Volume Title: Annals of Economic and Social Measurement,
Volume 5, number 4

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm76-4

Publication Date: October 1976

Chapter Title: Recursive Models with Qualitative Endogenous Variables

Chapter Author: G. S. Maddala, Lung-Fei Lee

Chapter URL: http://www.nber.org/chapters/c10494

Chapter pages in book: (p. 525 - 545)

Annals of Economic and Social Measurement, 5/4, 1976

RECURSIVE MODELS WITH QUALITATIVE ENDOGENOUS VARIABLES†

BY G. S. MADDALA AND LUNG-FEI LEE

The paper discusses the estimation procedures and identification problems for some simultaneous equations models involving underlying continuous unobservable variables for which the observed variables are qualitative. It also discusses the formulation of recursive models in the logit framework with an illustration of a five equation model.

1. INTRODUCTION

Models with qualitative endogenous variables have received a lot of attention by econometricians in recent years. Broadly speaking the models fall in two categories: those that start with a multivariate logistic distribution (see Goodman [2], Nerlove and Press [6]) and those that postulate certain underlying continuous response functions. In the latter class of models if y^* is the underlying continuous variable, we observe a qualitative variable y which (assuming it is binary) takes the value 1 if $y^* > 0$ and 0 if $y^* \leq 0$. When it comes to generalizations to many variables, models with underlying continuous variables are computationally more cumbersome than models considered by Nerlove and Press [6].[a] It is fruitful to investigate these models because the underlying causal structure is easier to understand, at least for econometricians used to thinking about recursive and non-recursive models and different types of simultaneous structures. Further, the extensions to models with discrete and continuous cases become more logical and easy to comprehend. In section 2 we present a set of simultaneous equation models involving underlying continuous unobservable variables for which the observed variables are qualitative. We consider the estimation procedures and the identification problems in these models. Some models are more convenient to present in a two equations framework (which is also useful to fix ideas on the nature of the problems involved) and hence we consider them in a two-equation framework. In section 3 we discuss the formulation of recursive models in the logit framework. The logit model has been discussed by Nerlove and Press [6] in the more general simultaneous framework where all endogenous variables are mutually interrelated. However, there will be many problems where one needs to postulate some special type of causality (in particular a recursive model). In section 4 we consider a logit model with such a causal structure. It is a five equation model analyzed earlier by Brown *et al.* [1] but we take into account the fact that some of the endogenous variables are qualitative. The final section presents the conclusions.

† Financial support from the National Science Foundation is gratefully acknowledged. We would like to thank Forrest Nelson for helpful comments on an earlier draft.

[a] Such continuous models have been considered by Heckman [3, 4].

2. Some Models With Underlying Continuous Variables:

In this section we will present three different models and discuss the problems of their logical consistency, identification and estimation. Models 1 and 2 are recursive models and model 3 is a particular type of simultaneous model. For ease of exposition we will discuss the first two models in a two-equation framework but model 3 is discussed in a general framework. This should not be interpreted to mean that models 1 and 2 are special cases of model 3. These three types of models are logically consistent models to analyze problems involving underlying continuous variables. It will be argued later that some other alternative formulations lead to logical inconsistencies.

Model 1 – A Simple Recursive Model with Qualitative Variables

Consider the two equations model:

$$y_1^* = X\beta_1 - \varepsilon_1$$
$$y_2^* = X\beta_2 + \gamma y_1 - \varepsilon_2$$

where ε_1, ε_2 have zero mean, unit variances and are serially independent, X is a vector of exogenous variables.[1] In general, at least one exogenous variable in equation 1 does not appear in equation 2 to guarantee the identification of β_2 and γ. If ε_1 and ε_2 are independent, the exclusion of one exogenous variable in X_2 is not necessary. Also in this model, y_1^*, y_2^* are not observable. Only the dichotomous variables y_1 and y_2 are observable. We assume that there exist constants μ_1 and μ_2 such that

$$y_1 = 1 \quad \text{iff } X\beta_1 - \varepsilon_1 \geq \mu_1 \quad \text{i.e. iff } X\beta_1 - \mu_1 \geq \varepsilon_1,$$
$$y_1 = 0 \quad \text{iff } X\beta_1 - \mu_1 < \varepsilon_1$$

and

$$y_2 = 1 \quad \text{iff } X\beta_2 + \gamma y_1 - \mu_2 \geq \varepsilon_2$$
$$y_2 = 0 \quad \text{iff } X\beta_2 + \gamma y_1 - \mu_2 < \varepsilon_2.$$

Denote the joint distribution function of $(\varepsilon_1, \varepsilon_2)$ by F. The probability function of (y_1, y_2) can easily be written down.

$$P_{11} = P(y_1 = 1, y_2 = 1) = F(X\beta_1 - \mu_1, X\beta_2 + \gamma - \mu_2)$$
$$P_{10} = P(y_1 = 1, y_2 = 0) = F(X\beta_1 - \mu_1, -X\beta_2 - \gamma + \mu_2)$$
$$P_{01} = P(y_1 = 0, y_2 = 1) = F(-X\beta_1 + \mu_1, X\beta_2 - \mu_2)$$
$$P_{00} = P(y_1 = 0, y_2 = 0) = F(-X\beta_1 + \mu_1, -X\beta_2 + \mu_2).$$

We get this simplified expression by assuming that ε_1, ε_2 are symmetrically distributed. (This assumption is used to simplify the notations only).

[1] ε_1 and ε_2 need not have unit variances but since y_1^* and y_2^* are not observable, these variances are not identified and β_i are identified only up to a proportionality factor $\sigma_i - (i = 1, 2)$.

The likelihood function to be maximized is

$$L(\beta_1, \beta_2, \gamma, \mu_1, \mu_2 | X, y_1, y_2) = \Pi P_{11}^{y_1 y_2} P_{10}^{y_1(1-y_2)} P_{01}^{(1-y_1)y_2} P_{00}^{(1-y_1)(1-y_2)}.$$

As with the identification problem in the ordinary logit or probit analysis, if X has a constant term, the coefficients of the constant terms are not identifiable since μ_1 and μ_2 are unknown constants.

For this model, consistent initial estimates for all the parameters are not easy to get. Except for the parameters β_1 which can be estimated consistently by applying the probit (if ε_1 is assumed to be standard normal) or logit analysis (if ε_1 is assumed to have the logistic distribution), the initial consistent estimates of the other parameters are not available. So what we can suggest is to use the consistent estimate $\hat{\beta}_1$ derived by the probit or logit analysis as an initial estimate for β_1 and try various values for the other parameters, study the values that they converge to and choose the one which maximizes the likelihood function. However, if the likelihood function involves numerical double integrals for some specified distributions for the error terms, the maximization procedure is expected to be difficult.

If ε_1 and ε_2 are independent, then the likelihood function reduces to

$$L = \prod_{y_1} [F_1(X\beta_1 - \mu_1)]^{y_1} [1 - F_1(X\beta_1 - \mu_1)]^{1-y_1}$$

$$\times \prod_{y_2} [F_2(X\beta_2 + \gamma y_1 - \mu_2)]^{y_2} [1 - F_2(X\beta_2 + \gamma y_1 - \mu_2)]^{1-y_2}$$

and maximizing L is equivalent to maximizing the likelihood functions for the first and second equations separately (as in a truly recursive model). In this case there will be no computational difficulty for the maximum likelihood procedure.

The extension of the two equations model to models with more equations is straightforward. The likelihood function can be written down theoretically but if it involves numerical multi-integrals, the computation will be intractable.

Model 2—A Recursive Model with Qualitative and Continuous Variables

Consider the model:

$$y_1^* = X\beta_1 - \varepsilon_1$$

$$y_2 = X\beta_2 + \gamma y_1 + \varepsilon_2$$

where ε_1, ε_2 are assumed to have zero mean and are serially independent, X are exogeneous variables, y_1 is an observed dichotomous variable, y_2 is an observed continuous variable and y_1^* is an underlying continuous variable. In fact,

$$y_1 = 1 \quad \text{iff } y_1^* > 0 \quad \text{or iff } X\beta_1 \geq \varepsilon_1$$

$$y_1 = 0 \quad \text{iff } X\beta_1 < \varepsilon_1.$$

Here we assume also that at least one exogeneous variable appears in equation 1 but not in equation 2 to guarantee the identification of the parameters β_2 and γ. If ε_1 and ε_2 are independent, the condition is unnecessary.

The joint density function of y_1, y_2 in this case is

$$g(y_1=1, y_2)=\int_{-\infty}^{X\beta_1} f(\varepsilon_1, y_2-X\beta_2-\gamma)\, d\varepsilon_1$$

and

$$g(y_1=0, y_2)=\int_{X\beta_1}^{\infty} f(\varepsilon_1, y_2-X\beta_2)\, d\varepsilon_1$$

where $f(\varepsilon_1, \varepsilon_2)$ is the joint density function of $(\varepsilon_1, \varepsilon_2)$. The likelihood function to be maximized is

$$L(\beta_1, \beta_2, \gamma|X, y_1, y_2)=\prod_{y_1 y_2} g(y_1=1, y_2)^{y_1} g(y_1=0, y_2)^{1-y_1}.$$

If ε_1 and ε_2 are independent, the likelihood function reduces to

$$L=\prod_{y_1} [F_1(X\beta_1)]^{y_1}[1-F_1(X\beta_1)]^{1-y_1} \prod_{y_2} f_2(y_2-X\beta_2-\gamma y_1)$$

and thus maximizing L is equivalent to maximizing the likelihood functions for both equations separately. In the case that ε_1 and ε_2 are normally distributed, the maximum likelihood procedure is equivalent to estimating the first equation by probit analysis and the second equation by ordinary least squares.

As for the maximum likelihood procedure for the case when ε_1 and ε_2 are not independent, we have to get some good initial estimates to start the iteration. For this model, we can get the initial consistent estimates easily if $(\varepsilon_1, \varepsilon_2)$ are assumed to be normally distributed, i.e.,

$$(\varepsilon_1, \varepsilon_2) \sim N\left(0, \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right).$$

Since the first equation is a standard probit model, β_1 can be estimated consistently by probit analysis. Rewrite the second equation as

$$\begin{aligned} y_2 &= X\beta_2+\gamma F_1(X\beta_1)+\varepsilon_2+\gamma(y_1-F_1(X\beta_1)) \\ &= X\beta_2+\gamma F_1(X\beta_1)+\omega \end{aligned}$$

where $\omega=\gamma(y_1-F_1(X\beta_1))+\varepsilon_2$. Since $E(\omega)=0$ and ω is uncorrelated with the regressors, we can estimate β_2 and γ by regressing y_2 on X and $F_1(X\hat{\beta}_1)$. Since $\hat{\beta}_1$ is a consistent estimate of β_1, under some general conditions, it can be shown that the estimates $\hat{\beta}_2$ and $\hat{\gamma}$ of β and γ are consistent estimators. Denote the estimated residual of the second equation by $\tilde{\varepsilon}_2$ i.e.,

$$\tilde{\varepsilon}_2=y_2-X\hat{\beta}_2-\hat{\gamma}y_1.$$

Then the variance σ_2^2 can be estimated consistently by $\hat{\sigma}_2^2$ where

$$\hat{\sigma}_2^2=\frac{1}{T}\sum_{t=1}^{T} \tilde{\varepsilon}_2^2, \qquad (T \text{ is the sample size}).$$

Finally it remains to find some consistent estimates for σ_{12}. Rewrite the two equations into a switching regression model.

$$y_2 = X\beta_2 + \gamma + \varepsilon_2 \qquad \text{iff } X\beta_1 > \varepsilon_1$$
$$y_2 = X\beta_2 + \varepsilon_2 \qquad \text{iff } X\beta_1 \leq \varepsilon_1.$$

With these specifications, it can easily be shown that

$$E(y_2 - X\beta_2 - \gamma | y_1 = 1) = E(y_1\varepsilon_2)/F_1(X\beta_1)$$
$$= \sigma_{12}\left[-\frac{1}{\sqrt{2\pi}}e^{-(X\beta_1)^2/2}\right]\Big/F_1(X\beta_1)$$

and

$$E(y_2 - X\beta_2 | y_1 = 0) = E((1-y_1)\varepsilon_2)/[1 - F_1(X\beta_1)]$$
$$= \sigma_{12}\left[-\frac{1}{\sqrt{2\pi}}e^{-(X\beta_1)^2/2}\right]\Big/[1 - F_1(X\beta_1)].$$

Thus σ_{12} can be estimated consistently either by using the sub-sample corresponding to $y_1 = 1$ and regressing

$$y_2 - X\hat{\beta}_2 - \hat{\gamma} \text{ on } \left[-\frac{1}{\sqrt{2\pi}}e^{-(X\hat{\beta}_1)^2/2}\right]\Big/F_1(X\hat{\beta}_1)$$

or by using the subsample corresponding to $y_1 = 0$ and regressing

$$y_2 - X\hat{\beta}_2 \text{ on } \left[-\frac{1}{\sqrt{2\pi}}e^{-(X\hat{\beta}_1)^2/2}\right]\Big/\left[1 - F_1(X\hat{\beta}_1)\right].$$

or by combining these two sub-samples. Thus, we can use these consistent estimates as the initial estimates to start the iteration for the maximum likelihood procedure.

In the above model, the observed dependent variable is dichotomous in the first equation and the observed dependent variable is continuous in the second equation. In the reverse case we have the model:

$$y_1 = X\beta_1 + \varepsilon_1$$
$$y_2^* = X\beta_2 + \gamma y_1 - \varepsilon_2$$

where ε_1, ε_2 are serially independent with zero means and variances σ_{11}, σ_{22} and covariance σ_{12}. Here now y_1 is the observed continuous variable and y_2^* is unobserved but the dichotomous variable y_2 is observed.

$$y_2 = 1 \qquad \text{iff } y_2^* \geq 0 \qquad \text{or } X\beta_2 + \gamma y_1 \geq \varepsilon_2$$
$$y_2 = 0 \qquad \text{iff } X\beta_2 + \gamma y_1 < \varepsilon_2.$$

Under the rank condition that at least one of the exogeneous variables appears in the first equation but not the second one, we can show (the proof can be found in

Model 3) that only

$$\beta_1, \sigma_{11}, \frac{\beta_2}{\sigma}, \frac{\gamma}{\sigma}, \frac{\sigma_{22}}{\sigma^2}, \frac{\sigma_{12}}{\sigma}$$

are identifiable[2] where $\sigma^2 = \text{var}\,(\gamma\varepsilon_1 - \varepsilon_2)$.

In this model, the joint densities are

$$g(y_1, y_2 = 1) = \int_{-\infty}^{X\beta_2 + \gamma y_1} f(y_1 - X\beta_1, \varepsilon_2)\, d\varepsilon_2$$

and

$$g(y_1, y_2 = 0) = \int_{X\beta_2 + \gamma y_1}^{\infty} f(y_1 - X\beta_1, \varepsilon_2)\, d\varepsilon_2$$

The likelihood function to be maximized is

$$L(\beta_1, \beta_2, \gamma | y, X) = \prod_{y_1, y_2} [g(y_1, y_2 = 1)]^{y_1} [g(y_1, y_2 = 0)]^{1-y_1}.$$

Again if the residuals are independent, maximizing L amounts to estimation of each equation separately.

If the residuals ε_1 and ε_2 are normally distributed, the consistent initial estimates can be found as follows. The first equation is a standard regression model, so β_1 and σ_{11} can be estimated consistently by the ordinary least squares estimators $\hat{\beta}_1$ and $\hat{\sigma}_{11}$. Rewrite the second equation into a probit model,

$$\frac{y_2^*}{\sigma} = X\frac{\beta_2}{\sigma} + \frac{\gamma}{\sigma}(X\hat{\beta}_1) - \frac{\omega}{\sigma}$$

where $\omega = \gamma X(\hat{\beta}_1 - \beta_1) + (\varepsilon_2 - \gamma\varepsilon_1)$. It is easily shown that ω/σ is asymptotically a standard normal variable, so β_2/σ, γ/σ can be estimated consistently by the probit analysis. As for the parameters σ_{12}/σ, σ_{22}/σ^2, we can use the relation

$$E(\varepsilon_1 y_2) = \text{cov}\left(\varepsilon_1, \frac{\varepsilon_2 - \gamma\varepsilon_1}{\sigma}\right)\left[-\frac{1}{\sqrt{2\pi}} e^{-(X(\beta_2/\sigma) + (\gamma/\sigma)X\beta_1)^2/2}\right]$$

$$= \left[\left(\frac{\gamma}{\sigma}\right)\sigma_{11} - \frac{\sigma_{12}}{\sigma}\right]\left[\frac{1}{\sqrt{2\pi}} e^{-(X(\beta_2/\sigma) + (\gamma/\sigma)X\beta_1)^2/2}\right]$$

or equivalently

$$E(\varepsilon_1 | y_2 = 1) = \left[\frac{\gamma}{\sigma}\sigma_{11} - \frac{\sigma_{12}}{\sigma}\right]\left[\frac{\frac{1}{\sqrt{2\pi}} e^{-(X(\beta_2/\sigma) + (\gamma/\sigma)X\beta_1)^2/2}}{F_1\left(X\frac{\beta_2}{\sigma} + \frac{\gamma}{\sigma}X\beta_1\right)}\right]$$

to estimate σ_{12}/σ. Regress the product of the least squares residuals and y_2 on $1/\sqrt{2\pi}\, e^{-(X(\hat{\beta}_2/\sigma) + (\hat{\gamma}/\sigma)X\hat{\beta}_1)^2/2}$ and use this least square estimate and $(\hat{\gamma}/\sigma)\hat{\sigma}_{11}$ to

[2] Though the likelihood function involves 5 parameters β_1, σ_{11}, β_{12}/σ, γ/σ and σ_{12}/σ and it appears as though only these parameters are estimable, it should be noted that $\sigma^2 = \text{var}\,(\gamma\varepsilon_1 - \varepsilon_2) = \gamma^2\sigma_{11} - 2\sigma_{12} + \sigma_{22}$ or $(\gamma^2/\sigma^2)\sigma_{11} - 2(\gamma/\sigma)(\sigma_{12}/\sigma) + \sigma_{22}/\sigma^2 = 1$ and hence σ_{22}/σ^2 is also estimable.

solve for $\hat{\sigma}_{12}/\sigma$. Finally since

$$\sigma^2 = E(\gamma\varepsilon_1 - \varepsilon_2)^2$$
$$= \gamma^2\sigma_{11} - 2\gamma\sigma_{12} + \sigma_{22},$$

it implies $(\sigma_{22}/\sigma^2) = 1-(\gamma/\sigma)^2\sigma_{11}+2(\gamma/\sigma)(\sigma_{12}/\sigma)$. Hence we can estimate σ_{22}/σ^2 by $1-(\hat{\gamma}/\sigma)^2\hat{\sigma}_{11}+2(\hat{\gamma}/\sigma)(\hat{\sigma}_{12}/\sigma)$. Thus this gives the initial consistent estimates for all the identifiable parameters and they can be used to start the iteration of the maximum likelihood procedure.

Model 3—Simultaneous Model with Unobservable Continuous Variables:

This qualitative model with simultaneous continuous and unobservable endogeneous variables has the following specification,

$$B\tilde{y}_t + \Gamma X_t = \varepsilon_t,$$

where ε_t is serially independent, has zero mean and covariance matrix Σ, B is a $G \times G$ non-singular matrix with unitary diagonal elements. Here

$$\tilde{y}_t = (y^*_{1t}, y^*_{2t}, \ldots, y^*_{G_1t}, y_{G_1+1t}, \ldots, y_{Gt})$$

is a vector and $y_{G_1+t}, \ldots, y_{Gt}$ are observable continuous endogeneous variables, $y^*_{1t}, \ldots, y^*_{G_1t}$ are unobservable variables but the dichotomous variables $y_{1t}, \ldots, y_{G_1t}$ are observed such that

$$y_{it} = 1 \leftrightarrow y^*_{it} \geq 0$$
$$= 0 \leftrightarrow y^*_{it} < 0.$$

So this model is a simultaneous model with continuous and qualitative variables when $0 < G_1 < G$ and it is a simultaneous model with only qualitative variables when $G_1 = G$.

This model is quite similar to the usual simultaneous structural equations model. As in the probit model, the model has its identification problems. In this section, we will consider which parameters can be identifiable under the usual conditions for the inclusion and exclusion of the variables in the simultaneous system. Other prior information can of course give the identification of the unknown parameters.

Consider the reduced form for this system which is

$$\tilde{y}_t = -B^{-1}\Gamma X_t + B^{-1}\varepsilon_t$$
$$= \Pi X_t + v_t$$

where

$$v_t = B^{-1}\varepsilon_t \text{ and } \Pi = -B^{-1}\Gamma.$$

It follows that the covariance matrix Ω of v_t is

$$\Omega = B^{-1}\Pi B'^{-1}$$

Denote

$$\Lambda = \left[\begin{array}{c|c} D & 0 \\ \hline 0 & I \end{array}\right] = \begin{bmatrix} \frac{1}{\sigma_1} \cdot & & & & \\ & \cdot \frac{1}{\sigma_{G1}} & & & 0 \\ & & \cdot & & \\ & & & 1 \cdot & \\ & & & & \cdot 1 \end{bmatrix}$$

a $G \times G$ diagonal matrix where $\sigma_i^2 = \text{var}(v_{ii})$, $i = 1, \ldots, G_1$.

For the parameters of the reduced form of the system, it can be shown easily that $\Lambda\Pi$, $\Lambda\Omega\Lambda$ are identifiable but not Γ and Ω without any further assumptions.

Now let us consider the identifiability of the parameters of the structural equations by the equations

$$B\Pi + \Gamma = 0$$

$$B\Omega B' = \Sigma.$$

To simplify the notation, we will show the identification of the parameters for the whole system. For the identification of the parameters in any single equation, it follows immediately. First let us consider the parameters B and Γ.

$$B\Pi + \Gamma = 0 \rightarrow (B\Lambda^{-1})(\Lambda\Pi) + \Gamma = 0.$$

Since Λ is a diagonal matrix, the usual rank conditions for Π are applicable for $\Lambda\Pi$. However the normalization rule $\beta_{ii} = 1$ for the first G_1 structural equation has no effect in the identification of $B\Lambda^{-1}$ and Γ. To see this, write the matrix B in a partitioned form.

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where B_{11} is a $G_1 \times G_1$ matrix, B_{22} is a $(G - G_1) \times (G - G_1)$ matrix, B_{12} is a $G_1 \times (G - G_1)$ matrix and B_{21} is a $(G - G_1) \times G_1$ matrix. Thus

$$B\Lambda^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} D^{-1} & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} B_{11}D^{-1} & B_{12} \\ B_{21}D^{-1} & B_{22} \end{bmatrix}.$$

It is easy to see now that the first $G_1 \times G_1$ elements in the diagonal elements of $B\Lambda^{-1}$ are not unitary elements any more but rather the unknown parameters $1/\sigma_1, \ldots, 1/\sigma_{G_1}$. Hence each row of $[B_{11}D^{-1}, B_{12}]$ is identifiable only up to a proportion. However, if we insist that the coefficient of y_{ii}^* in the ith structural equation must be unity, we can normalize them by dividing the corresponding row of $[B_{11}D^{-1}, B_{12}]$ by $1/\sigma_i$. Thus we have

$$(\Lambda B\Lambda^{-1})(\Lambda\Pi) + \Lambda\Gamma = 0$$

where $\Lambda B\Lambda^{-1}$ has unitary diagonal elements. Hence $\Lambda B\Lambda^{-1}$ and $\Lambda\Gamma$ are identifiable if the rank condition holds for each structural equation in the system. Also

$$
\begin{aligned}
&\Sigma = B\Omega B' \\
&\rightarrow \Sigma = (B\Lambda^{-1})(\Lambda\Omega\Lambda)(\Lambda^{-1}B') \\
&\rightarrow \Lambda\Sigma\Lambda = (\Lambda B\Lambda^{-1})(\Lambda\Omega\Lambda)(\Lambda B\Lambda^{-1})'.
\end{aligned}
$$

Thus, under the rank conditions $\Lambda\Sigma\Lambda$ is identifiable. By the same arguments, if the rank condition holds only for some structural equations, it follows that the corresponding parameters in $\Lambda B\Lambda^{-1}$, $\Lambda\Gamma$ and $\Lambda\Sigma\Lambda$ will be identifiable.

The identification of the structural parameters can also be improved upon if more information is available in the system. Instead of a constant threshold for the unobservable endogeneous variables, if some extraneous variable thresholds are available the identification of the parameters in the corresponding structural equation will be improved. Without loss of generality assume that there exist some extraneous variables z_{it} for the first $G_2 (G_2 \leq G_1)$ equations such that

$$
\begin{aligned}
&y_{it} = 1 \Leftrightarrow y^*_{it} \geq z_{it} \\
&y_{it} = 0, \text{ otherwise}, i = 1, \ldots, G_2; t = 1, \ldots, T
\end{aligned}
$$

where z_{it} $(i = 1, \ldots, G_2)$ are uncorrelated with errors ε_t. In this case, if the rank conditions hold for all structural equations, we have

$$
\Lambda B\Lambda^{-1}, \Lambda\Gamma \qquad \text{and } \Lambda\Sigma\Lambda
$$

are identifiable where now

$$
\Lambda = \begin{bmatrix}
1 & & & & & & & 0 \\
& \ddots & & & & & & \\
& & 1 & & & & & \\
& & & \frac{1}{\sigma_{G_2+1}} & & & & \\
& & & & \ddots & & & \\
& & & & & \frac{1}{\sigma_{G_1}} & & \\
& & & & & & 1 & \\
0 & & & & & & & \ddots\ 1
\end{bmatrix}
$$

Finally, if the extraneous variables z_{it} are available for all $i = 1, \ldots, G_1$, Λ is an identity and hence B, Γ and Σ are all identifiable.

Heckman [3] has recently proposed to use the full information ML estimation for this kind of system. Also he has suggested some initial estimates for the parameters when the disturbance terms are assumed to be normally distributed. However, if the system has many structural equations and $G_1 > 2$, there will be G_1 multi-integrals involved in the density function and the estimation procedure will be intractable. A feasible alternative to the FIML method is to estimate the unrestricted reduced form equations separately by Probit analysis and use a two stage least square analogue to estimate the structural equations. The test for the significance of these parameters can also be developed.

Rewrite the system with all the coefficients to be identifiable. The system is:

$$(\Lambda B\Lambda^{-1})y_t^{**}+\Lambda\Gamma x_t=\Lambda\varepsilon_t$$

where $y_t^{**}=\Lambda\tilde{y}_t$. With these y_t^{**} as the unobservable continuous endogeneous variables, it characterizes y_t in the same way as $\tilde{y}_t$ does, i.e.,

$$\begin{aligned} y_{it}&=1 \Leftrightarrow y_{it}^{**}>0\\ &=0, \text{ otherwise,}\end{aligned}$$

for all $i=1,\ldots,G_1$. The reduced form of the system is

$$y_t^{**}=\Lambda\Pi x_t+\Lambda v_t.$$

The first G_1 equations in this reduced form system are the usual Probit models and the last $G-G_1$ are the ordinary regression equations. Thus $\Lambda\Pi$ can be estimated consistently by $\widehat{\Lambda\Pi}$ which are derived by the Probit analysis and the least squares procedures. As for the estimation of the parameters $\Lambda B\Lambda^{-1}$ and $\Lambda\Gamma$ it is sufficient to illustrate the procedure by the first and the G_1+1th equation.

Written down explicitly, the first equation has the following expression

$$\begin{aligned} y_{1t}^{**}&+\frac{\beta_{12}\sigma_2}{\sigma_1}y_{2t}^{**}+\ldots+\frac{\beta_{1G_1}\sigma_{G_1}}{\sigma_1}y_{G_1t}^{**}+\frac{\beta_{1G_1+1}}{\sigma_1}y_{G_1+1t}^{**}\\ &+\ldots+\frac{\beta_{1G}}{\sigma_1}y_{Gt}^{**}+\frac{\gamma_{11}}{\sigma_1}x_{1t}+\frac{\gamma_{12}}{\sigma_1}x_{2t}+\ldots+\frac{\gamma_{1k}}{\sigma_1}x_{kt}-\frac{\varepsilon_{1t}}{\sigma_1}.\end{aligned}$$

Denote $\hat{y}_t^{**}=\hat{\Lambda}\Pi x_t$ and substitute for y_t^{**} into the structural equation, it becomes

$$\begin{aligned} y_{1t}^{**}=&-\frac{\beta_{12}\sigma_2}{\sigma_1}\hat{y}_{2t}^{**}-\ldots\frac{\beta_{1G_1}\sigma_{G_1}}{\sigma_1}\hat{y}_{G_1t}^{**}-\frac{\beta_{1G_1+1}}{\sigma_1}\hat{y}_{G_1+1t}^{**}\\ &-\frac{\beta_{1G}}{\sigma_1}\hat{y}_{Gt}^{**}-\frac{\gamma_{11}}{\sigma_1}x_{1t}-\frac{\gamma_{12}}{\sigma_1}x_{2t}\ldots\frac{\gamma_{1k}}{\sigma_1}x_{kt}+w_{1t}\end{aligned}$$

where w_{1t} can be shown to have the same distribution as v_{1t}/σ_1 is asymptotically and hence asymptotically standard normal. Thus the maximum likelihood procedure for the Probit model can be applied again to this equation. Thus we can estimate the structural parameters

$$\frac{\beta_{12}\sigma_2}{\sigma_1},\ldots,\frac{\beta_{1G_1}\sigma_{G_1}}{\sigma_1},\frac{\beta_{1G_1+1}}{\sigma_1},\ldots,\frac{\beta_{1G}}{\sigma_1},\frac{\gamma_{11}}{\sigma_1}\frac{\gamma_{1K}}{\sigma_1}.$$

consistently. It follows that the asymptotic t test can also be developed for the test of the significance of these parameters.

The G_1+1th equation is

$$\begin{aligned} y_{G_1+1,t}=&-\beta_{G_1+1,1}\sigma_1y_{1t}^{**}-\ldots\beta_{G_1+1,G_1}\sigma_{G_1}y_{G_1t}^{**}-\beta_{G_1+1,G_1+2}y_{G_1+2,t}\\ &-\ldots-\beta_{G_1+1,G}y_{Gt}-\gamma_{G_1+1,1}x_{1t}-\ldots-\gamma_{G_1+1,k}x_{kt}+\varepsilon_{G_1+1t}.\end{aligned}$$

Substitute $\hat{y}_{it}^{**}$ for y_{it}^{**} $(i=1,\ldots,G_1)$ in the equation and apply the ordinary least squares procedure. The parameters

$$\beta_{G_1+1,1}\sigma_1,\ldots,\beta_{G_1+1,G_1}\sigma_{G_1}\beta_{G_1+1,G_1+2},\ldots,\beta_{G_1+1,G}\quad\text{and } \gamma_{G_1+1,1},\ldots,\gamma_{G_1+1k}$$

can be estimated consistently and the usual t test for the significance of the parameters can also be applied.

Models of the type 1, 2 and 3 considered here are well-defined. But in the class of qualitative simultaneous equations models, some models are not valid. For example, the model

$$y_1^* = x\beta_1 + \alpha_1 y_2 + \varepsilon_1$$

$$y_2 = x\beta_2 + \alpha_2 y_1 + \varepsilon_2$$

is not valid.

It leads to logical inconsistencies[3] because it results in an equation of the form

$$y^* = x\gamma + \delta y + u$$

where the unobservable variable y^* is related to the dichotomous variable y through another relation of the form

$$\begin{aligned} y &= 1 \quad \text{if } y^* > 0 \\ &= 0 \quad \text{if } y^* < 0. \end{aligned}$$

Other models of the form

$$y_1^* = x\beta_1 + \alpha_1 y_2 + \varepsilon_1$$

$$y_2^* = x\beta_2 + \alpha_1 y_1^* + \varepsilon_2$$

and

$$y_1^* = x\beta_1 + \alpha_1 y_2 - \varepsilon_1$$

$$y_2^* = x\beta_2 + \alpha_2 y_1 - \varepsilon_2$$

are also inconsistent. To show the inconsistency of the last model, it is easy to check in general that

$$\sum_{y_1, y_2} P(y_1, y_2) \neq 1$$

whenever $\alpha_1 \neq 0$ and $\alpha_2 \neq 0$.

All these inconsistent models have a common feature that the reduced forms are not defined. Thus the endogenous variables can not be explained by the exogeneous variables and the disturbances.

Hence we can conclude that all the simultaneous equations models with qualitative endogeneous variables can be broadly divided into the category of the recursive type of models as model 1, model 2, or their combination, and the category of the model 3.

3. Simultaneous vs. Recursive Models in the Logit Framework[4]

Nerlove and Press [6] discuss a logit model where the endogenous variables are all completely interrelated; for instance, if there are three such variables y_1, y_2, y_3 then y_1 influences y_2 and y_3, y_2 influences y_3 and y_1, and y_3 influences y_1 and y_2.

[3] The inconsistencies of this model have been recently discussed by Heckman [3].

[4] This section is based on the discussion in Maddala and Nelson [5].

This type of mutual independence may not always be desirable and we should be able to analyze models that have any causal structure we desire.

For illustrative purposes we will consider the case of three dichotomous variables y_1, y_2, y_3, and a set of exogenous variables to be denoted by x.

$$\text{Let } P_{ijk} = \Pr(Y_1 = i, Y_2 = j, Y_3 = k) \qquad i, j, k = 0 \text{ or } 1.$$

We can then write

$$
\begin{aligned}
P_{000} &= 1/D \\
P_{100} &= e^{\beta'_1 x}/D \\
P_{010} &= e^{\beta'_2 x}/D \\
P_{001} &= e^{\beta'_3 x}/D \qquad (1)\\
P_{110} &= e^{\beta'_4 x}/D \\
P_{101} &= e^{\beta'_5 x}/D \\
P_{011} &= e^{\beta'_6 x}/D \\
P_{111} &= e^{\beta'_7 x}/D
\end{aligned}
$$

where

$$D = 1 + \sum_{i=1}^{7} e^{\beta'_i x}$$

These equations imply the following relations:

$$
\begin{array}{c|c|c}
\dfrac{P_{100}}{P_{000}} = e^{\beta'_1 x} & \dfrac{P_{010}}{P_{00}} = e^{\beta'_2 x} & \dfrac{P_{001}}{P_{000}} = e^{\beta'_3 x} \\
\dfrac{P_{110}}{P_{010}} = e^{(\beta_4-\beta_2)'x} & \dfrac{P_{110}}{P_{100}} = e^{(\beta_4-\beta_1)'x} & \dfrac{P_{101}}{P_{100}} = e^{(\beta_5-\beta_1)'x} \\
\dfrac{P_{101}}{P_{001}} = e^{(\beta_5-\beta_3)'x} & \dfrac{P_{011}}{P_{001}} = e^{(\beta_6-\beta_3)'x} & \dfrac{P_{011}}{P_{010}} = e^{(\beta_6-\beta_2)'x} \\
\dfrac{P_{111}}{P_{011}} = e^{(\beta_7-\beta_6)'x} & \dfrac{P_{111}}{P_{101}} = e^{(\beta_7-\beta_5)'x} & \dfrac{P_{111}}{P_{110}} = e^{(\beta_7-\beta_4)'x}
\end{array}
$$

These reactions can be written as

$$
\begin{aligned}
&\text{Log}\frac{P(y_1=1|y_2y_3)}{P(y_1=0|y_2y_3)} = \beta'_1x + (\beta_4-\beta_2-\beta_1)'xy_2 + (\beta_5-\beta_3-\beta_1)'xy_3 \\
&\qquad + (\beta_7-\beta_6-\beta_5-\beta_4+\beta_3+\beta_2+\beta_1)'xy_2y_3 \\
(2)\quad &\text{Log}\frac{P(y_2=1|y_1y_3)}{P(y_2=0|y_1y_3)} = \beta'_2x + (\beta_4-\beta_2-\beta_1)'xy_1 + (\beta_6-\beta_3-\beta_2)'xy_3 \\
&\qquad + (\beta_7-\beta_6-\beta_5-\beta_4+\beta_3+\beta_2+\beta_1)'xy_1y_3 \\
&\text{Log}\frac{P(y_3=1|y_1y_3)}{P(y_3=0|y_1y_2)} = \beta'_3x + (\beta_5-\beta_3-\beta_1)'xy_1 + (\beta_6-\beta_3-\beta_2)'xy_2 \\
&\qquad + (\beta_7-\beta_6-\beta_5-\beta_4+\beta_3+\beta_2+\beta_1)'xy_1y_2.
\end{aligned}
$$

Note the symmetry in the coefficients of the equations (2). This symmetry was discussed by Nerlove and Press [6]. To simplify the model we can impose:

$$(3) \qquad \begin{aligned} &(\beta_4-\beta_2-\beta_1)'x=\beta_{12}\\ &(\beta_5-\beta_3-\beta_1)'x=\beta_{13}\\ &(\beta_6-\beta_3-\beta_2)'x=\beta_{23}\\ &(\beta_7-\beta_6-\beta_5-\beta_4-\beta_3-\beta_2+\beta_1)'x=\gamma. \end{aligned}$$

We can get this model if the first element of x is 1, all but the first elements of the vector β_4 are equal to the sum of the corresponding elements of β_2 and β_1, with similar conditions holding for β_5 and β_6, and for β_7 all but the first element are equal to the sum of the corresponding elements of β_1, β_2 and β_3.

Thus, an important consequence of the multinomial logistic model (1) is that we get the well defined conditional distributions (2). In actual practice, if there are a number of categories, the complete multinomial model (1) involves too many parameters. That is why Nerlove and Press suggest estimating equations (2) by the logit method treating the right hand variables as exogenous. One can get consistent estimators for the parameters by this procedure (though these are not fully efficient because they ignore the cross equation constraints). This procedure reduces the number of parameters to be estimated considerably. Further reduction can be achieved by making some simplifying assumptions like (3). If we further impose the restriction $\beta_7-\beta_6-\beta_5-\beta_4+\beta_3+\beta_2+\beta_1=0$ we can also eliminate the product terms involving y_1y_2, y_2y_3, y_3y_1 in equations (2).

Unlike the usual simultaneous equations model where it is not possible to interpret each equation as a conditional expectation (except in a recursive system) the specification (1) permits well defined conditional probabilities (2). Also, it looks as if we cannot have causal chains in simultaneous equation logit models. This is indeed not so. Consider a situation where the causal relations between $y_1y_2y_3$ are as shown in Figures 1 and 2.

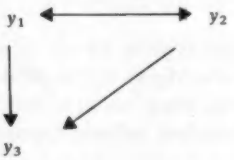

Figure 1

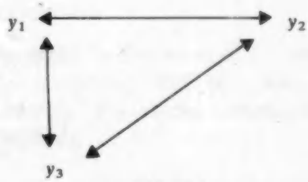

Figure 2

Suppose that y_1 and y_2 are variables that do precede (in time or in some other sense) variable y_3. Then a relationship as in Figure 2 obviously does not make sense and it is a relationship as in Figure 1 that we should be considering. It might be thought that the symmetry conditions in equations (2) imply that if y_3 depends on y_1, then the reverse must be true with the *same* effect. This is of course not true. What the symmetry conditions imply is that if y_1 depends on y_3 and y_3 depends on y_1 then the two effects should be equal. We have to interpret the conditional probability equations (2) as depicting the nature of the causal relationships between the variables. For the model in Figure 1 these causal relationships can be written in the following form

$$\text{Log}\frac{\Pr(y_1=1|y_2,x)}{\Pr(y_1=0|y_2,x)}=\delta y_2+\alpha_1'x$$

(4)
$$\text{Log}\frac{\Pr(y_2=1|y_1,x)}{\Pr(y_2=0|y_1,x)}=\delta y_1+\alpha_2'x$$

$$\text{Log}\frac{\Pr(y_3=1|y_1,y_2x)}{\Pr(y_3=0|y_1,y_2x)}=\beta_1 y_1+\beta_2 y_2+\alpha_3'x.$$

Note that the symmetry conditions have been imposed only for the first two equations in (4) since y_1 and y_2 are jointly determined. One can estimate δ, α_1, α_2 from the joint probability distribution of y_1 and y_2. These joint probabilities are:

$$P_{11}=e^{(\alpha_1+\alpha_2)'x+\delta}/\Delta$$
$$P_{01}=e^{\alpha_2'x}/\Delta$$
$$P_{10}=e^{\alpha_1'x}/\Delta$$
$$P_{00}=1/\Delta$$

where

(5)
$$\Delta=1+e^{\alpha_1'x}+e^{\alpha_2'x}+e^{(\alpha_1+\alpha_2)'x+\delta}.$$

As for the third equation in (4) its parameters are estimated separately. This equation implies

(6)
$$\text{Log}\frac{P_{111}}{P_{110}}=\beta_1+\beta_2+\alpha_3'x$$

$$\text{Log}\frac{P_{011}}{P_{010}}=\beta_2+\alpha_3'x$$

$$\text{Log}\frac{P_{101}}{P_{100}}=\beta_1+\alpha_3'x$$

$$\text{Log}\frac{P_{001}}{P_{000}}=\alpha_3'x$$

and equations (6) in conjunction with (5) will enable us to estimate the joint probabilities P_{ijk} for any goodness of fit tests. If we assume the causal relationship in Figure 2, the conditional probabilities will be given by equations (2), with any appropriate zero restrictions, and the joint probabilities will be given by (1), again with the appropriate zero restrictions.

Given any specification of the conditional odds ratios as in (2) one can deduce the joint probabilities (1). The ML estimation procedure based on the implied joint probabilities (1), has been called the full information ML procedure by Nerlove and Press [6]. They argue that it is computationally less cumbersome to estimate the conditional equations (2) and that in practice these should be adequate.

In the case of a recursive model, of course, as in the usual simultaneous equations context, the estimates from the conditional equations (2) would be fully efficient. As an illustration consider the causal model:

$$y_1 = f(x)$$
$$y_2 = f(x, y_1)$$

where y_1 and y_2 are binary.

$$\Pr(y_1 = 1) = \frac{e^{\beta_1' x}}{1 + e^{\beta_1' x}} \tag{7}$$

$$\Pr(y_2 = 1 | y_1) = \frac{e^{\beta_2' x + \gamma y_1}}{1 + e^{\beta_2' x + \gamma y_1}}.$$

These give the joint probabilities

$$\begin{aligned} P_{11} &= F(\beta_1' x) F(\beta_2' x + \gamma) \\ P_{01} &= F(\beta_2' x)[1 - F(\beta_1' x)] \\ P_{10} &= F(\beta_1' x)[1 - F(\beta_2' x + \gamma)] \\ P_{00} &= [1 - F(\beta_1' x)][1 - F(\beta_2' x)] \end{aligned} \tag{8}$$

where

$$F(z) = \frac{e^z}{1 + e^z}$$

The separate estimation of equations (7) and the joint estimation of equations (8) are the same.

4. An Application

The model we analyze here is a model analyzed by Brown *et al.* [1] on the effectiveness of the neighborhood youth corps programs (NYC program). We estimate here a model somewhat simpler than theirs.[5] The model consists of five endogeneous variables and ten exogeneous variables.

Endogeneous Variables

- y_1 Heard of the NYC, a dummy variable, 1—yes, 0—no.
- y_2 Dummy variable for participation in NYC program, 1—participated, 0—not participated.
- y_3 Dropout from high school a dummy variable, 1—dropout, 0—not dropout.
- y_4 Proportion of time involuntary unemployed in post-high school period.
- y_5 Current (or most recent) wage level of the individual in cents/hour.

[5] We are grateful to Stanley Horowitz for supplying us the data.

Exogeneous Variables

x_1 Constant term, $x_1 = 1$.

x_2 Western, Southern U.S. or else dummy variable 1—western or southern, 0—else.

x_3 Rural area, small city or medium city, big city dummy variable 1—rural area or small city, 0—medium or big city.

x_4 Family size while in high school.

x_5 Family income during high school.

x_6 Father's education.

x_7 Age of individual.

x_8 Sex of individual, a dummy variable, 1—male, 0—female.

x_9 Race of individual, a dummy variable, 1—white, 0—nonwhite.

x_{10} Number of friends of individual who dropped out of high school.

The NYC program is expected to influence the lives of its participants. It might be expected to affect their decisions about finishing high school, participating in the labor force, wage level and so on. In addition to the NYC, other factors may influence these activities and also their enrollment in NYC. We build a five equation recursive model to study the NYC participation and assess the effects of the NYC program on the individual's activities. The exogeneous variables x_2, x_3 differentiate the regions and communities in which the individual may live. Variables x_4, x_5, x_6 quantify factors of the home environment experienced by the individual while he was in high school. x_7, x_8, x_9 measure the individual characteristics that are expected to be important determinants of the person's activities and opportunities. The last variable captures the group status that might influence his activities. The structure of the model is given in Table 1. Table 2 presents the OLS estimates and Table 3 presents the 2SLS estimates.

TABLE 1

THE STRUCTURE OF THE MODEL

	y_1	y_2	y_3	y_4	y_5	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
y_1						✓		✓	✓	✓		✓		✓	
y_2	✓					✓			✓		✓	✓		✓	
y_3		✓				✓	✓	✓			✓	✓	✓		✓
y_4		✓	✓			✓		✓			✓	✓		✓	
y_5		✓	✓	✓		✓	✓	✓			✓		✓	✓	

i.e.,

$$y_1 = \alpha_{10} + \alpha_{11}x_3 + \alpha_{12}x_4 + \alpha_{13}x_5 + \alpha_{14}x_7 + \alpha_{15}x_9 + \varepsilon_1$$

$$y_2 = \beta_{21}y_1 + \alpha_{20} + \alpha_{21}x_4 + \alpha_{22}x_6 + \alpha_{23}x_7 + \alpha_{24}x_9 + \varepsilon_2$$

$$y_3 = \beta_{31}y_2 + \alpha_{30} + \alpha_{31}x_2 + \alpha_{32}x_3 + \alpha_{33}x_6 + \alpha_{34}x_7 + \alpha_{35}x_8 + \alpha_{36}x_{10} + \varepsilon_3$$

$$y_4 = \beta_{41}y_2 + \beta_{42}y_3 + \alpha_{40} + \alpha_{41}x_3 + \alpha_{42}x_6 + \alpha_{43}x_7 + \alpha_{44}x_9 + \varepsilon_4$$

$$y_5 = \beta_{51}y_2 + \beta_{52}y_3 + \beta_{53}y_4 + \alpha_{50} + \alpha_{51}x_2 + \alpha_{52}x_3 + \alpha_{53}x_6 + \alpha_{54}x_8 + \alpha_{55}x_9 + \varepsilon_5$$

TABLE 2

THE OLS ESTIMATES AND THEIR *t* STATISTICS

	y_1	y_2	y_3	y_4	1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
y_1					1.847	0.009	−0.043	0.011	−0.002	0.003	−0.05	0.001	−0.055	
					(7.27	(0.28)	(−1.02)	(1.77)	(−1.56)	(0.73)	(−4.09)	(0.04)	(−1.90)	
y_2	0.679				−0.676	0.024	−0.013	0.008	−0.0004	−0.004	0.033	−0.003	0.036	
	(15.83)				(−2.23)	(0.69)	(−0.27)	(1.16)	(−0.28)	(−1.12)	(2.27)	(−0.09)	(1.08)	
y_3		−0.028			1.187	−0.110	0.071	0.003	0.0002	−0.009	−0.045	0.063	−0.002	0.006
		(−0.94)			(4.33)	(−3.36)	(1.58)	(0.50)	(0.2)	(−2.47)	(−3.37)	(2.08)	(−0.06)	(1.7)
y_4		0.006	0.038		0.292	0.0001	0.016	0.0002	0.0005	0.003	−0.014	−0.009	−0.026	
		(0.43)	(2.03(		'2.12)	(0.01)	(0.73)	(0.07)	(0.78)	(1.7)	(−2.05)	(−0.62)	(−1.67)	
y_5		−3.289	−11.961	−19.672	164.251	−16.226	−21.676	0.330	0.051	1.037	1.450	43.010	−14.672	
		(−0.68)	(−1.88)	(−1.19)	(3.66)	(−3.06)	(−2.95)	(0.31)	(0.25)	(1.76)	(0.67)	(8.78)	(−2.9)	

TABLE 3

2SLS Estimates and their t Statistics

	y_1	y_2	y_3	y_4	1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
y_1					1.952 (7.42)		−0.037 (−0.82)	0.010 (1.57)	−0.002 (−1.75)		−0.054 (−4.11)		−0.066 (−2.21)	
y_2	0.757 (1.45)				−0.798 (−0.74)			0.007 (0.93)		−0.006 (−1.34)	0.036 (1.08)		0.041 (0.78)	
y_3		−0.141 (−0.29)			1.210 (2.81)	−0.102 (−2.69)	0.059 (1.17)			−0.010 (−2.16)	−0.042 (−3.02)	0.071 (2.17)		0.009 (1.76)
y_4		−0.058 (−0.33)	−0.022 (−0.21)		0.417 (1.92)		0.013 (0.55)			0.002 (1.02)	−0.016 (−2.07)		−0.026 (−1.72)	
y_5		86.528 (1.06)	−127.59 (−1.79)	−8.175 −(0.04)	163.95 (3.44)	−33.371 (−3.1)	−6.40 (−0.63)			0.360 (0.3)		54.891 (5.86)	−16.14 (−2.13)	

TABLE 4

Logit Estimates and their Chi-Square Test Statistics

	y_1	y_2	1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
y_1			9.068		−0.234 (0.61)	0.074 (2.48)	−0.015 (3.53)		−0.352 (16.18)		−0.494 (5.15)	
y_2	5.717 (31.98)		−9.459			0.041 (1.39)		−0.031 (1.72)	0.224 (5.63)		0.180 (0.92)	
y_3		−0.155 (0.53)	5.489	−0.726 (11.43)	0.402 (1.90)			−0.097 (9.02)	−0.297 (9.19)	0.491 (5.69)		0.041 (2.61)

TABLE 5

LOGIT 2SLS ESTIMATES AND THEIR TEST STATISTICS*

	y_1	y_2	y_3	y_4	1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
y_1														
y_2	3.926				−7.044			0.027		−0.017	0.197		0.218	
	(2.31)							(0.72)		(0.74)	(1.5)		(0.87)	
y_3		−4.312			8.544	−0.739	0.328			−0.130	−0.317	0.481		0.041
		(4.77)				(11.74)	(1.25)			(11.39)	(10.45)	(5.43)		(2.65)
y_4		−0.090	−0.016		0.434		0.013			0.002	−0.016		−0.026	
		(−0.48)	(−0.17)		(1.94)		(0.56)			(0.95)	(−2.10)		(−1.75)	
y_5		60.053	−36.834	−275.81	168.39	−21.847	−10.81			1.837		44.495	−22.157	
		(0.82)	(−0.7)	(−1.36)	(3.29)	(−2.64)	(−1.18)			(1.14)		(5.66)	(−2.83)	

* The test statistics for equations 1, 2, 3 are chi-square test statistics; the test statistics for equations 4, 5 are t-test statistics.

As is evident, even for the recursive models considered in section 2, the ML estimation involves bivariate integrals unless the residual's are independent. Extension to more variables involves higher order integrals. We could have used the methods outlined in section 3 which are straightforward adaptations of the Nerlove–Press procedure. However we chose to estimate our model by the following computationally simpler procedures. First we estimated the model by using the logit method separately on each equation treating all the right hand variables as exogeneous (which is valid if the residuals are independent). Next we used a 2SLS analogue which we call here logit 2SLS. In this method the endogenous dummy variables are replaced by their estimated values obtained by the application of the logit method to the reduced form. These estimates are presented in Tables 4 and 5.

If the NYC program is effective we would expect β_{31} and β_{41} to be negative and β_{51} to be positive. Also β_{42} is expected to be positive and β_{52} and β_{53} are expected to be negative. The OLS estimates reported in Table 2 have some wrong signs (β_{41} and β_{51}). The 2SLS estimates reported in Table 3 have the correct signs for the coefficient of y_2 but none of the coefficients are significant and β_{42} has the wrong sign (though the coefficient is not significant). The single equation logit estimates reported in Table 4 still indicate that the NYC program is not effective. The logit 2SLS estimates reported in Table 5 indicate a stronger effect of the NYC program—particularly on the dropout rate out of high school, though it has no additional effect on the post high school rate of involuntary unemployment and the wage rate earned. It appears to influence these variables only through its influence on the dropout rate.

5. Conclusions

The paper presents some models where some of the endogenous variables are unobserved continuous variables for which the observed variables are discrete, and discusses the identification and estimation problems in these models. The paper also discusses the formulation of simultaneous and recursive models in the logit framework. An empirical example concerning the effectiveness of the neighborhood youth corps program is presented. The model consists of five endogenous variables, and has a particular causal structure that resembles a recursive model in the simultaneous equations literature (or more precisely the matrix of coefficients of the endogenous variables is triangular). The 2SLS method where the discrete nature of the endogenous variables is taken into account leads to the conclusion that the neighborhood youth corps program has a significant effect on the rate of dropping out of high school, whereas the ordinary 2SLS method, where the discrete nature of the endogenous variables is not taken into account, showed no significant effect of the program.

University of Florida
University of Minnesota

References

[1] George F. Brown, Jr., Arlene Holen, Stanley A. Horowitz, Louis Jacobson, and Lester P. Silverman—"Analysis of the Neighborhood Youth Corps Program," Center for Naval Analysis: Dec. 1972.

[2] Leo Goodman, "The Multivariate Analysis of Qualitative Date: Interaction Among Multiple Classifications," *Journal of the American Statistical Association*, 1970, pp. 226–256.

[3] James J. Heckman, "Simultaneous Equations Models With Both Continuous and Discrete Endogenous Variables With and Without Structural Shift in the Equations," University of Chicago, Aug. 1975.

[4] ———, "Dummy Endogenous Variables," NBER Working Paper, April 1973. Forthcoming (subject to revision) in *Econometrica*.

[5] G. S. Maddala and Forrest D. Nelson, "Analysis of Qualitative Variables," Working Paper No. 70, Oct. 1974, NBER (Cambridge, Mass.).

[6] Marc Nerlove and James Press, "Univariate and Multivariate Log-linear and Logistic Models," Rand Corporation, R-1306-EDA/NIH, Dec. 1973.