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 a continuous response function, 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]. 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.

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

\( ^\ast \) 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 each 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 \( \varepsilon_1, \varepsilon_2 \) have zero mean, unit variances and are serially independent, \( X \) is a vector of exogenous variables. In general, at least one exogenous variable in equation 1 does not appear in equation 2 to guarantee the identification of \( \beta_2 \) and \( \gamma \). If \( \varepsilon_1 \) and \( \varepsilon_2 \) 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 \( \mu_1 \) and \( \mu_2 \) such that

\[ y_1 = 1 \quad \text{iff} \quad X\beta_1 - \varepsilon_1 \geq \mu_1 \]
\[ y_1 = 0 \quad \text{iff} \quad X\beta_1 - \varepsilon_1 < \mu_1 \]

and

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

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 \( \varepsilon_1, \varepsilon_2 \) are symmetrically distributed. (This assumption is used to simplify the notations only).

1 \( \varepsilon_1 \) and \( \varepsilon_2 \) need not have unit variances but since \( y_1^* \) and \( y_2^* \) are not observable, these variances are not identified and \( \beta_2 \) are identified only up to a proportionality factor \( \sigma_i - (i = 1, 2) \).

526
The likelihood function to be maximized is
\[
L(\beta_1, \beta_2, \gamma, \mu_1, \mu_2 | X, y_1, y_2) = \prod_{i=1}^n \{ [F_1(X\beta_1 - \mu_1)]^{y_1}[1 - F_1(X\beta_1 - \mu_1)]^{1-y_1}\} \times \prod_{i=1}^n \{ [F_2(X\beta_2 + \gamma y_1 - \mu_2)]^{y_2}[1 - F_2(X\beta_2 + \gamma y_1 - \mu_2)]^{1-y_2}\}
\]
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 \( \mu_1 \) and \( \mu_2 \) are unknown constants.

For this model, consistent initial estimates for all the parameters are not easy to get. Except for the parameters \( \beta_1 \) which can be estimated consistently by applying the probit (if \( e_1 \) is assumed to be standard normal) or logit analysis (if \( e_1 \) 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 \( \beta_1 \) 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 \( e_1 \) and \( e_2 \) are independent, then the likelihood function reduces to
\[
L = \prod_{i=1}^n \{ [F_1(X\beta_1 - \mu_1)]^{y_1}[1 - F_1(X\beta_1 - \mu_1)]^{1-y_1}\} \prod_{i=1}^n \{ [F_2(X\beta_2 + \gamma y_1 - \mu_2)]^{y_2}[1 - F_2(X\beta_2 + \gamma y_1 - \mu_2)]^{1-y_2}\}
\]
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 it involves numerical multi-integrals, the computation will be intractable.

Model 2—A Recursive Model with Qualitative and Continuous Variables

Consider the model:
\[
y_1^* = X\beta_1 - e_1 \\
y_2 = X\beta_2 + \gamma y_1 + e_2
\]
where \( e_1, e_2 \) 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,
\[
y_1 = 1 \quad \text{iff } y_1^* > 0 \quad \text{or iff } X\beta_1 \geq e_1 \\
y_1 = 0 \quad \text{iff } X\beta_1 < e_1.
\]
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 \( \beta_2 \) and \( \gamma \). If \( e_1 \) and \( e_2 \) are independent, the condition is unnecessary.
The joint density function of $y_1$, $y_2$ in this case is
\[ g(y_1 = 1, y_2) = \int_{-\infty}^{\infty} f(\varepsilon_1, y_2 - X\beta_2 - \gamma) \, d\varepsilon_1 \]
and
\[ g(y_1 = 0, y_2) = \int_{-\infty}^{\infty} f(\varepsilon_1, y_2 - X\beta_2) \, d\varepsilon_1 \]
where $f(\varepsilon_1, \varepsilon_2)$ is the joint density function of $(\varepsilon_1, \varepsilon_2)$. The likelihood function to be maximized is
\[ L(\beta_1, \beta_2, \gamma | X, y_1, y_2) = \prod_{Y_1} g(y_1 = 1, y_2) g(y_1 = 0, y_2) \]
If $\varepsilon_1$ and $\varepsilon_2$ are independent, the likelihood function reduces to
\[ L = \prod_{y_1} \left[ F_1(X\beta_1) \right]^{y_1} \left[ 1 - F_1(X\beta_1) \right]^{1-y_1} \prod_{y_2} f(y_2 - X\beta_2 - \gamma y_2) \]
and thus maximizing $L$ is equivalent to maximizing the likelihood functions for both equations separately. In the case that $\varepsilon_1$ and $\varepsilon_2$ 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 $\varepsilon_1$ and $\varepsilon_2$ 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.,
\[ (\varepsilon_1, \varepsilon_2) \sim N(0, \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2 \end{bmatrix}) \]
Since the first equation is a standard probit model, $\beta_1$ can be estimated consistently by probit analysis. Rewrite the second equation as
\[ y_2 = X\beta_2 + \gamma F_1(X\beta_1) + \varepsilon_2 + \gamma (y_1 - F_1(X\beta_1)) = X\beta_2 + \gamma F_1(X\beta_1) + \omega \]
where $\omega = \gamma (y_1 - F_1(X\beta_1)) + \varepsilon_2$. Since $E(\omega) = 0$ and $\omega$ is uncorrelated with the regressors, we can estimate $\beta_2$ and $\gamma$ by regressing $y_2$ on $X$ and $F_1(X\beta_1)$. Since $\hat{\beta}_2$ is a consistent estimate of $\beta_2$, under some general conditions, it can be shown that the estimates $\hat{\beta}_2$ and $\hat{\gamma}$ of $\beta_2$ and $\gamma$ are consistent estimators. Denote the estimated residual of the second equation by $\hat{\varepsilon}_2$ i.e.,
\[ \hat{\varepsilon}_2 = y_2 - X\hat{\beta}_2 - \hat{\gamma} y_1. \]
Then the variance $\sigma^2_2$ can be estimated consistently by
\[ \sigma^2_2 = \frac{1}{T-1} \sum_{t=1}^T \hat{\varepsilon}_2^2, \quad (T \text{ is the sample size}). \]
Finally it remains to find some consistent estimates for \( \sigma_{12} \). Rewrite the two equations into a switching regression model.

\[
y_2 = X_2 \beta_2 + \gamma + \varepsilon_2 \quad \text{iff} \quad X \beta_1 > \varepsilon_1
\]

\[
y_2 = X_2 \beta_2 + \varepsilon_2 \quad \text{iff} \quad X \beta_1 \leq \varepsilon_1
\]

With these specifications, it can easily be shown that

\[
E(y_2 - X \beta_2 - y | y_1 = 1) = E(y_1 \varepsilon_2) / F_1(X \beta_1)
\]

\[
= \sigma_{12} \left[ \frac{1}{\sqrt{2\pi}} e^{-(X \beta_1)^2/2} \right] / F_1(X \beta_1)
\]

and

\[
E(y_2 - X \beta_2 | y_1 = 0) = E((1 - y_1) \varepsilon_2) / [1 - F_1(X \beta_1)]
\]

\[
= \sigma_{12} \left[ \frac{1}{\sqrt{2\pi}} e^{-(X \beta_1)^2/2} \right] / [1 - F_1(X \beta_1)]
\]

Thus \( \sigma_{12} \) can be estimated consistently either by using the sub-sample corresponding to \( y_1 = 1 \) and regressing

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

or by using the subsample corresponding to \( y_1 = 0 \) and regressing

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

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 \( \varepsilon_1, \varepsilon_2 \) are serially independent with zero means and variances \( \sigma_{11}, \sigma_{22} \) and covariance \( \sigma_{12} \). Here now \( y_1 \) is the observed continuous variable and \( y_2^* \) is unobserved but the dichotomous variable \( y_2 \) is observed.

\[
y_2 = 1 \quad \text{iff} \quad y^*_2 \geq 0 \quad \text{or} \quad X \beta_2 + \gamma y_1 \geq \varepsilon_2
\]

\[
y_2 = 0 \quad \text{iff} \quad X \beta_2 + \gamma y_1 < \varepsilon_2
\]

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 
\[
\beta_1, \sigma_{11}, \beta_2, \gamma, \sigma_{22}, \sigma_{12}
\]
are identifiable where \( \sigma^2 = \text{var}(\gamma \epsilon_1 - \epsilon_2) \).
In this model, the joint densities are
\[
g(y_1, y_2 = 1) = \int_{-\infty}^{X_{y_1} + y_{y_1}} f(y_1 - XB_1, \epsilon_2) \, d\epsilon_2
\]
and
\[
g(y_1, y_2 = 0) = \int_{-\infty}^{X_{y_2} + y_{y_2}} f(y_1 - XB_1, \epsilon_2) \, d\epsilon_2
\]
The likelihood function to be maximized is
\[
L(\beta_1, \beta_2, \gamma | y, X) = \prod_{y_1, y_2} [g(y_1, y_2 = 1)]^y [g(y_1, y_2 = 0)]^{1-y}.
\]
Again if the residuals are independent, maximizing \( L \) amounts to estimation of each equation separately.

If the residuals \( \epsilon_1 \) and \( \epsilon_2 \) are normally distributed, the consistent initial estimates can be found as follows. The first equation is a standard regression model, so \( \beta_1 \) and \( \sigma_{11} \) 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,
\[
\frac{y_2}{\sigma} = \frac{X\beta_2 + \gamma (X\beta_1)}{\sigma} - \frac{\omega}{\sigma}
\]
where \( \omega = \gamma X(\beta_1 - \beta_0) + (\epsilon_2 - \gamma \epsilon_1) \). It is easily shown that \( \omega/\sigma \) is asymptotically a standard normal variable, so \( \beta_2/\sigma, \gamma/\sigma \) can be estimated consistently by the probit analysis.

As for the parameters \( \sigma_{12}/\sigma, \sigma_{22}/\sigma^2 \), we can use the relation
\[
E(\epsilon_1, y_2) = \text{cov}(\epsilon_1, \frac{\epsilon_2 - \gamma \epsilon_1}{\sigma}) = \left[ \frac{\gamma}{\sigma} \sigma_{11} - \frac{\sigma_{12}}{\sigma} \right] \left[ \frac{e^{-\frac{1}{2}(X\beta_1 + \gamma \epsilon_{\gamma}X\beta_1)^2}}{2\pi} \right]
\]
or equivalently
\[
E(\epsilon_1 | y_2 = 1) = \left[ \frac{\gamma}{\sigma} \sigma_{11} - \frac{\sigma_{12}}{\sigma} \right] \left[ \frac{e^{-\frac{1}{2}(X\beta_2 + \gamma \epsilon_{\gamma}X\beta_1)^2}}{2\pi} \right]
\]
to estimate \( \sigma_{12}/\sigma \). Regress the product of the least squares residuals and \( y_2 \) on
\[
1/\sqrt{2\pi} e^{-\frac{1}{2}(X\beta_2 + \gamma \epsilon_{\gamma}X\beta_1)^2}
\]
and use this least square estimate and \((\gamma/\sigma) \hat{\sigma}_{11}\) to

\[530]
solve for $\sigma_{12}/\sigma$. Finally since

$$\sigma^2 = E(y_{e1} - e_2)^2$$

$$= \gamma^2 \sigma_{11} - 2y\sigma_{12} + \sigma_{22},$$

it implies $(\sigma_{12}/\sigma^2) = 1 - (\gamma/\sigma)^2 \sigma_{11} + 2(\gamma/\sigma)(\sigma_{12}/\sigma)$. Hence we can estimate $\sigma_{12}/\sigma^2$ by $1 - (\gamma/\sigma)^2 \sigma_{11} + 2(\gamma/\sigma)(\sigma_{12}/\sigma)$. 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,

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

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

$$\tilde{y}_t = (y_{1t}^+, y_2^+, \ldots, y_{G1t}^+, y_{G1t+1}, \ldots, y_{Glt})$$

is a vector and $y_{G1t+1}, \ldots, y_{G1t}$ are observable continuous endogeneous variables, $y_{1t}^+, \ldots, y_{G1t}^+$ are unobservable variables but the dichotomous variables $y_{1t}, \ldots, y_{G1t}$ are observed such that

$$y_a = 1 \leftrightarrow y_a^+ \geq 0$$

$$= 0 \leftrightarrow y_a^+ < 0.$$

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

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

$$\tilde{y}_t = -B^{-1}\Gamma X_t + B^{-1}\epsilon_t$$

$$= \Pi X_t + \nu_t,$$

where

$$\nu_t = B^{-1}\epsilon_t,$$

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

It follows that the covariance matrix $\Omega$ of $\nu_t$ is

$$\Omega = B^{-1}\Pi B^{-1}.$$
For the parameters of the reduced form of the system, it can be shown easily that \( A \Pi, AD \) are identifiable but not \( \Pi \) and \( D \) 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 \( \Gamma \).

\[
B \Pi + \Gamma = 0 \rightarrow (BA^{-1})(A\Pi) + \Gamma = 0.
\]

Since \( A \) is a diagonal matrix, the usual rank conditions for \( \Pi \) are applicable for \( A \Pi \). However the normalization rule \( B_{0i} = 1 \) for the first \( G_1 \) structural equation has no effect in the identification of \( BA^{-1} \) and \( \Gamma \). 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, \( B_{12} \) 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

\[
BA^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} D^{-1} & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} B_{11}D^{-1} & B_{12} \\ B_{21}D^{-1} & B_{22} \end{bmatrix}
\]

It is easy to see now that the first \( G_1 \times G_1 \) elements in the diagonal elements of \( BA^{-1} \) 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 \( y^*_i \) in the \( i \)th 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

\[
(ABA^{-1})(A\Pi) + \Lambda \Gamma = 0
\]

where \( ABA^{-1} \) has unitary diagonal elements. Hence \( ABA^{-1} \) and \( A\Gamma \) are identifi-
able if the rank condition holds for each structural equation in the system. Also
\[ \Sigma = B\Omega B' \]
\[ \rightarrow \Sigma = (BA^{-1})(A\Omega A)(A^{-1}B') \]
\[ \rightarrow \Lambda \Sigma \Lambda = (ABA^{-1})(A\Omega A)(ABA^{-1})'. \]

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 \( ABA^{-1}, \Lambda \Gamma \) and \( \Lambda \Sigma A \) 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 \( z_i \) for the first \( G_2 (G_2 < G_1) \) equations such that
\[
y_{it} = 1 \Leftrightarrow y^*_t = z_i, \\
y_{it} = 0, \text{otherwise, } i = 1, \ldots, G_2; t = 1, \ldots, T
\]
where \( z_i (i = 1, \ldots, G_2) \) are uncorrelated with errors \( e_t \). In this case, if the rank conditions hold for all structural equations, we have
\[ \Lambda B \Gamma^{-1}, \Lambda \Gamma \text{ and } \Lambda \Sigma A \]
are identifiable where now \[ \Lambda = \begin{bmatrix}
1 & & & \\
1 & \frac{1}{G_2 + 1} & & \\
& & \ddots & \ddots & \ddots & \\
& & & \frac{1}{G_1} & & \\
0 & & & & \ddots & \\
& & & & & 1
\end{bmatrix} \]

Finally, if the extraneous variables \( z_i \) are available for all \( i = 1, \ldots, G_1, \Lambda \) is an identity and hence \( B, \Gamma \) and \( \Sigma \) 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:

\((AB\Lambda^{-1})\gamma_{t}^{**} + \Lambda\Gamma x_{t} = \Lambda\eta_{t}\)

where \(\gamma_{t}^{**} = \Lambda\tilde{y}_{t}\). With these \(\gamma_{t}^{**}\) as the unobservable continuous endogenous variables, it characterizes \(y_{t}\) in the same way as \(\tilde{y}_{t}\) does, i.e.,

\[ y_{it} = 1 \Leftrightarrow \gamma_{it}^{**} > 0 \]

\[ = 0, \text{ otherwise,} \]

for all \(i = 1, \ldots, G_{1}\). The reduced form of the system is

\[ y_{t}^{**} = \Lambda\Pi x_{t} + \Lambda\nu_{t}. \]

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 \(\tilde{\Lambda}\Pi\) which are derived by the Probit analysis and the least squares procedures. As for the estimation of the parameters \(\Lambda\beta\Lambda^{-1}\) and \(\Lambda\Gamma\) it is sufficient to illustrate the procedure by the first and the \(G_{1} + 1\)th equation.

Written down explicitly, the first equation has the following expression

\[
\frac{\beta_{12} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{12} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \ldots + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**}
\]

Denote \(\gamma_{t}^{**} = \tilde{\Lambda}\Pi x_{t}\) and substitute for \(\gamma_{t}^{**}\) into the structural equation, it becomes

\[
y_{t}^{**} = \frac{\beta_{12} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} - \ldots - \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**} + \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}} \gamma_{t}^{**}
\]

where \(w_{it}\) can be shown to have the same distribution as \(\nu_{it}/\sigma_{1}\) 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

\[
\frac{\beta_{12} \sigma_{2}}{\sigma_{1}}, \ldots, \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}}, \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}}, \ldots, \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}}, \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}}, \ldots, \frac{\beta_{1G_{1}2} \sigma_{2}}{\sigma_{1}}
\]

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

The \(G_{1} + 1\)th equation is

\[
y_{G_{1}+1} = -\beta_{G_{1}+1,1} \sigma_{1} \gamma_{G_{1}}^{**} - \ldots - \beta_{G_{1}+1,G_{1}} \sigma_{G_{1}} \gamma_{G_{1}}^{**} - \beta_{G_{1}+1,G_{1}+2} y_{G_{1}+2} - \ldots - \beta_{G_{1}+1,G_{1}+2} y_{G_{1}+2} - \ldots - \beta_{G_{1}+1,G_{1}+2} y_{G_{1}+2} - \ldots - \gamma_{G_{1}+1,k \gamma_{G_{1}+1,k} + \epsilon_{G_{1}+1}}.
\]

Substitute \(\gamma_{G_{1}}^{**}\) for \(\gamma_{G_{1}}^{**}\) \((i = 1, \ldots, G_{1})\) in the equation and apply the ordinary least squares procedure. The parameters

\[
\beta_{G_{1}+1,1} \sigma_{1}, \ldots, \beta_{G_{1}+1,G_{1}} \sigma_{1}, \beta_{G_{1}+1,G_{1}+2}, \ldots, \beta_{G_{1}+1,G_{1}+2}, \ldots, \beta_{G_{1}+1,G_{1}+2}, \ldots, \beta_{G_{1}+1,G_{1}+2} \quad \text{and} \quad \gamma_{G_{1}+1,1}, \ldots, \gamma_{G_{1}+1,1}, \ldots, \gamma_{G_{1}+1,1}
\]
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
\[
\begin{align*}
y_1^* &= x_{11} + \alpha_1 y_2 + \epsilon_1 \\
y_2^* &= x_{22} + \alpha_2 y_1 + \epsilon_2 
\end{align*}
\]
is not valid.

It leads to logical inconsistencies because it results in an equation of the form
\[
y^* = xy + \delta y + u
\]
where the unobservable variable $y^*$ is related to the dichotomous variable $y$
through another relation of the form
\[
y = 1 \quad \text{if } y^* > 0 \\
= 0 \quad \text{if } y^* < 0.
\]

Other models of the form
\[
\begin{align*}
y_1^* &= x_{11} + \alpha_1 y_1 + \epsilon_1 \\
y_2^* &= x_{22} + \alpha_2 y_2 + \epsilon_2
\end{align*}
\]
and
\[
\begin{align*}
y_1^* &= x_{11} + \alpha_1 y_2 - \epsilon_1 \\
y_2^* &= x_{22} + \alpha_2 y_2 - \epsilon_2
\end{align*}
\]
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

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 \( x \).

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

We can then write

\[
P_{000} = \frac{e^{\theta_{000}}}{D} \\
P_{001} = \frac{e^{\theta_{001}}}{D} \\
P_{010} = \frac{e^{\theta_{010}}}{D} \\
P_{100} = \frac{e^{\theta_{100}}}{D} \\
P_{011} = \frac{e^{\theta_{011}}}{D} \\
P_{101} = \frac{e^{\theta_{101}}}{D} \\
P_{110} = \frac{e^{\theta_{110}}}{D} \\
P_{111} = \frac{e^{\theta_{111}}}{D}
\]

where

\[
D = 1 + \sum_{i=1}^{7} e^{\theta_{0i}}
\]

These equations imply the following relations:

\[
\begin{align*}
P_{000} &= e^{\theta_{000}} \\
P_{001} &= e^{\theta_{001}} \\
P_{010} &= e^{\theta_{010}} \\
P_{100} &= e^{\theta_{100}} \\
P_{011} &= e^{\theta_{011}} \\
P_{101} &= e^{\theta_{101}} \\
P_{110} &= e^{\theta_{110}} \\
P_{111} &= e^{\theta_{111}}
\end{align*}
\]

These reactions can be written as

\begin{align*}
\log \frac{P(y_1 = 1 | y_2 y_3)}{P(y_1 = 0 | y_2 y_3)} &= \beta_{11} x + (\beta_{4} - \beta_{2} - \beta_{3})y_2 + (\beta_{5} - \beta_{3} - \beta_{4})y_3 + (\beta_{5} - \beta_{4} - \beta_{3} - \beta_{2})y_2 y_3 \\
\log \frac{P(y_2 = 1 | y_1 y_3)}{P(y_2 = 0 | y_1 y_3)} &= \beta_{11} x + (\beta_{4} - \beta_{2} - \beta_{3})y_1 + (\beta_{5} - \beta_{3} - \beta_{4})y_3 + (\beta_{5} - \beta_{4} - \beta_{3} - \beta_{2})y_2 y_3 \\
\log \frac{P(y_3 = 1 | y_1 y_2)}{P(y_3 = 0 | y_1 y_2)} &= \beta_{11} x + (\beta_{4} - \beta_{2} - \beta_{3})y_1 y_3 + (\beta_{5} - \beta_{3} - \beta_{4} - \beta_{2})y_2 y_3
\end{align*}
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:

\[(3)\]
\[
(\beta_1 - \beta_2 - \beta_3) x = \beta_{12} \\
(\beta_1 - \beta_2 - \beta_3) x = \beta_{11} \\
(\beta_1 - \beta_2 - \beta_3) x = \beta_{13} \\
(\beta_1 - \beta_2 - \beta_3 - \beta_4 - \beta_5 + \beta_6) x = \gamma.
\]

We can get this model if the first element of \(x\) is 1, all but the first elements of the vector \(\beta_4\) are equal to the sum of the corresponding elements of \(\beta_2\) and \(\beta_1\), with similar conditions holding for \(\beta_5\) and \(\beta_3\), and for \(\beta_7\) all but the first element are equal to the sum of the corresponding elements of \(\beta_5\), \(\beta_3\) and \(\beta_2\).

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 \(\beta_1 - \beta_2 - \beta_3 - \beta_4 + \beta_5 + \beta_6 = 0\) 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 \(y_1, y_2, y_3\) are as shown in Figures 1 and 2.

\[
\begin{align*}
\text{Figure 1} & : y_1 \rightarrow y_2 \rightarrow y_3 \\
\text{Figure 2} & : y_1 \rightarrow y_2 \leftarrow y_3
\end{align*}
\]

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 \(y_2\) 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

\[
\begin{align*}
\log \frac{\Pr(y_1 = 1 | y_2, x)}{\Pr(y_1 = 0 | y_2, x)} &= \delta y_2 + \alpha_1 x \\
\log \frac{\Pr(y_2 = 1 | y_1, x)}{\Pr(y_2 = 0 | y_1, x)} &= \delta y_1 + \alpha_2 x \\
\log \frac{\Pr(y_3 = 1 | y_1, y_2, x)}{\Pr(y_3 = 0 | y_1, y_2, x)} &= \beta_1 y_1 + \beta_2 y_2 + \alpha_3 x.
\end{align*}
\]  

(4)

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 \(\delta, \alpha_1, \alpha_2\) from the joint probability distribution of \(y_1\) and \(y_2\). These joint probabilities are:

\[
\begin{align*}
P_{11} &= e^{\alpha_1 + \alpha_2 x + \delta x} / \Delta \\
P_{01} &= e^{\alpha_2 x} / \Delta \\
P_{10} &= e^{\alpha_1 x} / \Delta \\
P_{00} &= 1 / \Delta
\end{align*}
\]

where

\[
\Delta = 1 + e^{\alpha_1 x} + e^{\alpha_2 x} + e^{\alpha_1 + \alpha_2 x + \delta x}.
\]  

(5)

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

\[
\begin{align*}
\log \frac{P_{111}}{P_{110}} &= \beta_1 + \beta_2 + \alpha_3 x \\
\log \frac{P_{011}}{P_{010}} &= \beta_2 + \alpha_3 x \\
\log \frac{P_{101}}{P_{100}} &= \beta_1 + \alpha_3 x \\
\log \frac{P_{001}}{P_{000}} &= \alpha_3 x
\end{align*}
\]

(6)

and equations (6) in conjunction with (5) will enable us to estimate the joint probabilities \(P_{ik}\) 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.

\[
\Pr(y_1 = 1) = \frac{e^{\beta_1 x}}{1 + e^{\beta_1 x}}
\]
\[
\Pr(y_2 = 1|y_1) = \frac{e^{\beta_2 x + \gamma}}{1 + e^{\beta_2 x + \gamma}}
\]

These give the joint probabilities

\[
P_{11} = F(\beta_1 x)F(\beta_2 x + \gamma)
\]
\[
P_{01} = F(\beta_2 x)[1 - F(\beta_1 x)]
\]
\[
P_{10} = F(\beta_1 x)[1 - F(\beta_2 x + \gamma)]
\]
\[
P_{00} = [1 - F(\beta_1 x)][1 - F(\beta_2 x)]
\]

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. 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.

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

<table>
<thead>
<tr>
<th>( y_1 )</th>
<th>( y_2 )</th>
<th>( y_3 )</th>
<th>( y_4 )</th>
<th>( y_5 )</th>
<th>( y_6 )</th>
<th>( y_7 )</th>
<th>( y_8 )</th>
<th>( y_9 )</th>
<th>( y_{10} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
</tr>
</tbody>
</table>

i.e.,

\[
y_1 = \alpha_{10} + \alpha_{11} x_1 + \alpha_{12} x_4 + \alpha_{13} x_5 + \alpha_{14} x_7 + \alpha_{15} x_9 + \epsilon_1 \\
y_2 = \beta_{21} y_1 + \alpha_{20} + \alpha_{21} x_4 + \alpha_{22} x_5 + \alpha_{23} x_7 + \alpha_{24} x_9 + \epsilon_2 \\
y_3 = \beta_{31} y_2 + \alpha_{30} + \alpha_{31} x_2 + \alpha_{32} x_3 + \alpha_{33} x_4 + \alpha_{34} x_7 + \alpha_{35} x_9 + \epsilon_3 \\
y_4 = \beta_{41} y_2 + \beta_{42} y_3 + \alpha_{40} + \alpha_{41} x_4 + \alpha_{42} x_5 + \alpha_{43} x_7 + \alpha_{44} x_8 + \epsilon_4 \\
y_5 = \beta_{51} y_2 + \beta_{52} y_3 + \beta_{53} y_4 + \alpha_{50} + \alpha_{51} x_2 + \alpha_{52} x_3 + \alpha_{53} x_4 + \alpha_{54} x_8 + \epsilon_5
\]
<table>
<thead>
<tr>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
<th>$y_5$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
<th>$x_6$</th>
<th>$x_7$</th>
<th>$x_8$</th>
<th>$x_9$</th>
<th>$x_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.847</td>
<td>0.009</td>
<td>-0.043</td>
<td>0.011</td>
<td>-0.002</td>
<td>0.003</td>
<td>-0.05</td>
<td>0.001</td>
<td>-0.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7.27)</td>
<td>(0.28)</td>
<td>(-1.02)</td>
<td>(1.77)</td>
<td>(-1.56)</td>
<td>(0.73)</td>
<td>(-4.09)</td>
<td>(0.04)</td>
<td>(-1.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_2$</td>
<td>0.679</td>
<td>(15.83)</td>
<td>-0.676</td>
<td>0.024</td>
<td>-0.013</td>
<td>0.008</td>
<td>-0.0004</td>
<td>-0.004</td>
<td>0.033</td>
<td>-0.003</td>
<td>0.035</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-2.23)</td>
<td>(0.69)</td>
<td>(-0.27)</td>
<td>(1.16)</td>
<td>(-0.28)</td>
<td>(-1.12)</td>
<td>(2.27)</td>
<td>(-0.09)</td>
<td>(1.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_3$</td>
<td>-0.028</td>
<td>(-0.94)</td>
<td>1.187</td>
<td>-0.110</td>
<td>0.071</td>
<td>0.003</td>
<td>0.0002</td>
<td>0.009</td>
<td>-0.045</td>
<td>0.063</td>
<td>-0.002</td>
<td>0.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4.33)</td>
<td>(-3.36)</td>
<td>(1.58)</td>
<td>(0.50)</td>
<td>(0.2)</td>
<td>(-2.47)</td>
<td>(-3.37)</td>
<td>(2.08)</td>
<td>(-0.06)</td>
<td>(1.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_4$</td>
<td>0.006</td>
<td>0.038</td>
<td>0.292</td>
<td>0.0001</td>
<td>0.016</td>
<td>0.0002</td>
<td>0.0005</td>
<td>0.000</td>
<td>-0.014</td>
<td>-0.009</td>
<td>-0.026</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.43)</td>
<td>(2.03)</td>
<td>(2.12)</td>
<td>(0.01)</td>
<td>(0.73)</td>
<td>(0.07)</td>
<td>(0.78)</td>
<td>(1.7)</td>
<td>(-2.05)</td>
<td>(-0.62)</td>
<td>(-1.67)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_5$</td>
<td>-3.289</td>
<td>-11.961</td>
<td>-19.672</td>
<td>164.251</td>
<td>-16.226</td>
<td>-21.676</td>
<td>0.330</td>
<td>0.051</td>
<td>1.037</td>
<td>1.450</td>
<td>43.010</td>
<td>-14.672</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.68)</td>
<td>(-1.88)</td>
<td>(-1.19)</td>
<td>(3.66)</td>
<td>(-3.06)</td>
<td>(-2.95)</td>
<td>(0.31)</td>
<td>(0.25)</td>
<td>(1.76)</td>
<td>(0.67)</td>
<td>(8.78)</td>
<td>(-2.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 4

**Logit Estimates and their Chi-Square Test Statistics**

<table>
<thead>
<tr>
<th></th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
<th>$l$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
<th>$x_6$</th>
<th>$x_7$</th>
<th>$x_8$</th>
<th>$x_9$</th>
<th>$x_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>5.489</td>
<td>-0.155</td>
<td>-0.262</td>
<td>-0.402</td>
<td>-0.097</td>
<td>0.297</td>
<td>0.491</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_2$</td>
<td>-0.494</td>
<td>-0.155</td>
<td>5.489</td>
<td>-0.262</td>
<td>-0.402</td>
<td>-0.097</td>
<td>0.297</td>
<td>0.491</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_3$</td>
<td>-0.494</td>
<td>-0.155</td>
<td>5.489</td>
<td>-0.262</td>
<td>-0.402</td>
<td>-0.097</td>
<td>0.297</td>
<td>0.491</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 3

**2SLS Estimates and their t Statistics**

<table>
<thead>
<tr>
<th></th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
<th>$l$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
<th>$x_6$</th>
<th>$x_7$</th>
<th>$x_8$</th>
<th>$x_9$</th>
<th>$x_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>1.952</td>
<td>-0.037</td>
<td>0.010</td>
<td>-0.002</td>
<td>-0.054</td>
<td>-0.066</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_2$</td>
<td>0.757</td>
<td>-0.798</td>
<td>0.007</td>
<td>-0.066</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_3$</td>
<td>-0.014</td>
<td>1.21</td>
<td>-0.102</td>
<td>0.059</td>
<td>-0.010</td>
<td>-0.042</td>
<td>0.071</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_4$</td>
<td>-0.058</td>
<td>-0.022</td>
<td>0.417</td>
<td>0.013</td>
<td>0.002</td>
<td>-0.016</td>
<td>-0.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_5$</td>
<td>86.528</td>
<td>-127.59</td>
<td>-163.95</td>
<td>-33.37</td>
<td>-6.40</td>
<td>0.360</td>
<td>54.891</td>
<td>-16.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 3 (Continued)

<table>
<thead>
<tr>
<th></th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
<th>$l$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
<th>$x_6$</th>
<th>$x_7$</th>
<th>$x_8$</th>
<th>$x_9$</th>
<th>$x_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>9.068</td>
<td>-0.234</td>
<td>0.074</td>
<td>-0.015</td>
<td>-0.352</td>
<td>-0.494</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_2$</td>
<td>5.717</td>
<td>-9.459</td>
<td>0.041</td>
<td>-0.031</td>
<td>0.224</td>
<td>0.180</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_3$</td>
<td>-0.155</td>
<td>5.489</td>
<td>-0.726</td>
<td>-0.402</td>
<td>-0.097</td>
<td>0.297</td>
<td>0.491</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( y_1 )</td>
<td>( y_2 )</td>
<td>( y_3 )</td>
<td>( y_4 )</td>
<td>( 1 )</td>
<td>( x_2 )</td>
<td>( x_3 )</td>
<td>( x_4 )</td>
<td>( x_5 )</td>
<td>( x_6 )</td>
<td>( x_7 )</td>
<td>( x_8 )</td>
<td>( x_9 )</td>
<td>( x_{10} )</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>( y_1 )</td>
<td>3.926</td>
<td>-7.044</td>
<td>0.027</td>
<td>-0.017</td>
<td>0.197</td>
<td>0.218</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( y_2 )</td>
<td>-4.312</td>
<td>8.544</td>
<td>-0.739</td>
<td>0.328</td>
<td>-0.130</td>
<td>-0.317</td>
<td>0.481</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( y_3 )</td>
<td>60.053</td>
<td>-36.834</td>
<td>-257.81</td>
<td>168.39</td>
<td>-21.847</td>
<td>-10.81</td>
<td>1.837</td>
<td>44.495</td>
<td>-22.157</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( y_4 )</td>
<td>-0.090</td>
<td>-0.016</td>
<td>0.434</td>
<td>0.013</td>
<td>0.002</td>
<td>-0.016</td>
<td>-0.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( y_5 )</td>
<td>-0.048</td>
<td>-0.17</td>
<td>1.94</td>
<td>0.56</td>
<td>-0.95</td>
<td>-2.10</td>
<td>-1.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 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 residuals 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 $\beta_{31}$ to be negative and $\beta_{41}$ to be positive. Also $\beta_{51}$ is expected to be positive and $\beta_{52}$ and $\beta_{53}$ are expected to be negative. The OLS estimates reported in Table 2 have some wrong signs ($\beta_{41}$ and $\beta_{51}$). The 2SLS estimates reported in Table 3 have the correct signs for the coefficient of $y_1$ but none of the coefficients are significant and $\beta_{52}$ 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


545