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.

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

(1)
$$
\begin{aligned}
P_{000} &= 1/D \\
P_{100} &= e^{\beta_1' x}/D \\
P_{010} &= e^{\beta_2' x}/D \\
P_{001} &= e^{\beta_3' x}/D \\
P_{110} &= e^{\beta_4' x}/D \\
P_{101} &= e^{\beta_5' x}/D \\
P_{011} &= e^{\beta_6' x}/D \\
P_{111} &= e^{\beta_7' x}/D
\end{aligned}
$$

where

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

These equations imply the following relations:

$\frac{P_{100}}{P_{000}} = e^{\beta_1' x}$	$\frac{P_{010}}{P_{00}} = e^{\beta_2' x}$	$\frac{P_{001}}{P_{000}} = e^{\beta_3' x}$
$\frac{P_{110}}{P_{010}} = e^{(\beta_4-\beta_2)' x}$	$\frac{P_{110}}{P_{100}} = e^{(\beta_4-\beta_1)' x}$	$\frac{P_{101}}{P_{100}} = e^{(\beta_5-\beta_1)' x}$
$\frac{P_{101}}{P_{001}} = e^{(\beta_5-\beta_3)' x}$	$\frac{P_{011}}{P_{001}} = e^{(\beta_6-\beta_3)' x}$	$\frac{P_{011}}{P_{010}} = e^{(\beta_6-\beta_2)' x}$
$\frac{P_{111}}{P_{011}} = e^{(\beta_7-\beta_6)' x}$	$\frac{P_{111}}{P_{101}} = e^{(\beta_7-\beta_5)' x}$	$\frac{P_{111}}{P_{110}} = e^{(\beta_7-\beta_4)' x}$

These reactions can be written as

$$\text{Log}\,\frac{P(y_1 = 1|y_2y_3)}{P(y_1 = 0|y_2y_3)} = \beta_1'x + (\beta_4 - \beta_2 - \beta_1)'xy_2 + (\beta_5 - \beta_3 - \beta_1)'xy_3 + (\beta_7 - \beta_6 - \beta_5 - \beta_4 + \beta_3 + \beta_2 + \beta_1)'xy_2y_3$$

(2)
$$\text{Log}\,\frac{P(y_2 = 1|y_1y_3)}{P(y_2 = 0|y_1y_3)} = \beta_2'x + (\beta_4 - \beta_2 - \beta_1)'xy_1 + (\beta_6 - \beta_3 - \beta_2)'xy_3 + (\beta_7 - \beta_6 - \beta_5 - \beta_4 + \beta_3 + \beta_2 + \beta_1)'xy_1y_3$$

$$\text{Log}\,\frac{P(y_3 = 1|y_1y_3)}{P(y_3 = 0|y_1y_2)} = \beta_3'x + (\beta_5 - \beta_3 - \beta_1)'xy_1 + (\beta_6 - \beta_3 - \beta_2)'xy_2 + (\beta_7 - \beta_6 - \beta_5 - \beta_4 + \beta_3 + \beta_2 + \beta_1)'xy_1y_2.$$

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:

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

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.

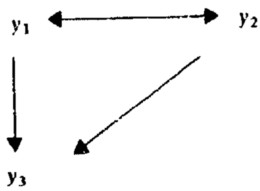

Figure 1

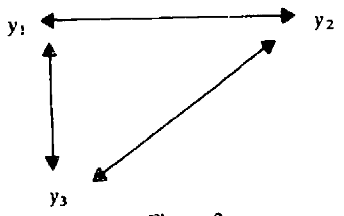

Figure 2

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.

(7)
$$\Pr(y_1 = 1) = \frac{e^{\beta_1' x}}{1 + e^{\beta_1' x}}$$
$$\Pr(y_2 = 1 | y_1) = \frac{e^{\beta_2' x + \gamma y_1}}{1 + e^{\beta_2' x + \gamma y_1}}.$$

These give the joint probabilities

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

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.[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 (2.31)				−7.044			0.027 (0.72)		−0.017 (0.74)	0.197 (1.5)		0.218 (0.87)	
y_3		−4.312 (4.77)			8.544	−0.739 (11.74)	0.328 (1.25)			−0.130 (11.39)	−0.317 (10.45)	0.481 (5.43)		0.041 (2.65)
y_4		−0.090 (−0.48)	−0.016 (−0.17)		0.434 (1.94)		0.013 (0.56)			0.002 (0.95)	−0.016 (−2.10)		−0.026 (−1.75)	
y_5		60.053 (0.82)	−36.834 (−0.7)	−275.81 (−1.36)	168.39 (3.29)	−21.847 (−2.64)	−10.81 (−1.18)			1.837 (1.14)		44.495 (5.66)	−22.157 (−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., Ariene 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.