estimates. Let $\mathbf{R} = \mathbf{X}\mathbf{J}_1$, where $\mathbf{J}_1$ consists of unit and zero column vectors. From (9.4) one has

$$Y_i = \mathbf{X}_i(\mathbf{J}_1\boldsymbol{\delta}_0 + \mathbf{\Pi}^*\boldsymbol{\delta}_1) + v_i. \tag{9.6}$$

Let $\mathbf{c}$ be the corresponding reduced form parameter vector of Y_i in (9.2). It is obvious that

$$\mathbf{c} = \mathbf{J}_1\boldsymbol{\delta}_0 + \boldsymbol{\Pi}^*\boldsymbol{\delta}_1. \tag{9.7}$$

The estimates suggested by Amemiya are ordinary least squares, OLS, or generalized least squares, GLS, estimates derived from

$$\hat{\mathbf{c}} = \mathbf{J}_1\boldsymbol{\delta}_0 + \hat{\boldsymbol{\Pi}}^*\boldsymbol{\delta}_1 + \xi, \tag{9.8}$$

where $\xi = \hat{\mathbf{c}} - \mathbf{c} - (\hat{\boldsymbol{\Pi}}^* - \boldsymbol{\Pi}^*)\boldsymbol{\delta}_1$.

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

9.3 Structural Equations with Probit Structure

The $G_3 - G_2$ equations have unobservable latent variables at the left-hand side that belong to this category. In our model the endogenous variable Y_i is an unobservable latent variable. The two-stage estimates are derived from maximizing the function ln L in (9.9) w.r.t. $\boldsymbol{\Theta}_1' = (\boldsymbol{\delta}_0', \boldsymbol{\delta}_1')$ which are the identifiable parameters under the normalization $\sigma_v^2 = 1$.

$$\begin{aligned}\ln L = \sum_{i=1}^{N} \{ & I_i \ln \Phi(\mathbf{R}_i\boldsymbol{\delta}_0 + (\mathbf{X}_i\hat{\boldsymbol{\Pi}}^*)\boldsymbol{\delta}_1) \\ & + (1 - I_i)\ln(1 - \Phi(\mathbf{R}_i\boldsymbol{\delta}_0 + (\mathbf{X}_i\hat{\boldsymbol{\Pi}}^*)\boldsymbol{\delta}_1))\},\end{aligned} \tag{9.9}$$

where I_i is the observed dichotomous indicator of Y_i, Φ is the standard normal c.d.f. Let $\mathbf{P} = [\mathbf{X}\delta_{11}, \ldots, \mathbf{X}\delta_{1M}]$, $\boldsymbol{\Theta}_2' = (\boldsymbol{\Pi}_1^{*\prime}, \ldots, \boldsymbol{\Pi}_M^{*\prime})$, $\mathbf{S} = [\mathbf{R}\mathbf{X}\boldsymbol{\Pi}^*]$, where $\boldsymbol{\delta}_1' = [\delta_{11}, \ldots, \delta_{1M}]$ and $\boldsymbol{\Pi}^* = (\boldsymbol{\Pi}_1^*, \ldots, \boldsymbol{\Pi}_M^*)$. Let

$$\boldsymbol{\Lambda}_1 = \begin{bmatrix} \dfrac{\phi_1}{1 - \Phi_1} & & 0 \\ & \ddots & \\ 0 & & \dfrac{\phi_N}{1 - \Phi_N} \end{bmatrix}$$

and

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

where ϕ_i and Φ_i are standard normal density and distribution functions evaluated at $\mathbf{R}_i\boldsymbol{\delta}_0 + \mathbf{X}_i\boldsymbol{\Pi}^*\boldsymbol{\delta}_1$. Following Amemiya, the asymptotic distribution of this two-stage estimator $\hat{\boldsymbol{\Theta}}_1$ can be derived from

$$\hat{\boldsymbol{\Theta}}_1 - \boldsymbol{\Theta}_1 \triangleq (\mathbf{S}'\Lambda\mathbf{S})^{-1}(\mathbf{S}'\Lambda_1(\mathbf{I} - \boldsymbol{\Phi}) - \mathbf{S}'\Lambda\mathbf{P}(\hat{\boldsymbol{\Theta}}_2 - \boldsymbol{\Theta}_2)), \tag{9.10}$$

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

PROPOSITION 9.1: For equation (9.3) with unobservable latent variable Y_i and its dichotomous realization I_i, the two-stage estimate $\hat{\boldsymbol{\Theta}}_1$ derived from maximizing equation (9.9) is asymptotically less efficient than the GLS estimate $\hat{\boldsymbol{\Theta}}_1^A$ derived from Amemiya's principle.

PROOF: From (9.10), the asymptotic variance of $\hat{\boldsymbol{\Theta}}_1$ is

$$\mathbf{V}_{\boldsymbol{\Theta}_1} = (\mathbf{S}'\Lambda\mathbf{S})^{-1}\{\mathbf{S}'\Lambda\mathbf{S} + \mathbf{S}'\Lambda\mathbf{P}\mathbf{V}_{\boldsymbol{\Theta}_2}\mathbf{P}'\Lambda\mathbf{S} - \mathbf{S}'\Lambda_1\mathbf{E}_2'\mathbf{P}'\Lambda\mathbf{S} \\ - \mathbf{S}'\Lambda\mathbf{P}\mathbf{E}_2\Lambda_1\mathbf{S}\}(\mathbf{S}'\Lambda\mathbf{S})^{-1},$$

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

$$\mathbf{V}_{\boldsymbol{\Theta}_1^A} = (\mathbf{Z}'\boldsymbol{\Omega}_\xi^{-1}\mathbf{Z})^{-1},$$

where $\mathbf{Z} = [\mathbf{J}_1\ \boldsymbol{\Pi}^*]$ and $\boldsymbol{\Omega}_\xi$ in this probit structural equation is

$$\boldsymbol{\Omega}_\xi = [\mathbf{I}\ \mathbf{P}_1]\begin{bmatrix} (\mathbf{X}'\Lambda\mathbf{X})^{-1} & * \\ -\mathbf{E}_2\Lambda_1\mathbf{X}(\mathbf{X}'\Lambda\mathbf{X})^{-1} & \mathbf{V}_{\boldsymbol{\Theta}_2} \end{bmatrix}\begin{bmatrix} \mathbf{I} \\ \mathbf{P}_1' \end{bmatrix},$$

with $\mathbf{P} = \mathbf{X}\mathbf{P}_1$. The two expressions $\mathbf{V}_{\boldsymbol{\Theta}_1}$ and $\boldsymbol{\Omega}_\xi$ follow, because $\mathbf{c}$ is the probit maximum likelihood estimate of the reduced form equation and $\hat{\mathbf{c}} - \mathbf{c} \triangleq (\mathbf{X}'\Lambda\mathbf{X})^{-1}\mathbf{X}'\Lambda_1(\mathbf{I} - \boldsymbol{\Phi})$. On the other hand, $\mathbf{V}_{\boldsymbol{\Theta}_1}$ can be rewritten as

$$\mathbf{V}_{\Theta_1} = (\mathbf{Z}'\mathbf{X}'\Lambda\mathbf{X}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}'\Lambda\mathbf{X}\Omega_\xi\mathbf{X}'\Lambda\mathbf{X}\mathbf{Z}(\mathbf{Z}'\mathbf{X}'\Lambda\mathbf{X}\mathbf{Z})^{-1}.$$

It follows that $\mathbf{V}_{\Theta_1} - \mathbf{V}_{\Theta_1{}^A}$ is nonnegative definite.

9.4 Structural Equations with Observable Continuous Endogenous Variables

The first G_1 equations in our model (9.1) are in this category. Y_i in (9.3) is an observable continuous variable. Let $\Theta_1' = (\delta_0', \delta_1')$, $\mathbf{S} = [\mathbf{R} \quad \mathbf{X}\Pi^*]$ be a matrix with $(\mathbf{R}_i \quad \mathbf{X}_i\Pi^*)$ in its ith row, and $\tilde{\mathbf{S}}$ be its estimated value. An OLS procedure can be applied to (9.5). The two-stage estimate is thus

$$\hat{\Theta}_1 = (\tilde{\mathbf{S}}'\tilde{\mathbf{S}})^{-1}\tilde{\mathbf{S}}'\mathbf{Y}. \tag{9.11}$$

This two-stage method is similar to Theil's two-stage least squares method (1971) and was used in Heckman (1976, 1977). Amemiya GLS estimate derived from (9.8) is

$$\hat{\Theta}_1^A = (\tilde{\mathbf{Z}}'\tilde{\Omega}_\xi^{-1}\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\tilde{\Omega}'^{-1}_\xi\hat{\mathbf{c}}, \tag{9.12}$$

where $\mathbf{Z} = [\mathbf{J}_1 \quad \Pi^*]$, Ω_ξ is the variance matrix of ξ, and $\tilde{\mathbf{Z}}$ and $\tilde{\Omega}_\xi$ are their estimated values.

Using the two-equation Nelson-Olson and Heckman models, Amemiya derived the asymptotic variance matrices for $\hat{\Theta}_1$ and $\hat{\Theta}_1^A$. He also gave separate proofs in the two models that $\hat{\Theta}_1^A$ is more efficient than $\hat{\Theta}_1$. For our model the asymptotic covariance matrices are quite lengthy but can be derived in a straightforward manner. The interesting thing is to compare their efficiency. This follows in proposition 9.2.

PROPOSITION 9.2: For equation (9.3) with observable continuous variable Y_i the estimate $\hat{\Theta}_1^A$ in (9.12) is asymptotically more efficient than the estimate $\hat{\Theta}_1$ in (9.11).

PROOF: Let $\mathbf{P} = [\mathbf{X}\delta_{11}, \ldots, \mathbf{X}\delta_{1M}]$, $\Theta_2' = (\Pi_1^{*\prime}, \ldots, \Pi_M^{*\prime})$, where $\delta_1' = [\delta_{11}, \ldots, \delta_{1M}]$ and $\Pi^* = (\Pi_1^*, \ldots, \Pi_M^*)$. Denote the asymptotic variance matrix of $\hat{\Theta}_2' = (\hat{\Pi}_1^{*\prime}, \ldots, \hat{\Pi}_M^{*\prime})$ by $\mathbf{V}_{\Theta_2}$ and the asymptotic covariance of $\hat{\Theta}_2$ and v by $\mathbf{E}_1$. From (9.5) it is obvious that the variance matrix of $\hat{\Theta}_1$ is

$$\mathbf{V}_{\Theta_1} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\{\sigma_v^2\mathbf{I} + \mathbf{P}\mathbf{V}_{\Theta_2}\mathbf{P}' - \mathbf{P}\mathbf{E}_1 - \mathbf{E}_1'\mathbf{P}'\}\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}.$$

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

$$\mathbf{V}_{\Theta_1^A} = (\mathbf{Z}'\boldsymbol{\Omega}_\xi^{-1}\mathbf{Z})^{-1}.$$

To compare $\mathbf{V}_{\bar{\Theta}_1}$ and $\mathbf{V}_{\Theta_1^A}$, one notes from (9.8) that

$$\boldsymbol{\Omega}_\xi = [\mathbf{I} - \mathbf{P}_1]\begin{bmatrix} \sigma_v^2(\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}_1' \\ * & \mathbf{V}_{\Theta_2} \end{bmatrix}\begin{bmatrix} \mathbf{I} \\ -\mathbf{P}_1' \end{bmatrix},$$

where $\mathbf{P}_1 = \boldsymbol{\delta}_1' \otimes \mathbf{I}$, $\otimes$ being the Kronecker product. Since $\mathbf{P} = \mathbf{X}\mathbf{P}_1$ and $\mathbf{S} = \mathbf{X}\mathbf{Z}$,

$$\begin{aligned} \mathbf{V}_{\Theta_1} &= (\mathbf{Z}'\mathbf{X}'\mathbf{X}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}'\{\sigma_v^2\mathbf{I} + \mathbf{X}\mathbf{P}_1\mathbf{V}_{\Theta_2}\mathbf{P}_1'\mathbf{X}' - \mathbf{X}\mathbf{P}_1\mathbf{E}_1 \\ &\quad - \mathbf{E}_1'\mathbf{P}_1'\mathbf{X}'\}\mathbf{X}\mathbf{Z}(\mathbf{Z}'\mathbf{X}'\mathbf{X}\mathbf{Z})^{-1} \\ &= (\mathbf{Z}'\mathbf{X}'\mathbf{X}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}'\mathbf{X}\boldsymbol{\Omega}_\xi\mathbf{X}'\mathbf{X}\mathbf{Z}(\mathbf{Z}'\mathbf{X}'\mathbf{X}\mathbf{Z})^{-1}. \end{aligned}$$

It follows $\mathbf{V}_{\Theta_1} - \mathbf{V}_{\Theta_1^A}$ is nonnegative definite, and $\hat{\boldsymbol{\Theta}}_1^A$ is more efficient.

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

9.5 Structural Equations with Censored Dependent Variables

The last $G - G_3$ equations belong to this category. The variable Y_i is censored. When $\mathbf{S}_l^*$ and its complement set $\mathbf{S}_l/\mathbf{S}_l^*$ are nonempty, there are switching systems. For the variable Y in $\mathbf{S}_l^*$ which is observed when $Y_{G_2+l,i} > 0$, based on observed subsamples, equation (9.4) can be rewritten

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + (\mathbf{X}_i\boldsymbol{\Pi}^*)\boldsymbol{\delta}_1 + \lambda\frac{\phi(\mathbf{X}_i\boldsymbol{\alpha})}{\Phi(\mathbf{X}_i\boldsymbol{\alpha})} + \xi_i, \tag{9.13}$$

where $E(\xi_i \mid I_i = 1) = 0$, I_i is the dichotomous indicator of the underlying latent variable that activates the censoring, and $\boldsymbol{\alpha}$ is the reduced form parameters of that latent variable. The two-stage estimator is to find $\boldsymbol{\alpha}$ and $\boldsymbol{\Pi}^*$ in the first stage from probit maximum likelihood and similar equations in (9.19) and estimate $\boldsymbol{\Theta}_1' = (\boldsymbol{\delta}_0', \boldsymbol{\delta}_1', \lambda)$ in the second stage from

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + (\mathbf{X}_i\hat{\boldsymbol{\Pi}}^*)\boldsymbol{\delta}_1 + \lambda\frac{\phi(\mathbf{X}_i\hat{\boldsymbol{\alpha}})}{\Phi(\mathbf{X}_i\hat{\boldsymbol{\alpha}})} + \eta_i, \tag{9.14}$$

where

$$\eta_i = \xi_i - \lambda\left(\frac{\phi(\mathbf{X}_i\hat{\boldsymbol{\alpha}})}{\Phi(\mathbf{X}_i\hat{\boldsymbol{\alpha}})} - \frac{\phi(\mathbf{X}_i\boldsymbol{\alpha})}{\Phi(\mathbf{X}_i\boldsymbol{\alpha})}\right) - \mathbf{X}_i(\hat{\boldsymbol{\Pi}}^* - \boldsymbol{\Pi}^*)\boldsymbol{\delta}_1.$$

For the variable Y in $\mathbf{S}_i/\mathbf{S}_i^*$, (9.14) should be modified to

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + (\mathbf{X}_i\hat{\mathbf{\Pi}}^*)\boldsymbol{\delta}_1 + \lambda\frac{\phi(\mathbf{X}_i\hat{\boldsymbol{\alpha}})}{1 - \Phi(\mathbf{X}_i\hat{\boldsymbol{\alpha}})} + \eta_i, \tag{9.15}$$

where

$$\eta_i = \xi_i - \lambda\left(\frac{\phi(\mathbf{X}_i\hat{\boldsymbol{\alpha}})}{1 - \Phi(\mathbf{X}_i\hat{\boldsymbol{\alpha}})} - \frac{\phi(\mathbf{X}_i\boldsymbol{\alpha})}{1 - \Phi(\mathbf{X}_i\boldsymbol{\alpha})}\right) - \mathbf{X}_i(\hat{\mathbf{\Pi}}^* - \mathbf{\Pi}^*)\boldsymbol{\delta}_1.$$

All these expressions can be represented as

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + (\mathbf{X}_i\hat{\mathbf{\Pi}}^*)\boldsymbol{\delta}_1 + \lambda G(\mathbf{X}_i\hat{\boldsymbol{\alpha}}) + \eta_i, \tag{9.16}$$

where $G(\mathbf{X}_i\boldsymbol{\alpha})$ is a nonlinear function of $\mathbf{X}_i\boldsymbol{\alpha}$ and

$$\eta_i = \xi_i - \lambda(G(\mathbf{X}_i\hat{\boldsymbol{\alpha}}) - G(\mathbf{X}_i\boldsymbol{\alpha})) - \mathbf{X}_i(\hat{\mathbf{\Pi}}^* - \mathbf{\Pi}^*)\boldsymbol{\delta}_1.$$

Let $\mathbf{S} = [\mathbf{R} \quad \mathbf{X}\mathbf{\Pi}^* \quad \mathbf{G}]$ and $\tilde{\mathbf{S}}$ be its estimated value, where $\mathbf{G}$ is the vector consisted of $G(\mathbf{X}_i\boldsymbol{\alpha})$. The two-stage estimator $\hat{\mathbf{\Theta}}_1$ from (9.16) is

$$\hat{\mathbf{\Theta}}_1 = (\tilde{\mathbf{S}}'\tilde{\mathbf{S}})^{-1}\tilde{\mathbf{S}}'\mathbf{Y}. \tag{9.17}$$

Let $\mathbf{P} = \mathbf{X}(\boldsymbol{\delta}_1' \otimes \mathbf{I})$ and $\rho = \xi - \lambda\mathbf{D}_1(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$, where $\mathbf{D}_1$ is the gradient matrix of G evaluated at $\boldsymbol{\alpha}$. The asymptotic covariance matrix can be derived from

$$\hat{\mathbf{\Theta}}_1 - \mathbf{\Theta}_1 \triangleq (\tilde{\mathbf{S}}'\tilde{\mathbf{S}})^{-1}\tilde{\mathbf{S}}'(\rho - \mathbf{P}(\hat{\mathbf{\Theta}}_2 - \mathbf{\Theta}_2)). \tag{9.18}$$

To derive the Amemiya's GLS, one has to estimate the reduced form parameter $\mathbf{c}' = (\mathbf{c}_1', \mathbf{c}_2')$ for Y_i from

$$Y_i = \mathbf{X}_i\mathbf{c}_1 + G(\mathbf{X}_i\hat{\boldsymbol{\alpha}})\mathbf{c}_2 + \rho_i. \tag{9.19}$$

It is $\hat{\mathbf{c}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{Y}$, where $\mathbf{W} = [\mathbf{X} \quad \mathbf{G}]$. Let $\mathbf{Z} = [\mathbf{J}_1 \quad \mathbf{J}_2\mathbf{\Pi}^* \quad \mathbf{J}_3]$ be defined from $\mathbf{S} = \mathbf{W}[\mathbf{J}_1 \quad \mathbf{J}_2\mathbf{\Pi}^* \quad \mathbf{J}_3]$. The GLS derived from Amemiya's principle is

$$\hat{\mathbf{\Theta}}_1{}^A = (\tilde{\mathbf{Z}}'\tilde{\mathbf{\Omega}}_\omega^{-1}\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\tilde{\mathbf{\Omega}}_\omega^{-1}\hat{\mathbf{c}}, \tag{9.20}$$

where $\mathbf{\Omega}_\omega$ is the asymptotic covariance matrix of $\boldsymbol{\omega} = \hat{\mathbf{c}} - \mathbf{c} + \mathbf{J}_2(\mathbf{\Pi}^* - \hat{\mathbf{\Pi}}^*)\boldsymbol{\delta}_1$. The comparisons of $\hat{\mathbf{\Theta}}_1$ and $\hat{\mathbf{\Theta}}_1^A$ follow from proposition 9.3.

PROPOSITION 9.3: In the equation with censored dependent variable Y_i, the GLS estimate (9.20) derived from Amemiya's principle is asymptotically more efficient than the two-stage estimator $\hat{\mathbf{\Theta}}_1$ in (9.17)

PROOF: From (9.18) the asymptotic covariance matrix of $\hat{\mathbf{\Theta}}_1$ is

$$\mathbf{V}_{\mathbf{\Theta}_1} = (\mathbf{Z}'\mathbf{W}'\mathbf{W}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{W}'(\mathbf{V}_\rho + \mathbf{P}\mathbf{V}_{\mathbf{\Theta}_2}\mathbf{P}' - \mathbf{P}\mathbf{E} - \mathbf{E}'\mathbf{P}')\mathbf{W}\mathbf{Z}(\mathbf{Z}'\mathbf{W}'\mathbf{W}\mathbf{Z})^{-1},$$

where $\mathbf{V}_\rho$ is the asymptotic variance matrix of $\boldsymbol{\rho}$, $\mathbf{E}$ is the asymptotic covariance matrix of $\hat{\mathbf{\Theta}}_2 - \mathbf{\Theta}_2$ and $\boldsymbol{\rho}$. On the other hand, $\mathbf{V}_{\mathbf{\Theta}_1^A} = (\mathbf{Z}'\mathbf{\Omega}_\omega^{-1}\mathbf{Z})^{-1}$ from (9.20). Since $\hat{\mathbf{c}} - \mathbf{c} \triangleq (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\rho}$, it implies

$$\begin{aligned}\mathbf{\Omega}_\omega = {} & (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}_\rho\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1} + \mathbf{J}_2(\boldsymbol{\delta}_1' \otimes \mathbf{I})\mathbf{V}_{\theta_2}(\boldsymbol{\delta}_1 \otimes \mathbf{I})\mathbf{J}_2' \\ & - \mathbf{J}_2(\boldsymbol{\delta}_1' \otimes \mathbf{I})\mathbf{E}\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1} - (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{E}'(\boldsymbol{\delta}_1 \otimes \mathbf{I})\mathbf{J}_2'.\end{aligned}$$

Since $\mathbf{P} = \mathbf{W}\mathbf{J}_2(\boldsymbol{\delta}_1' \otimes \mathbf{I})$, $\mathbf{V}_{\mathbf{\Theta}_1}$ can be rewritten as

$$\mathbf{V}_{\mathbf{\Theta}_1} = (\mathbf{Z}'\mathbf{W}'\mathbf{W}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{W}'\mathbf{W}\mathbf{\Omega}_\omega\mathbf{W}'\mathbf{W}\mathbf{Z}(\mathbf{Z}'\mathbf{W}'\mathbf{W}\mathbf{Z})^{-1}.$$

It follows that $\mathbf{V}_{\mathbf{\Theta}_1} - \mathbf{V}_{\mathbf{\Theta}_1^A}$ is nonnegative definite.

It is interesting to point out that this conclusion applies not only to the censored dependent variables in (9.14) and (9.15) but also to other models. Amemiya (1977b) derives the asymptotic covariance matrix for a two-equation Heckman model with structural shift but fails to apply his principle to that model. Consider the following equation with structural shift;

$$Y_{2i} = \mathbf{R}_i\boldsymbol{\delta}_0 + Y_{1i}\delta_1 + I_i\delta_2 + \varepsilon_i, \tag{9.21}$$

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

$$Y_{1i} = \mathbf{X}_i\mathbf{c} + u_i. \tag{9.22}$$

Equation (9.21) is modified to

$$Y_{2i} = \mathbf{R}_i\boldsymbol{\delta}_0 + (\mathbf{X}_i\mathbf{c})\delta_1 + \Phi(\mathbf{X}_i\mathbf{c})\delta_2 + \xi_i, \tag{9.23}$$

where Φ is standard normal c.d.f. With consistent estimates $\hat{\mathbf{c}}$ available, Heckman's two-stage estimator for $\mathbf{\Theta}_1' = (\boldsymbol{\delta}_0', \boldsymbol{\delta}_1', \boldsymbol{\delta}_2)$ is derived by least squares applied to

$$Y_{2i} = \mathbf{R}_i\boldsymbol{\delta}_0 + (\mathbf{X}_i\hat{\mathbf{c}})\delta_1 + \Phi(\mathbf{X}_i\hat{\mathbf{c}})\delta_2 + \eta_i. \tag{9.24}$$

It is obvious that equation (9.24) is a special case of (9.16). Hence it is possible to apply Amemiya's principle to equation (9.21), and more efficient estimates can be derived.

9.6 Structural Equations with Tobit Structure

This is the case where Y_i in (9.3) is limited dependent. Equations $G_1 + 1, \ldots, G_2$ in the general model (9.2) are in this category. A two-stage estimator $\hat{\boldsymbol{\Theta}}_1$ was proposed by Nelson and Olson (1977). This estimation method can be generalized to our model. The two-stage estimators are derived by maximizing the following function:

$$\ln L(\boldsymbol{\Theta}_1) = \sum_{i=1}^{N} \left\{ -\frac{1}{2} I_i \ln \sigma_v^2 - \frac{1}{2\sigma^2} I_i (Y_i - \mathbf{R}_i \boldsymbol{\delta}_0 - (\mathbf{X}_i \hat{\boldsymbol{\Pi}}^*)\boldsymbol{\delta}_1)^2 \right.$$
$$\left. + (1 - I_i) \ln (1 - \Phi(\mathbf{R}_i \boldsymbol{\delta}_0 + (\mathbf{X}_i \hat{\boldsymbol{\Pi}}^*)\boldsymbol{\delta}_1)) \right\}. \qquad (9.25)$$

Define

$$a_{11i} = \frac{\dfrac{S_i\Theta_1}{\sigma_v}\phi_i - \dfrac{\phi_i^2}{1 - \Phi_i} - \Phi_i}{\sigma_v^2},$$

$$a_{22i} = \frac{1}{4\sigma_v^4}\left(\phi_i\left(\frac{S_i\Theta_1}{\sigma_v}\right)^3 + \frac{S_i\Theta_1}{\sigma_v}\phi_i - \left(\frac{S_i\Theta_1}{\sigma_v}\right)^2 \frac{\phi_i^2}{1 - \Phi_i} - 2\Phi_i\right),$$

$$a_{12i} = \frac{-\phi_i}{2\sigma_v^3}\left(\left(\frac{S_i\Theta_1}{\sigma_v^2}\right)^2 + 1 - \frac{S_i\Theta_1}{\sigma_v}\frac{\phi_i}{1 - \Phi_i}\right),$$

where $S_i = [\mathbf{R}_i \mathbf{X}_i \boldsymbol{\Pi}^*]$ and $\boldsymbol{\Theta}_1' = (\boldsymbol{\delta}_0', \boldsymbol{\delta}_1')$.

Let $\mathbf{A}_{ij}$ be the $N \times N$ diagonal matrix whose kth diagonal element is a_{ijk}, $i,j = 1, 2$, and

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

It is easy to show that

$$E\left(\frac{\partial^2 \ln L(\boldsymbol{\Theta}_1)}{\partial \boldsymbol{\Theta}_1 \partial \boldsymbol{\Theta}_1'}\right) = \mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*\mathbf{Z}^*,$$

$$E\left(\frac{\partial^2 \ln L(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2)}{\partial \boldsymbol{\Theta}_1 \partial \boldsymbol{\Theta}_2'}\right) = -\mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{P}^*,$$

and

$$\frac{\partial \ln L(\Theta_1)}{\partial \Theta_1} = \mathbf{Z}^{*\prime}\mathbf{X}^{*}\mathbf{u},$$

where

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{l} \end{bmatrix}, \quad \mathbf{Z}^* = \begin{bmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}, \quad \mathbf{P}^* = \begin{bmatrix} \mathbf{P} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{P} = \mathbf{X}\mathbf{P}_1, \quad \mathbf{P}_1 = \delta_1' \otimes \mathbf{I},$$

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

$$\mathbf{u}' = \Bigg[\frac{1}{\sigma_v}\frac{\phi_1}{1-\Phi_1}(1-I_1) + \frac{1}{\sigma_v^2}I_1 v_1, \ldots, \frac{1}{\sigma_v}\frac{\phi_N}{1-\Phi_N}(1-I_N) + \frac{1}{\sigma_v^2}I_N v_N,$$
$$\frac{S_1\Theta_1}{2\sigma_v^3}\frac{\phi_1}{1-\Phi_1}(1-I_1) - \frac{1}{2\sigma_v^2}I_1 + \frac{1}{2\sigma_v^4}I_1 v_1^2,$$
$$\ldots, \frac{S_N\Theta_N}{2\sigma_v^3}\frac{\phi_N}{1-\Phi_N}(1-I_N) - \frac{1}{2\sigma_v^2}I_N + \frac{1}{2\sigma_v^4}I_N v_N^2\Bigg].$$

It follows from the Taylor series expansion

$$\hat{\Theta}_1 - \Theta_1 \triangleq \mathbf{L}_1[\mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*\mathbf{Z}^*]^{-1}(\mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{u} - \mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{P}^*(\hat{\Theta}_2 - \Theta_2)), \tag{9.26}$$

where $\mathbf{L}_1 = [\mathbf{I}, \mathbf{0}]$ is an identity matrix augmented with a column of zeros. The detailed expression for $\mathbf{V}_{\Theta_1}$ is quite lengthy but can be derived in a straightforward manner as in Amemiya (1977a).

With consistent estimator $\hat{\Theta}_2$ from the first stage and $\hat{\mathbf{c}}$ from the tobit maximum likelihood, one can compare the two-stage estimate in (9.26) with Amemiya's generalized least squares estimate $\hat{\Theta}_1^A$.

PROPOSITION 9.4: For equation (9.3) with limited dependent variable Y_i, the two-stage estimator $\hat{\Theta}_1$ derived by maximizing equation (9.25) is asymptotically less efficient than $\hat{\Theta}_1^A$.

PROOF: From (9.26) the asymptotic variance of $\hat{\Theta}_1$ is

$$\begin{aligned}\mathbf{V}_{\Theta_1} = \mathbf{L}_1(&\mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*\{(\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1} \\ &+ \mathbf{P}_1^*\mathbf{V}_{\theta_2}\mathbf{P}_1^{*\prime} - (\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{E}\mathbf{P}_1^* \\ &- \mathbf{P}_1^*\mathbf{E}'\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1}\}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*\mathbf{Z}^*)^{-1}\mathbf{L}_1',\end{aligned}$$

where $\mathbf{P}_1^* = [\begin{smallmatrix}\mathbf{P}_1\\ 0\end{smallmatrix}]$ and $\mathbf{E}$ is the asymptotic covariance matrix of $\mathbf{u}$ and $(\hat{\boldsymbol{\Theta}}_2 - \boldsymbol{\Theta}_2)'$. To compare the asymptotic variance of $\hat{\boldsymbol{\Theta}}_1^A$, one notes that $\hat{\mathbf{c}}$ is a tobit maximum likelihood estimate, and hence

$$\hat{\mathbf{c}} - \mathbf{c} \triangleq - \mathbf{L}_2(\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{u},$$

where $\mathbf{L}_2$ is an appropriate identity matrix augmented with a zero column.

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

$$\mathbf{V}_{\boldsymbol{\Theta}_1^A} = (\mathbf{Z}'\boldsymbol{\Omega}_\xi^{-1}\mathbf{Z})^{-1},$$

where

$$\boldsymbol{\Omega}_\xi = \mathbf{L}_2(\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1}\mathbf{L}_2' + \mathbf{P}_1\mathbf{V}_{\boldsymbol{\Theta}_2}\mathbf{P}_1' - \mathbf{P}_1\mathbf{E}'\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1}\mathbf{L}_2' - \mathbf{L}_2(\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{E}\mathbf{P}_1'.$$

To compare $\mathbf{V}_{\boldsymbol{\Theta}_1}$ with $\mathbf{V}_{\boldsymbol{\Theta}_1^A}$, one has to evaluate $\mathbf{L}_1(\mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*\mathbf{Z}^*)^{-1}$ and $\mathbf{L}_2(\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1}\mathbf{L}_2'$. This can be done with the well-known formulas for finding the inverse of a partitioned matrix: let

$$\mathbf{B} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{l}(\mathbf{l}'\mathbf{A}_{22}\mathbf{l})^{-1}\mathbf{l}'\mathbf{A}_{21}.$$

It is easy to check that the following equalities hold:

$$\mathbf{L}_1(\mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*\mathbf{Z}^*)^{-1} = (\mathbf{Z}'\mathbf{X}'\mathbf{B}\mathbf{X}\mathbf{Z})^{-1}[\mathbf{I} - \mathbf{Z}'\mathbf{X}'\mathbf{A}_{12}\mathbf{l}(\mathbf{l}'\mathbf{A}_{22}\mathbf{l})^{-1}]; \tag{9.27}$$

$$\mathbf{L}_2(\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1} = (\mathbf{X}'\mathbf{B}\mathbf{X})^{-1}[\mathbf{I} - \mathbf{Z}'\mathbf{X}'\mathbf{A}_{12}\mathbf{l}(\mathbf{l}'\mathbf{A}_{22}\mathbf{l})^{-1}]; \tag{9.28}$$

$$\mathbf{L}_2(\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime} = (\mathbf{X}'\mathbf{B}\mathbf{X})^{-1}[\mathbf{X}' - \mathbf{Z}'\mathbf{X}'\mathbf{A}_{12}\mathbf{l}(\mathbf{l}'\mathbf{A}_{22}\mathbf{l})^{-1}\mathbf{l}']; \tag{9.29}$$

$$[\mathbf{I} - \mathbf{Z}'\mathbf{X}'\mathbf{A}_{12}\mathbf{l}(\mathbf{l}'\mathbf{A}_{22}\mathbf{l})^{-1}]\mathbf{Z}^{*\prime}\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^{*\prime}\mathbf{P}_1^* = \mathbf{Z}'\mathbf{X}'\mathbf{B}\mathbf{X}\mathbf{P}_1; \tag{9.30}$$

$$[\mathbf{I} - \mathbf{Z}'\mathbf{X}'\mathbf{A}_{12}\mathbf{l}(\mathbf{l}'\mathbf{A}_{22}\mathbf{l})^{-1}]\mathbf{Z}^{*\prime}\mathbf{X}^{*\prime} = \mathbf{Z}'\mathbf{X}'\mathbf{B}\mathbf{X}\mathbf{L}_2(\mathbf{X}^{*\prime}\mathbf{A}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}. \tag{9.31}$$

It follows from (9.27) through (9.31) that

$$\mathbf{V}_{\boldsymbol{\Theta}_1} = (\mathbf{Z}'\mathbf{X}'\mathbf{B}\mathbf{X}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}'\mathbf{B}\mathbf{X}\boldsymbol{\Omega}_\xi\mathbf{X}'\mathbf{B}\mathbf{X}\mathbf{Z}(\mathbf{Z}'\mathbf{X}'\mathbf{B}\mathbf{X}\mathbf{Z})^{-1},$$

and hence $\mathbf{V}_{\boldsymbol{\Theta}_1} - \mathbf{V}_{\boldsymbol{\Theta}_1^A}$ is nonnegative definite.

9.7 Switching and Censored Models with Sample Separation Information

Switching and censored simultaneous equations systems with sample separation are special cases of the general model introduced in section 9.2. These models correspond to the cases in which $G_2 = 0$, $G_3 = 1$, $\mathbf{S}_1 = \{G_3 + 1, \ldots, G\}$, and the unobserved latent variable Y_1 does not appear explicitly in other equations. The system is a switching simultaneous equation model when $\mathbf{S}_1^*$ is nonempty; $\mathbf{S}_1^* \neq \mathbf{S}_1$, the endogenous variables in $\mathbf{S}_1^*$ and $\mathbf{S}_1/\mathbf{S}_1^*$, form a complete simultaneous equation system in each regime; and the endogenous variables in one regime do not appear in another one. This switching simultaneous equation model differs from the models studied in Goldfeld and Quandt (1972, 1973, 1976), since the sample separation information is available.[2] The censored simultaneous equation model introduced by Heckman (1974, 1976) can also be regarded as a special case in which either $\mathbf{S}_1^*$ is empty or $\mathbf{S}_1^* = \mathbf{S}_1$.

In this section we would consider procedures such as two-stage least squares, instrumental variables methods, and Amemiya's principle in the estimation of single structural equations in the system.[3] For each structural equation (9.3) in regime $\mathbf{S}_1^*$, based on observed subsamples corresponding to that regime, equation (9.3) can be rewritten as

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + \mathbf{Y}_i^*\boldsymbol{\delta}_1 + \lambda\frac{\phi(\mathbf{X}_i\boldsymbol{\alpha})}{\Phi(\mathbf{X}_i\boldsymbol{\alpha})} + \varepsilon_i^*, \tag{9.32}$$

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

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + \mathbf{Y}_i^*\boldsymbol{\delta}_1 + \lambda\frac{\phi(\mathbf{X}_i\boldsymbol{\alpha})}{1 - \Phi(\mathbf{X}_i\boldsymbol{\alpha})} + \varepsilon_i^*, \tag{9.33}$$

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

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + \mathbf{Y}_i^*\boldsymbol{\delta}_1 + \lambda G(\mathbf{X}_i\boldsymbol{\alpha}) + \varepsilon_i^*, \tag{9.34}$$

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

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

and therefore it is enough to consider the estimation of (9.34). Let $\hat{\alpha}$ be the probit maximum likelihood estimator of the reduced form Y_1. Equation (9.34) can be modified to

$$Y_i = \mathbf{R}_i\boldsymbol{\delta}_0 + \mathbf{Y}_i^*\boldsymbol{\delta}_1 + \lambda G(\mathbf{X}_i\boldsymbol{\alpha}) + \eta_i. \tag{9.35}$$

Let $\mathbf{H}$ be a matrix consisting of $(\mathbf{R}_i, \mathbf{Y}_i^*, G(\mathbf{X}_i\boldsymbol{\alpha}))$ as its ith row and $\tilde{\mathbf{H}}$ be its estimated value evaluated at $\hat{\boldsymbol{\alpha}}$. In matrix form equation (9.35) is

$$\mathbf{Y} = \tilde{\mathbf{H}}\boldsymbol{\Theta} + \boldsymbol{\eta}, \tag{9.36}$$

where $\boldsymbol{\Theta}' = (\boldsymbol{\delta}_0', \boldsymbol{\delta}_1', \lambda)$. The disturbances η_i in (9.36) are heteroscedastic and autocorrelated as pointed out in Lee, Maddala, and Trost (1977). There are several methods to estimate equation (9.36).

METHOD 9.1: Let $\tilde{\mathbf{X}}$ be a matrix with $(\mathbf{X}_i, G(\mathbf{X}_i\hat{\boldsymbol{\alpha}}))$ as its rows. Premultiply (9.36) by $\tilde{\mathbf{X}}$:

$$\tilde{\mathbf{X}}'\mathbf{Y} = \tilde{\mathbf{X}}'\tilde{\mathbf{H}}\boldsymbol{\Theta} + \tilde{\mathbf{X}}'\boldsymbol{\eta}. \tag{9.37}$$

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

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

with the asymptotic covariance matrix

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

where $\mathbf{V}_\eta$ is the covariance matrix of $\boldsymbol{\eta}$ and $\tilde{\mathbf{V}}_\eta$ is a consistent estimate of $\mathbf{V}_\eta$. Method 9.1 is similar to the usual two-stage least squares procedures; this method has been discussed in Heckman (1976) and Lee et al. (1977).

METHOD 9.2: Equation (9.37) is estimated by GLS with covariance matrix $\tilde{\mathbf{X}}'\tilde{\mathbf{V}}_\eta\tilde{\mathbf{X}}$. The estimator is

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

with the asymptotic covariance matrix $V(\hat{\boldsymbol{\Theta}}_{(2)}) = [\tilde{\mathbf{H}}'\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{V}}_\eta\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{H}}]^{-1}$. Method 9.2 differs from method 9.1 in that the correct asymptotic covariance matrix of $\tilde{\mathbf{X}}'\boldsymbol{\eta}$ is used.

METHOD 9.3: Premultiply equation (9.36) by $\tilde{\mathbf{X}}'\tilde{\mathbf{V}}_\eta^{-1}$;

$$\tilde{\mathbf{X}}'\tilde{\mathbf{V}}_\eta^{-1}\mathbf{Y} = \tilde{\mathbf{X}}'\tilde{\mathbf{V}}_\eta^{-1}\tilde{\mathbf{H}}\boldsymbol{\Theta} + \tilde{\mathbf{X}}'\tilde{\mathbf{V}}_\eta^{-1}\boldsymbol{\eta}. \tag{9.38}$$

The GLS procedure is then applied to (9.38). The estimator is

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

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

METHOD 9.4: Choose an instrumental variables matrix $\tilde{V}_\eta^{-1}\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{H}$. The following instrumental variables (IV) estimator can be derived;

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

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

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

It is of interest to compare these various estimators. Since $H = \tilde{X}Z + [0, V^*, 0]$, with $Z = [J_1, \Pi^*, J_3]$ as in (9.6), it can be easily shown that

$$\text{p}\lim \frac{1}{N_1}\tilde{X}'H = \text{p}\lim \frac{1}{N_1}\tilde{X}'\tilde{X}Z,$$

$$\text{p}\lim \frac{1}{N_1}\tilde{X}'V_\eta^{-1}H = \text{p}\lim \frac{1}{N_1}\tilde{X}'V_\eta^{-1}\tilde{X}Z,$$

where N_1 is the number of sample observations in the relevant regime. It follows that

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

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

and

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

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

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

4. $A \geq B$ means that $A - B$ is nonnegative definite.

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

The estimator $\hat{\Theta}_{(4)}$ derived from method 9.4 has an optimal property. It is the most efficient IV estimator in the set of IV estimators in the estimation of equation (9.36). This can be shown as follows. let $\mathbf{W}$ be an arbitrary instrumental variables matrix. The corresponding IV estimator is

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

with asymptotic covariance matrix $V(\hat{\Theta}_W) = (\mathbf{W}'\mathbf{H})^{-1}\mathbf{W}'\mathbf{V}_\eta\mathbf{W}(\mathbf{H}'\mathbf{W})^{-1}$. Since

$$\text{p}\lim N_1 V(\hat{\Theta}_W) = \text{p}\lim N_1[\mathbf{Z}'\tilde{\mathbf{X}}'\mathbf{W}(\mathbf{W}'\mathbf{V}_\eta\mathbf{W})^{-1}\mathbf{W}'\tilde{\mathbf{X}}\mathbf{Z}]^{-1}$$

and

$$\tilde{\mathbf{X}}'\mathbf{V}_\eta^{-1}\tilde{\mathbf{X}} \geq \tilde{\mathbf{X}}'\mathbf{W}(\mathbf{W}'\mathbf{V}_\eta\mathbf{W})^{-1}\mathbf{W}'\tilde{\mathbf{X}},$$

$\hat{\Theta}_{(4)}$ is asymptotically more efficient than $\hat{\Theta}_W$.

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

$$\hat{\mathbf{C}} = \tilde{\mathbf{Z}}\Theta + \omega, \tag{9.39}$$

where $\tilde{\mathbf{Z}} = [\mathbf{J}_1 \quad \hat{\mathbf{\Pi}}^* \quad \mathbf{J}_3]$, with $\hat{\mathbf{C}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{Y}$ and $\hat{\mathbf{\Pi}}^* = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{Y}^*$. The GLS estimator derived from Amemiya's principle is

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

As derived in section 9.4, $\hat{\mathbf{\Omega}}_\omega = \tilde{\mathbf{X}}'\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\hat{\mathbf{V}}_\eta\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$. It is obvious that $\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{H} = \tilde{\mathbf{X}}\tilde{\mathbf{Z}}$. It follows $\hat{\Theta}_G^A = \hat{\Theta}_{(2)}$, that is Amemiya's GLS procedure is exactly method 9.2. Therefore one concludes that the estimators $\hat{\Theta}_{(3)}$ and $\hat{\Theta}_{(4)}$ are asymptotically more efficient than the GLS estimator derived from Amemiya's principle. Let us now analyze the OLS estimator derived from Amemiya's principle. The OLS estimator is

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

It follows

$$\begin{aligned}\hat{\Theta}_L^A &= [\tilde{\mathbf{Z}}'\tilde{\mathbf{X}}'\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\tilde{\mathbf{Z}}]^{-1}\tilde{\mathbf{Z}}'\tilde{\mathbf{X}}'\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\hat{\mathbf{C}} \\ &= [\mathbf{H}'\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-2}\tilde{\mathbf{X}}'\mathbf{H}]^{-1}\mathbf{H}'\tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-2}\tilde{\mathbf{X}}'\mathbf{Y},\end{aligned}$$

which is the GLS procedure applied to 9.37 as if $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})$ is the covariance matrix. This estimator is less efficient than $\hat{\Theta}_{(2)}$, $\hat{\Theta}_{(3)}$, and $\hat{\Theta}_{(4)}$, but in general $\hat{\Theta}_L^A$ and $\hat{\Theta}_{(1)}$ will not dominate each other.

As a final remark we would like to point out that $\hat{\boldsymbol{\theta}}_{(3)}$ and $\hat{\boldsymbol{\theta}}_{(4)}$ are computationally as simple as the GLS estimator derived from Amemiya's principle. As demonstrated in Lee et al. (1977), $\mathbf{V}_\eta$ in the first regime is a sum of two matrices;

$$\mathbf{V}_\eta = \mathbf{V}_1 + \mathbf{D}_1(\mathbf{X}'\boldsymbol{\Lambda}\mathbf{X})^{-1}\mathbf{D}_1',$$

where $\mathbf{V}_1$ is a diagonal matrix. Hence the following inversion relation can be used;

$$\mathbf{V}_\eta^{-1} = \mathbf{V}_1^{-1} - \mathbf{V}_1^{-1}\mathbf{D}_1(\mathbf{X}'\boldsymbol{\Lambda}\mathbf{X} + \mathbf{D}_1'\mathbf{V}_1^{-1}\mathbf{D}_1)^{-1}\mathbf{D}_1'\mathbf{V}_1^{-1},$$

and we do not need to invert numerically an $N_1 \times N_1$ matrix. Similarly this is true for $\mathbf{V}_\eta$ in the second regime.

9.8 Conclusion

We have analyzed an estimation principle of Amemiya in a general simultaneous equation model. The model consists of observable continuous endogenous variables, unobservable latent endogenous variables with dichotomous indicator, and limited and censored dependent variables in a simultaneous equation framework. This general model contains the Nelson and Olson simultaneous tobit model, the Heckman simultaneous dummy endogenous variables model, the censored simultaneous equation model, and the switching simultaneous equations models as special cases. Various consistent two-stage estimation methods are generalized. Amemiya's principles are investigated in this general model, using an arbitrary number of equations. Amemiya's generalized two-stage estimators are compared with the other two-stage estimators. It was shown that Amemiya's estimators are more efficient in all the cases. Contrary to Amemiya (1977a) his principle can also be applied to Heckman's model with structural shift. A generalized two-stage estimator derived from his principle is also found to be more efficient than Heckman's approach. The proofs are general and do not depend on a case-by-case analysis.

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

References

Amemiya, T. 1974. Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal. *Econometrica*. 42; 999–1012.

Amemiya, T. 1976. The Specification and Estimation of A Multivariate Logit Model. Technical report 211. Institute for Mathematical Studies in the Social Sciences, Stanford University.

Amemiya, T. 1977a. The Estimation of a Simultaneous Equation Tobit Model. Technical report 236. Institute for Mathematical Studies in the Social Sciences, Stanford University. *International Economic Review*, in press.

Amemiya, T. 1977b. The Estimation of a Simultaneous Equation Generalized Probit Model. *Econometrica*. 46; 1193–1205.

Brundy, J. M., and D. W. Jorgenson. 1974. Consistent and Efficient Estimation of Systems of Simultaneous Equations by Means of Instrumental Variables. In *Frontiers in Econometrics*, ed. P. Zarembka. New York: Academic Press; pp. 215–244.

Goldfeld, S. M., and R. E. Quandt. 1972. *Nonlinear Methods in Econometrics*. Amsterdam; North-Holland.

Goldfeld, S. M., and R. E. Quandt. 1975. Estimation in a Disequilibrium Model and the Value of Information. *Journal of Econometrics*. 3; 325–348.

Goldfeld, S. M., and R. E. Quandt. 1976. Estimation of Structural Change in Simultaneous Equation Models. In *Studies in Nonlinear Estimation*, ed. S. M. Goldfeld and R. E. Quandt. Cambridge, Mass.; Ballinger, chapter 2.

Heckman, J. J. 1974. Shadow-Price Market Wages and Labor Supply. *Econometrica*. 42; 679–694.

Heckman, J. J. 1976a. The Common Structure of a Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*. 5; 475–492.

Heckman, J. J. 1976b. Simultaneous Equation Models with Both Continuous and Discrete Endogenous Variables with and Without Structural Shift in the Equations. In *Studies in Nonlinear Estimation*, ed. S. M. Goldfeld and R. E. Quandt. Cambridge, Mass.: Ballinger.

Heckman, J. J. 1977. Dummy Endogenous Variable in a Simultaneous Equation System. *Econometrica*, in Press.

Hendry, D. F. 1976. The Structure of Simultaneous Equation Estimators. *Journal of Econometrics*. 4; 51–88.

Kiefer, N. M. 1978. On the Value of Sample Separation Information. University of Chicago. *Econometrica*, in press.

Lee, L. F. 1976. Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables. Discussion paper 76–77–04, University of Florida. *International Economic Review*, in press.

Lee, L. F. 1977. Identification and Estimation in Binary Choice Models with Limited Dependent Variables. Discussion paper. 77–85. Center for Economic Research, University of Minnesota. *Econometrica*, in press.

Lee, L. F., G. S., Maddala and R. P. Trost. 1977. Instrumental Variable Estimation of Simultaneous Equation System with Selectivity. *Econometrica*, in press.

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

Maddala, G. S., and L. Lee. 1976. Recursive Models with Qualitative Endogenous Variables. *Annals of Economic and Social Measurement*. 5: 525–545.

Nelson, F. D. 1977. Censored Regression Models with Unobserved, Stochastic, Censored Thresholds. *Journal of Econometrics*. 6; 309–327.

Nelson, F. D., and L. Olson. 1977. Specification and Estimation of a Simultaneous-Equation Model with Limited Dependent Variables. Social Science Working Paper 149. California Institute of Technology. *International Economic Review*; in press.

Sargan, J. D. 1958. The Estimation of Economic Relationships Using Instrumental Variables. *Econometrica*. 26; 393–415.

Theil, H. 1971. *Principles of Econometrics*. New York: Wiley.