NBER Working Paper Series

DUMMY ENDOGENOUS VARIABLES IN A SIMULTANEOUS EQUATION SYSTEM

James J. Heckman

Working Paper No. 177

CENTER FOR ECONOMIC ANALYSIS OF HUMAN BEHAVIOR
AND SOCIAL INSTITUTIONS

National Bureau of Economic Research, Inc.
204 Junipero Serra Boulevard, Stanford, CA 94305

May 1977

First draft April 1973
Second draft July 1975

This research was supported by NSF and ASPER Department of Labor grants to the National Bureau of Economic Research. Takeshi Amemiya, R. Bahadur, Linda Edwards, Zvi Griliches, Tom MaCurdy, Marc Nerlove, Randy Olsen, Donald Sant, Peter Schmidt and Arnold Zellner all made valuable comments on drafts of this paper. I assume all responsibility for any remaining errors. The first draft of this paper appeared under the same title as an unpublished NBER paper in April 1973. A second draft was circulated under a different title and presented at the World Econometric Society Meetings at Toronto, 1975.

This paper considers the formulation and estimation of simultaneous equation models with both discrete and continuous endogenous variables. The statistical model proposed here is sufficiently rich to encompass the classical simultaneous equation model for continuous endogenous variables and more recent models for purely discrete endogenous variables as special cases of a more general model.

Interest in discrete data has been fueled by a rapid growth in the availability of microeconomic data sets coupled with a growing awareness of the importance of discrete choice models for the analysis of microeconomic problems (see McFadden, 1976). To date, the only available statistical models for the analysis of discrete endogenous variables have been developed for the purely discrete case. The log-linear or logistic model of Goodman (1970) as expanded by Haberman (1974) and Nerlove and Press (1976) is one such model that has been widely used. The multivariate probit model of Ashford and Sowden (1970), Amemiya (1975) and Zellner and Lee (1965) is another widely used model. This paper expands the multivariate probit structure to accomodate continuous endogenous variables. Alternatively, the model presented here expands the classical simultaneous equation theory to encompass multivariate probit models.

The models developed below rely critically on the notion that discrete endogenous variables are generated by continuous latent variables crossing thresholds. Such models have an honored place in the history of statistics and were first advanced by Pearson (1900). The theory of biserial and tetrachoric correlation is based on this idea. (See Kendall and Stuart,

Vol. II, 1967; Lord and Novick, Chs. 16-20, 1967.) It is argued in this paper that this class of statistical models provides a natural framework for generating simultaneous equation models with both discrete and continuous random variables.

In contrast, the framework of Goodman, while convenient for formulating descriptive models for discrete data, offers a much less natural apparatus for analyzing econometric structural equation models. This is so primarily because the simultaneous equation model is inherently an unconditional representation of behavioral equations while the model of Goodman is designed to facilitate the analysis of conditional representations, and does not lend itself to the unconditional formulations required in simultaneous equation theory.

The structure of this paper is in four parts. In part one general models are discussed. Dummy endogenous variables are introduced in two distinct roles: (1) as proxies for unobserved latent variables and (2) as direct shifters of behavioral equations. Five models incorporating such dummy variables are discussed. Part two, also the longest section, presents a complete analysis of the most novel and most general of the five models presented in part one. This is a model with both continuous and discrete endogenous variables. The issues of identification and estimation are discussed together by proving the existence of consistent estimators. Maximum likelihood estimators and alternative estimators are discussed. In part three, a brief discussion of a multivariate probit model with structural shift is presented. Part four presents a comparison between the models developed in this paper and the models of Goodman and Nerlove and Press.

I. A General Model for the Two Equation Case

Since few new issues arise in the multiple equation case, for expositional simplicity the bulk of the analysis in this paper is conducted for a two equation system. All of the models considered in this paper can be subsumed as special cases of the following pair of simultaneous equations for continuous latent random variables y^*_{1i} and y^*_{2i},

$$(1a) \qquad y^*_{1i} = x_{1i}\alpha_1 + d_i\beta_1 + y^*_{2i}\gamma_1 + U_{1i}$$

$$(1b) \qquad y^*_{2i} = x_{2i}\alpha_2 + d_i\beta_2 + y^*_{1i}\gamma_2 + U_{2i}$$

where dummy variable d_i is defined by

$$(1c) \qquad d_i = 1 \quad \text{iff} \quad y^*_{2i} > 0$$

$$d_i = 0 \quad \text{otherwise},^1$$

and

$$E(U_{ji}) = 0, \quad E(U^2_{ji}) = \sigma_{jj}, \quad E(U_{1i}U_{2i}) = \sigma_{12}, \quad j=1,2, \; i=1,\ldots,I.$$

$$E(U_{ji}U_{j'i'}) = 0, \quad \text{for} \quad j, j' = 1,2, \; i \neq i'.$$

"X_{1i}" and "X_{2i}" are, respectively, $1xK_1$ and $1xK_2$ row vectors of bounded exogenous variables. The joint density of continuous random variables U_{1i}, U_{2i} is $g(U_{1i}, U_{2i})$ which is assumed to be a bivariate normal density in the analysis of Sections II and III. In order to focus attention on the essential features of the argument, the conventional assumptions of classical

[1] Clearly, a second dummy variable could be defined as arising from y^*_{1i} crossing a threshold. Note, too, that the choice of the zero threshold is an arbitrary normalization.

simultaneous equation theory are maintained. In particular, it is assumed that equations (1a) and (1b) are identified if $\beta_1 = \beta_2 = 0$ and both y_{1i}^* and y_{2i}^* are observed for each of the I observations.[1] In this special case, which conforms to the classical simultaneous equation model, standard methods are available to estimate all of the parameters of the structure.

The full model of equations (1a)-(1c) is sufficiently novel to require some discussion. First, note that the model is cast in terms of latent variables y_{1i}^* and y_{2i}^* which may or may not be directly observed. Even if y_{2i}^* is never observed, the event $y_{2i}^* > 0$ is observed and its occurrence is recorded by setting a dummy variable, d_i, equal to one. If $y_{2i}^* < 0$, the dummy variable assumes the value zero. Second, note that if $y_{2i}^* > 0$, structural equations (1a) and (1b) are shifted by an amount β_1 and β_2 respectively.

To fix ideas, several plausible economic models are discussed that may be described by equation system (1a)-(1c). First, suppose that both y_{1i}^* and y_{2i}^* are observed outcomes of a market at time i, say quantity and price. Equation (1a) is the demand curve while equation (1b) is the supply curve. If the price exceeds some threshold (zero inequality (1c), but this can be readily amended to be any positive constant), the government takes certain actions that shift both the supply curve and the demand curve, say a subsidy to consumers and a per unit subsidy to producers. These actions shift the demand curve and the supply curve by the amount β_1 and β_2 respectively.

As another example, consider a model of the effect of laws on the status of blacks. Let y_{1i}^* be the measured income of blacks in state i while

[1] For reasons that become clearer in the analysis of Section II, identification is assumed to be secured through exclusion restrictions or through restrictions on reduced forms for covariance parameters that are estimable.

y^*_{2i} is an unmeasured variable that reflects the state's population sentiment toward blacks. If sentiment for blacks is sufficiently favorable, ($y^*_{2i} > 0$), the state may enact antidiscrimination legislation and the presence of such legislation in state i, a variable that can be measured, is denoted by a dummy variable $d_i = 1$. In the income equation (1a), both the presence of a law and the population sentiment towards blacks is assumed to affect the measured income of blacks. The first effect is assumed to operate discretely while the second effect is assumed to operate in a more continuous fashion. An important question for the analysis of policy is to determine whether or not measured effects of legislation are due to genuine consequences of legislation ($\beta_1 \neq 0$) or to the spurious effect that the presence of legislation favorable to blacks merely proxies the presence of pro-black sentiment that would lead to higher status for blacks in any event ($\gamma_1 \neq 0$). In Section II, methods for consistently estimating the separate effects (β_1 and γ_1) are presented.[1] This example is valuable because it illustrates two conceptually distinct roles for dummy variables: (1) as indicators of latent variables that cross thresholds and (2) as direct shifters of behavioral functions. These two roles must be carefully distinguished in the ensuing analysis.

The model of equations (1a)-(1c) subsumes a wide variety of interesting econometric models. These special cases are briefly discussed in turn.

Case 1. The Classical Simultaneous Equation Model

This model arises when y^*_{1i} and y^*_{2i} are observed, and there is no structural shift in the equations ($\beta_1 = \beta_2 = 0$).

[1] Note that even if sentiment were measured (i.e., y^*_{2i} were known), least squares estimators of equation (1a) are inconsistent because of the correlation of d_i and y^*_{2i} with U_{1i}.

Case 2. The Classical Simultaneous Equation Model with Structural Shift

This model is the same as that of Case 1 except that structural shift is permitted in each equation. It will be shown below that certain restrictions must be imposed on the model in order to generate a sensible statistical structure for this case.

Case 3. The Multivariate Probit Model

This model arises when y_{1i}^* and y_{2i}^* are not observed but the events $y_{1i}^* \lessgtr 0$ and $y_{2i}^* \lessgtr 0$ are observed (i.e., one knows whether or not the latent variables have crossed a threshold). The notation of equations (1a)-(1c) must be altered to accommodate two dummy variables but that modification is obvious. No structural shift is permitted ($\beta_1 = \beta_2 = 0$). This is the model of Ashford and Sowden (1970), Amemiya (1975) and Zellner and Lee (1965).

Case 4. The Multivariate Probit Model with Structural Shift

This model is the same as that of Case 3 except that structural shift is permitted ($\beta_1, \beta_2 \neq 0$).

Case 5. The Hybrid Model

This model arises when y_{1i}^* is observed and y_{2i}^* is not, but the event $y_{2i}^* \gtrless 0$ is observed. No structural shift is permitted.

Case 6. The Hybrid Model with Structural Shift

This model is the same as that of Case 5 except that structural shifts in the equations are permitted.

The hybrid models of Cases 5 and 6 are the most novel and general. Accordingly, these cases receive the greatest attention in the ensuing analysis. Models 2 and 4 are also new but since the analysis of these models follow

directly from the analysis of the hybrid model they receive less attention in this paper. Case 4 is briefly discussed in Section III while Case 2 is never explicitly developed.

II. The Hybrid Model with Structural Shift

In this section, a model with one observed continuous random variable, and one latent random variable is analyzed for the general case of structural shift in the equations. The argument proceeds in the following steps. First, a condition for the existence of a meaningful statistical model is derived. Second, consistent estimators of identified parameters are presented. Third, maximum likelihood estimators are discussed. Finally, some alternative estimators are presented and discussed.

To facilitate the discussion, equations (1a) and (1b) may be written in semi-reduced form as

$$y_{1i} = X_{1i}\pi_{11} + X_{2i}\pi_{12} + d_i\pi_{13} + V_{1i}$$

$$y^*_{2i} = X_{1i}\pi_{21} + X_{2i}\pi_{22} + d_i\pi_{23} + V_{2i}$$

$$d_i = 1 \quad \text{iff} \quad y^*_{2i} \geq 0$$

$$= 0 \text{ otherwise,}$$

where

$$(2) \qquad \pi_{11} = \frac{\alpha_1}{1 - \gamma_1\gamma_2}, \ \pi_{21} = \frac{\alpha_1\gamma_2}{1 - \gamma_1\gamma_2}, \ \pi_{12} = \frac{\alpha_2\gamma_1}{1 - \gamma_1\gamma_2}, \ \pi_{22} = \frac{\alpha_2}{1 - \gamma_1\gamma_2}$$

$$\pi_{13} = \frac{\beta_1 + \gamma_1\beta_2}{1 - \gamma_1\gamma_2}, \ \pi_{23} = \frac{\gamma_2\beta_1 + \beta_2}{1 - \gamma_1\gamma_2}, \quad V_{1i} = \frac{U_{1i} + \gamma_1 U_{2i}}{1 - \gamma_1\gamma_2}$$

$$V_{2i} = \frac{\gamma_2 U_{1i} + U_{2i}}{1 - \gamma_1\gamma_2}$$

In the ensuing analysis it is assumed that exogenous variables included in both X_{1i} and X_{2i} are allocated to either X_{1i} or X_{2i}, but not both. The absence of an asterisk on y_{1i} denotes that this variable is observed. "y^*_{2i}" is not observed. Random variables U_{1i} and U_{2i} are assumed to be bivariate normal random variables. Accordingly, the joint distribution of V_{1i}, V_{2i}, $h(V_{1i}, V_{2i})$, is a bivariate normal density fully characterized by the following assumptions:

$$E(V_{1i}) = 0 \qquad E(V_{2i}) = 0$$

$$E(V_{1i}^2) = \omega_{11} \quad E(V_{1i}V_{2i}) = \omega_{12} \quad E(V_{2i}^2) = \omega_{22}$$

To obtain the true reduced form equations, assume that the conditional probability that d_i is unity given X_{1i} and X_{2i} exists, and denote this probability by P_i. Then the true reduced forms may be written

(3a) $$y_{1i} = X_{1i}\pi_{11} + X_{2i}\pi_{12} + P_i\pi_{13} + V_{1i} + (d_i - P_i)\pi_{13}$$

(3b) $$y^*_{2i} = X_{1i}\pi_{21} + X_{2i}\pi_{22} + P_i\pi_{23} + V_{2i} + (d_i - P_i)\pi_{23}$$

(3c) $$d_i = 1 \quad \text{iff} \quad y^*_{2i} \geq 0$$

$$d_i = 0 \quad \text{otherwise.}$$

The error term in each equation consists of the sum of continuous and discrete random variables that are correlated. The errors have zero conditional mean but if P_i is a nontrivial function of X_{1i}, X_{2i}, heteroscedasticity is present in the errors.

(i) Conditions for Existence of the Model[1]

The first order of business is to determine whether or not the model of equations (1a)-(1b) as represented in reduced form by equations (3a)-(3b) makes sense. Without imposing a further restriction, it does not. The restriction required is precisely the restriction implicitly assumed in writing equations (3a) and (3b), i.e., the restriction that permits one to define a unique probability statement for the events $d_i = 1$ and $d_i = 0$ so that P_i in fact exists. A necessary and sufficient condition for this to be so is that $\pi_{23} = 0$, i.e., that the probability of the event $d_i = 1$ is not a determinant of the event. Equivalently, this assumption can be written as the requirement that $\gamma_2\beta_1 + \beta_2 = 0$. This condition is critical to the analysis and thus deserves some discussion. The argument supporting this assumption is summarized in the following proposition.

Proposition. A necessary and sufficient condition for the model of equations (1a)-(1c) or (3a)-(3c) to be defined is that $\pi_{23} = 0 = \gamma_2\beta_1 + \beta_2$. This assumption is termed the Principle Assumption.

Proof. Sufficiency is obvious. Thus, only necessary conditions are discussed. Denote the joint density of V_{2i}, d_i by $t(V_{2i}, d_i)$ which is assumed to be a proper density in the sense that

$$\sum_{d_i=0,1} \int_{-\infty}^{\infty} t(V_{2i}, d_i)dV_{2i} = 1.$$

From equations (3b) and (3c), the probability that $y^*_{2i} \geq 0$ given $d_i = 1$

[1] I am grateful to Peter Schmidt for correcting an important error in the argument of this section in a previous draft.

must be unity, so that one may write,

$$P_r(V_{2i} > \ell_i | d_i = 1) = 1$$

where the symbols ℓ_i and ℓ_i' are defined by

$$\ell_i = -(X_{1i}\pi_{21} + X_{2i}\pi_{22} + \pi_{23})$$

and

$$\ell_i' = \ell_i + \pi_{23}.$$

Alternatively, one may write this probability as

$$(4a) \quad \int_{\ell_i}^{\infty} t(V_{2i}, 1)dV_{2i} = P_i$$

and obviously

$$(4b) \quad \int_{-\infty}^{\ell_i} t(V_{2i}, 1)dV_{2i} = 0.$$

Using similar reasoning, one can conclude that

$$(4c) \quad \int_{-\infty}^{\ell_i'} t(V_{2i}, 0)dV_{2i} = 1 - P_i$$

and

$$(4d) \quad \int_{\ell_i'}^{\infty} t(V_{2i}, 0)dV_{2i} = 0.$$

The sum of the left hand side terms of equations (4a)-(4d) equals the sum of the right hand side terms which should equal one if the probability of the event $d_i = 1$ is meaningfully defined. If $\pi_{23} = 0$, this is the case. But if $\pi_{23} < 0$, the sum of the left hand side terms falls short of one while if $\pi_{23} > 0$, this sum exceeds one. Q.E.D. Notice that this argument does not rely on the assumption that V_{2i} is normally distributed but does

rely on the assumption that V_{2i} has positive density at almost all points on the real line.

An intuitive motivation for this condition is possible. Suppose that one rewrites equations (1a)-(1c) to exclude d_i, i.e., write

$$y^*_{1i} = X_{1i}\alpha_1 + y^*_{2i}\gamma_1 + U_{1i}$$

$$y^*_{2i} = X_{2i}\alpha_2 + y^*_{1i}\gamma_2 + U_{2i}.$$

$$d_i = 1 \quad \text{iff} \quad y^*_{2i} > 0.$$

$$d_i = 0 \text{ otherwise.}$$

Note that y^*_{1i} is an unobserved latent variable. The random variable y_{1i} is observed and is defined by the following equation

$$y_{1i} = y^*_{1i} + d_i\beta_1.$$

Making the appropriate substitutions of y_{1i} for y^*_{1i} in the system given above, one concludes that

$$y_{1i} = X_{1i}\alpha_1 + d_i\beta_1 + y^*_{2i}\gamma_1 + U_{1i}$$

$$y^*_{2i} = X_{2i}\alpha_2 + (y_{1i} - d_i\beta_1)\gamma_2 + U_{2i}.$$

Invoking the Principle Assumption, one reaches equations (1a)-(1c) <u>including</u> d_i. Thus the dummy shift variable $d_i\beta_1$ may be viewed as a veil that obscures measurement of the latent variable y^*_{1i}. But, as will be shown, the veil can be removed, i.e., β_1 can be estimated.

It is important to note that the principle assumption does not rule out structural shift in equations (1a) and (1b). It simply restricts the nature of the shift. However, the Principle Assumption does exclude any

structural shift in the reduced form equation that determines the probability of a shift (equation (3b)).

(ii) Identification of Parameters: Indirect Least Squares Estimators

Given the principle assumption, equation system (3a)-(3c) may be written as

$$(5a) \qquad y_{1i} = X_{1i}\pi_{11} + X_{2i}\pi_{12} + P_i\pi_{13} + V_{1i} + (d_i - P_i)\pi_{13}$$

$$(5b) \qquad y^*_{2i} = X_{1i}\pi_{21} + X_{2i}\pi_{22} + V_{2i}$$

$$(5c) \qquad d_i = 1 \quad \text{iff} \quad y^*_{2i} \geq 0$$

$$d_i = 0 \text{ otherwise.}$$

Estimation of equation (5b) is a problem in probit analysis. Subject to the standard requirements for identification and existence of probit estimates (see Nerlove and Press, 1976), one may normalize by $\omega_{22}^{1/2}$ and estimate

$$(6) \qquad \pi^*_{21} = \frac{\pi_{21}}{(\omega_{22})^{1/2}}, \quad \pi^*_{22} = \frac{\pi_{22}}{(\omega_{22})^{1/2}},$$

by using the reduced form probit function to estimate the conditional probabilities of the events $d_i = 1$ and $d_i = 0$.

To determine how to estimate the parameters of the equation (5a), it is useful to write the conditional expectation of y_{1i} given d_i, X_{1i} and X_{2i}, i.e.,

$$(7) \qquad E(y_{1i}|X_{1i}, X_{2i}, d_i) = X_{1i}\pi_{11} + X_{2i}\pi_{12} + d_i\pi_{13} + E(V_{1i}|d_i, X_{1i}, X_{2i}).$$

Utilizing a result familiar in the theory of biserial correlation (see, e.g., Tate, 1954 or Johnson and Kotz, 1972, Vol. 4),

$$E(V_{1i}|d_i, X_{1i}, X_{2i}) = \frac{\omega_{12}}{(\omega_{22})^{1/2}} (\lambda_i d_i + \tilde{\lambda}_i(1 - d_i))$$

where $$\lambda_i = \frac{\phi(c_i)}{1 - \Phi(c_i)}$$

with $c_i = -(X_{1i}\pi^*_{21} + X_{2i}\pi^*_{22})$, where ϕ and Φ are the density and distribution function of a standard normal random variable and

$$\tilde{\lambda}_i = -\lambda_i\left[\frac{\Phi(-c_i)}{\Phi(c_i)}\right].$$

If one knew, or could estimate, $E(V_{1i}|d_i, X_{1i}, X_{2i})$, it could be entered as a regressor in equation (7) and parameters π_{11}, π_{12}, π_{13} and $\omega^*_{12} = \frac{\omega_{12}}{(\omega_{22})^{1/2}}$ could be estimated by standard least squares methods. Since the normalized parameters of equation (5b) are estimable, so is c_i and hence λ_i, $\tilde{\lambda}_i$ and $E(V_{1i}|d_i, X_{1i}, X_{2i})$. Elsewhere (Heckman, 1977) it is shown that use of estimated values of λ_i and $\tilde{\lambda}_i$ instead of actual values as regressors in equation (7) leads, under general conditions, to consistent parameter estimates of all the regression coefficients in that equation.

Given this result, all of the parameters of equation (7) are estimable. Note in particular that the correlation between V_{1i} and $\frac{V_{2i}}{(\omega_{22})^{1/2}} = V^*_{2i}$ is also estimable even though there are no direct observations on y^*_{2i}. This result is a familiar one in the theory of biserial correlation.[1]

To see how to estimate the reduced form variance, ω_{11}, note that the general model, of which equation (7) is the conditional expectation, may be written as

---

[1]Note that the use of the estimated value of $E(V_{1i}|d_i, X_{1i}, X_{2i})$ as a regressor to estimate the parameters of the disturbance covariance structure closely parallela Telser's (1964) procedure of utilizing least squares residuals from other equations in a system of equations to estimate the parameters of the inter-equation covariance structure.

$$y_{1i} = E(y_{1i}|X_{1i}, X_{2i}, d_i) + \eta_i \tag{8}$$

where

$$E(\eta_i|X_{1i}, X_{2i}, d_i) = 0$$

and

$$E(\eta_i^2|X_{1i}, X_{2i}, d_i) = \omega_{11}[(1 - \rho^2) + \rho^2(d_i q_i + (1 - d_i)s_i]$$

where

$$\rho = \frac{\omega_{12}}{(\omega_{11}\omega_{22})^{1/2}}$$

$$q_i = 1 + \lambda_i c_i - \lambda_i^2$$

$$s_i = 1 - \tilde{\lambda}_i c_i - \tilde{\lambda}_i^2.$$

(See Johnson and Kotz, 1972, Vol. 4).

Since ω_{12}^* can be consistently estimated, and since $\omega_{12}^* = (\omega_{11})^{1/2}\rho$, a consistent estimate of ω_{11} is possible using the estimated residuals from least squares estimates of equation (8). Denote the estimated residual for observation i by $\hat{\eta}_i$. Then estimate ω_{11} from the following formula:

$$\hat{\omega}_{11} = \frac{1}{I}\sum_{i=1}^{I}\hat{\eta}_i^2 + (\omega_{12}^*)^2 \quad [1 - \frac{1}{I}(\sum_{i=1}^{I} d_i q_i + (1 - d_i)s_i)]$$

where estimated values of q_i and s_i are used in place of actual values. This estimate is consistent.[1]

Given consistent estimators of reduced form coefficients, estimators of the structural parameters are easily obtained. Since the coefficient of equation (5b) can only be estimated up to an unknown constant of proportionality,

[1]Further, it is guaranteed to be positive. One can prove that the second term on the right hand side must be positive.

$\omega_{22}^{1/2}$, it is not possible to estimate all of the coefficients of equations (1a) and (1b). Some of these coefficients can only be estimated up to an unknown constant of proportionality.

From equation (2), it is clear that if some exogenous variables appear in equation (1a) that do not appear in (1b) it is possible to estimate $\gamma_2^* = \gamma_2/\omega_{22}^{1/2}$. Take the j^{th} variable in X_{1i}, denoted $X_{1_j i}$, and its associated estimable reduced form coefficients π_{11_j} and $\pi^*_{21_j}$. Assume that this variable is not included in X_{2i}. Taking the ratio of the estimate of the second coefficient to the estimate of the first yields a consistent estimator of γ_2^*,

$$\hat{\gamma}_2^* = \frac{\hat{\pi}^*_{21_j}}{\hat{\pi}_{11_j}}, \quad j = 1, \ldots, J_1$$

where "^" denotes an estimate and where J_1 is the number of variables in X_{1i} not contained in X_{2i}, adopting the harmless convention that the first J_1 variables in X_{1i} are such variables. Similarly, one may consistently estimate $\gamma_1(\omega_{22})^{1/2} = \gamma_1^*$ if some variables included in X_{2i} do not appear in X_{1i}. Utilizing notation previously introduced,

$$\frac{\hat{\pi}_{12_j}}{\hat{\pi}^*_{22_j}} = \widehat{\gamma_1(\omega_{22})^{1/2}} = \gamma_1^*, \quad j = 1, \ldots, J_2$$

where J_2 is the number of variables in X_{2i} not contained in X_{1i} and the first J_2 variables in X_{2i} are assumed not to be included in X_{1i}.

In general, the model will be overidentified if it is identified at all. The procedure for resolving the overidentification problem is entirely conventional and will be discussed below. Assume, for the moment, that this problem can be resolved. Given unique estimates of γ_1^* and γ_2^*,

one can exploit the information in equations (2) and (6) to estimate

$$\alpha_1,\ \alpha_2(\omega_{22})^{-1/2} = \alpha_2^*,\ \beta_2(\omega_{22})^{-1/2} = \beta_2^*,\ \text{and}\ \beta_1.$$

The only parameters that remain to be identified are the disturbance covariance terms of the structural equations. Without further information, it is not possible to estimate all of the parameters of the structural equation covariance matrix, just the normalized parameters σ_{11}, $\sigma_{12}^* = \sigma_{12}\omega_{22}^{-1/2}$, $\sigma_{22}^* = \sigma_{22}\omega_{22}^{-1}$.

To see this, note that

$$\omega_{11} = E(V_{1i}^2) = \frac{\sigma_{11} + 2\gamma_1\sigma_{12} + \gamma_1^2\sigma_{22}}{(1 - \gamma_1\gamma_2)^2} = \frac{\sigma_{11} + 2\gamma_1^*\sigma_{12}^* + (\gamma_1^*)^2\sigma_{22}^*}{(1 - \gamma_1^*\gamma_2^*)^2} \tag{9}$$

$$\omega_{12} = E(V_{1i}V_{2i}) = \frac{\gamma_2\sigma_{11} + (1 + \gamma_1\gamma_2)\sigma_{12} + \gamma_1\sigma_{22}}{(1 - \gamma_1\gamma_2)^2}$$

$$\frac{\omega_{12}}{\omega_{22}^{1/2}} = \frac{E(V_{1i}V_{2i})}{\omega_{22}^{1/2}} = \frac{\gamma_2^*\sigma_{11} + (1 + \gamma_1^*\gamma_2^*)\sigma_{12}^* + \gamma_1^*\sigma_{22}^*}{(1 - \gamma_1^*\gamma_2^*)^2} \tag{10}$$

$$\omega_{22} = E(V_{wi}^2) = \frac{\gamma_2^2\sigma_{11} + 2\gamma_2\sigma_{12} + \sigma_{22}}{(1 - \gamma_1\gamma_2)^2}$$

$$1 = \frac{E(V_{2i}^2)}{\omega_{22}} = \frac{(\gamma_2^*)^2\sigma_{11} + 2(\gamma_2^*)\sigma_{12}^* + \sigma_{22}^*}{(1 - \gamma_1^*\gamma_2^*)^2} \text{ 1/} \tag{11}$$

Since $\gamma_1\gamma_2$, γ_1^* and γ_2^* are estimable parameters, and since consistent estimators of the left hand side terms of equations (9) and (10) are available, these

---

[1] This final restriction was suggested to me by Professor L. Lee.

equations, supplemented by equation (11) provide three linear equations in the three unknowns σ_{11}, σ^*_{12}, σ^*_{22}. In general, these equations can be solved for unique estimators.

(iii) Maximum Likelihood Estimators

The preceding analysis not only yields criteria for the identification of structural coefficients but also produces consistent estimators for identified coefficients. These estimators are useful for providing estimates enroute to deriving maximum likelihood estimators, but they are not, in general, efficient. The maximum likelihood estimator that is discussed next is asymptotically efficient.

The density function for the disturbances V_{1i}, V_{2i} is bivariate normal. For notational simplicity normalize V_{2i} by $\omega_{22}^{1/2}$ and define $V^*_{2i} = V_{2i}\omega_{22}^{-1/2}$. The joint density of V_{1i}, V^*_{2i} is $h(V_{1i}, V^*_{2i})$. Since $d_i = 1$ iff $y^*_{2i} > 0$ and $d_i = 0$ otherwise, the density of y_{1i}, d_i is given by

$$(12) \qquad f(y_{1i}, d_i) = \left[\int_{c_i}^{\infty} h(y_{1i} - X_{1i}\pi_{11} - X_{2i}\pi_{12} - d_i\pi_{13}, V^*_2)\, dV^*_2\right]^{d_i} \cdot \left[\int_{-\infty}^{c_i} h(y_{1i} - X_{1i}\pi_{11} - X_{2i}\pi_{12} - d_i\pi_{13}, V^*_2)\, dV^*_2\right]^{1-d_i}$$

where c_i has previously been defined as

$$c_i = -(X_{1i}\pi^*_{21} + X_{2i}\pi^*_{22}).^1$$

Using equation (2), the density may be rewritten in terms of identified structural parameters.

---

[1] For a discussion of this rather unusual density see Appendix A.

Assuming random sampling, the likelihood function for the hybrid model with structural shift is

$$(13)\quad \mathcal{L} = \prod_{i=1}^{I} f(y_{1i}, d_i),$$

where I is sample size. Under conditions specified below, this function possesses an optimum, and the maximum likelihood estimators have desirable large sample properties. The identification procedure previously discussed provides an algorithm for generating initial consistent estimators so that one Newton step produces asymptotically efficient estimators. These initial estimators are particularly valuable because likelihood function (13) is not a globally concave function of the structural parameters.

Note that if $\omega_{12} = 0$, so that the reduced form disturbances V_{1i} and V^*_{2i} are independent, density $h(V_{1i}, V^*_{2i})$ factors into a product of marginal densities $h_1(V_{1i})h_2(V^*_{2i})$ and $f(y_{1i}, d_i)$ becomes

$$f(y_{1i}, d_i) = h_1(y_{1i} - X_{1i}\pi_{11} - X_{2i}\pi_{12} - \pi_{13}d_i) \cdot \left[\int_{c_i}^{\infty} h_2(V^*_2)dV^*_2\right]^{d_i} \left[\int_{-\infty}^{c_i} h_2(V^*_2)dV^*_2\right]^{1-d_i}$$

so that regression estimators of equation (5a) and probit estimators of equation (5b) are maximum likelihood estimators. In most practical problems the assumption that $\omega_{12} = 0$ is untenable.

In addition to the ordinary identification conditions previously discussed, another condition is required in order for likelihood function (13) to possess a well defined maximum with respect to its parameters. In order to understand this condition, it is helpful to use conditional normal theory to write density $f(y_{1i}, d_i)$ as

$$f(y_{1i}, d_i) = h_1(y_{1i} - X_{1i}\pi_{11} - X_{2i}\pi_{12} - d_i\pi_{13})[\Phi(r_i)]^{d_i}[\Phi(-r_i)]^{1-d_i}$$

where Φ is the cumulative distribution of the univariate normal, and

$$r_i = \frac{c_i - \rho(V_{1i}/\omega_{11}^{1/2})}{(1-\rho^2)^{1/2}}$$

where $V_{1i} = y_{1i} - X_{1i}\pi_{11} - X_{2i}\pi_{12} - d_i\pi_{13}$

and $$\rho = \frac{\omega_{12}}{(\omega_{11}\omega_{22})^{1/2}} .$$

This representation of the density is both computationally and theoretically convenient.

In a sample of size I, classify the observations into two groups depending on whether or not the dummy variable d_i is one. Among the I_1 observations with $d_i = 1$, denote the smallest r_i by $r_i^{(1)}\text{min.}$, and the largest by $r_i^{(1)}\text{max}$. Among the $I_0 (= I - I_1)$ observations with $d_i = 0$, denote the largest r_i by $r_i^{(0)}\text{max}$, and the smallest by $r_i^{(0)}\text{min}$. Then, likelihood function (13) possesses no interior optimum in a compact parameter set if

$$(14) \qquad r_i^{(0)}\text{max} \leq r_i^{(1)}\text{min} \quad \text{or if} \quad r_i^{(0)}\text{min} \geq r_i^{(1)}\text{max}$$

so that there is no overlap among the values of the r_i classified by the occurrence of the event.[1]

This condition arises in standard probit analysis (see Nerlove and Press, 1976). To understand it, consider estimates of an ordinary probit model. If any variable perfectly classifies the outcomes of a discrete experiment, the coefficient of that variable becomes unbounded and is not

[1] A similar condition arises in the conditional logit model. See McFadden (1974, Appendix).

identified. This phenomenon may arise in the current problem even if no linear exogenous variable (or/combination of exogenous variables) perfectly classifies outcomes because of the presence of random variable V_{1i} in r_i. There is a positive probability that condition (14) will be met and maximum likelihood estimators will fail to exist. However, using standard results in the theory of order statistics, it is trivial to establish that the probability that condition (14) is met tends to zero if sample size expands by "fixed in repeated samples" sampling.[1] Given their existence, maximum likelihood estimators are consistent, asymptotically normal and efficient.[2]

---

[1] Briefly consider the condition $r_i^{(0)}\max \leq r_i^{(1)}\min$. If one data configuration (i.e., a choice of X_{1i}, X_{2i}) is considered and the number of observations on that configuration becomes large, the condition becomes

$$\max \left\{ - \frac{\rho}{(1-\rho^2)^{1/2}} V_{1i}^1 \right\}_{i=1}^{I_1} \leq \min \left\{ - \frac{\rho}{(1-\rho^2)^{1/2}} V_{1i}^0 \right\}_{i=1}^{I_0}$$

where V_{1i}^1 is random variable V_{1i} conditional on $d_i = 1$ while V_{1i}^0 is random variable V_{1i} conditional on $d_i = 0$. One may write

$$V_{1i}^1 = k_1 + \eta_{1i} \text{ and } V_{1i'}^0 = k_0 + \eta_{2i'}, \; i = 1,\ldots, I_1, \; i' = 1,\ldots, I_0,$$

where η_{1i} and $\eta_{2i'}$ are continuous independent random variables. The probability that $\eta_{1i} - \eta_{2i'} > k_0 - k_1$ is less than one for any pair of observations from the disjoint sets. Hence, in large samples, condition $r_i^{(0)}\max \leq r_i^{(1)}\min$ occurs with probability zero.

[2] It is a straightforward exercise to verify that LeCam's (1953) generalization of the classical Cramer conditions applies to the model in the text for suitably chosen values of the parameters. These conditions are local in nature and imply the existence of some root of the likelihood equations that is consistent and asymptotically normally distributed. Olsen (1977) verifies the LeCam conditions for a model based on the one in the text. Olsen's structural model is the reduced form of the current model with the exception that π_{13} is set at zero in his work. His proofs carry over to the more general case.

One point is not obvious, and is not covered in Olsen's work. Since a model with $\pi_{13} \neq 0$ superficially resembles the model advanced by Amemiya (1974)

(iv) Some Alternative Estimators

Since maximum likelihood estimation is computationally cumbersome, it is useful to consider alternative estimators for the hybrid model. In addition, the problem of the multiplicity of consistent estimators that arises in an overidentified model remains to be resolved. It will be shown that the fact that y_{2i}^* is never observed has important consequences that cause the analysis in this paper to differ dramatically from conventional simultaneous equation theory.

Consider equations (1a)-(1c), rewritten below to facilitate the exposition. Because it is assumed that y_{1i}^* is observed, it is replaced by y_{1i}. With this change, the equations become

---

to demonstrate the inconsistency of maximum likelihood estimators for the parameters of a discontinuous density, it is worth verifying that the expectation of the first partial of the log of the density of $f(y_{1i}, d_i)$ with respect to π_{13} vanishes when the expectation is taken with respect to the density evaluated at the true parameter values.

Denote E_0 as the expectation taken with respect to the density of y_{1i} and d_i when the true parameters are used in the density. Let $V_{1i} = y_{1i} - X_{1i}\pi_{11} - X_{2i}\pi_{12} - d_i\pi_{13}$. Then

$$E_0\left[\frac{\partial \ln f(y_{1i}, d_i)}{\partial \pi_{13}}\right] = E_0\left[\frac{d_i}{(1-\rho^2)} \frac{\int_{c_i}^{\infty} \left(\frac{V_{1i}}{\omega_{11}^{1/2}} - \rho \frac{V_{2i}^*}{\omega_{11}^{1/2}}\right) h(V_{1i}, V_{2i}^*)dV_{2i}^*}{\int_{c_i}^{\infty} h(V_{1i}, V_{2i}^*)dV_{2i}^*}\right]$$

Since

$$\int_{-\infty}^{\infty}\int_{c_i}^{\infty} \frac{V_{1i}}{\omega_{11}} h(V_{1i}, V_{2i}^*)dV_{2i}^*dV_{1i} = \rho \int_{-\infty}^{\infty}\int_{c_i}^{\infty} \frac{V_{2i}^*}{(\omega_{11})^{1/2}} h(V_{1i}, V_{2i}^*)dV_{2i}^*dV_{1i}$$

(See, e.g., Johnson and Kotz, 1972, Vol. 4), it follows that

$$E_0\left[\frac{\partial \ln f(y_{1i}, d_i)}{\partial \pi_{13}}\right] = 0$$

as desired.

(1a) $y_{1i} = X_{1i}\alpha_1 + d_i\beta_1 + y^*_{2i}\gamma_1 + U_{1i}$

(1b) $y^*_{2i} = X_{2i}\alpha_2 + d_i\beta_2 + y_{1i}\gamma_2 + U_{2i}$

where dummy variable d_i is defined by

(1c) $d_i = 1$ iff $y^*_{2i} > 0$

$d_i = 0$ otherwise,

and the Principle Assumption is invoked so that $\gamma_2\beta_1 + \beta_2 = 0$.

Using the results of the previous analysis, it is possible to estimate equations (1a) and (1b) directly using the reduced form coefficient estimates to generate instruments. To see this, note that it is possible to use equation (5b) to estimate the expectation of $y^*_{2i}/\omega_{22}^{1/2}$ conditional on X_{1i} and X_{2i}.

$$\hat{y}^*_{2i}/\omega_{22}^{1/2} = X_{1i}\hat{\pi}^*_{21} + X_{2i}\hat{\pi}^*_{22}.$$

From the probit estimates of (5b) it is possible to estimate the probability of the event $d_i = 1$ conditional on values of X_{1i} and X_{2i} ($\hat{P}_i$). Replacing y^*_{2i} and d_i by their estimated expectations, equation (1a) becomes

(1a') $y_{1i} = X_{1i}\alpha_1 + \hat{P}_i\beta_1 + (\hat{y}^*_{2i}\omega_{22}^{-1/2})\gamma^*_1$

$$+ U_{1i} + (d_i - \hat{P}_i)\beta_1 + \gamma^*_1(y^*_{2i}\omega_{22}^{-1/2} - \hat{y}^*_{2i}\omega_{22}^{-1/2}).$$

Least squares applied to equation (1a') yields unique consistent estimators of α_1, β_1 and γ^*_1. The proof is trivial and hence is omitted. Estimation of this equation resolves the problem of the multiplicity of estimators that arise from the application of indirect least squares discussed in Section (ii).

Precisely the same procedure may be used to estimate the parameters of equation (1b). There is one new point. The choice of a normalization

rule in a simultaneous equation system is usually arbitrary. Here the choice is important because y^*_{2i} is never observed, although its expectation can be estimated. In equations (1a) and (1b), y_{1i} is selected as the dependent variable. Substituting estimated conditional means for endogenous variables, equation (1b) may be written as

$$(1b') \quad y_{1i} = -\frac{1}{\gamma^*_2}(X_{2i}\alpha^*_2 + \hat{P}_i\beta^*_2 - (\hat{y^*_{2i}}\omega_{22}^{-1/2})) -$$

$$\frac{1}{\gamma^*_2}(U_{2i}\omega_{22}^{-1/2} + (d_i - \hat{P}_i)\beta^*_2 - (y^*_{2i}\omega_{22}^{-1/2} - \hat{y^*_{2i}}\omega_{22}^{-1/2})).$$

It is straightforward to prove that least squares estimators of equation (1b') are consistent.

There is one further problem.[1] Although the Principle Assumption requires $\gamma_2\beta_1 + \beta_2 = 0$, this constraint is not imposed in the preceding analysis. One way to impose the constraint is to estimate equation (1a') and use the fitted value of β^*_2/γ^*_2 $(= -\beta_1)$ as a parameter in equation (1b'). A more satisfactory approach that is computationally more burdensome is to impose the constraint directly in formulating joint least squares estimators for equations (1a') and (1b').

It is tempting to use the residuals computed from the fitted equations (1a') and (1b') to directly estimate the structural covariance terms σ_{11}, σ^*_{12} and σ^*_{22}. A direct application of structural two stage least squares formulae will not work precisely because y^*_{2i} is never observed. If estimated values of y^*_{2i} are used in place of actual values, the residuals from (1c) and (1b) will not permit identification of the structural covariances.

One method for circumventing this difficulty is to use the estimated structural parameters to solve for the reduced form parameters π^*_{21}, π^*_{22},

---

[1] I owe this point to Tom MaCurdy.

π_{11}, π_{12} and π_{13}. These estimates may be treated as known parameters in estimating equation (7). Hence it is possible to estimate ω_{11} and $\omega_{12}\omega_{22}^{-1/2}$, and by use of equations (9), (10) and (11) to obtain unique consistent estimates of σ_{11}, σ_{12}^{*} and σ_{22}^{*}.

While these estimators are simply computed, consistent, and free of the nonuniqueness problem that plagues indirect least squares estimators in the overidentified case, they are not asymptotically efficient nor are their asymptotic distributions simple. The standard formulae used for the computation of large sample parameter variances is inappropriate. This is so because $y_{2i}^{*}\omega_{22}^{-1/2}$ is never directly observed and only an estimated value of this variable is available.

To understand these difficulties, it is useful to discuss two special cases that are of interest in practical empirical work. First consider estimation of equation (1a) when $\beta_1 = 0$ so that no direct structural shift is present in that equation. Then consider a case in which $\gamma_1 = 0$ so that no unobserved latent variable is present in equation (1a). In both cases consistent parameter estimates are available, as has been shown.

Consider the disturbance from equation (1a') under the assumption that $\beta_1 = 0$. Denote the composite disturbance by $\tilde{U}_{1i}$

$$\tilde{U}_{1i} = U_{1i} + \gamma_1^{*}(y_{2i}^{*}\omega_{22}^{-1/2} - \hat{y}_{2i}^{*}\omega_{22}^{-1/2}).$$

The crucial feature of these residuals is that they are not independent across observations nor are they identically distributed. Accordingly, standard central limit theorems do not apply to regression coefficient estimates of equation (1a'). In particular, it is not the case that the standard estimator of the regression parameter variance-covariance matrix,

$$\hat{\sigma}_{11}\begin{bmatrix}\Sigma\, X_{1i}X_{1i}' & \Sigma\, X_{1i}\hat{y}_{2i}^{*}\omega_{22}^{-1/2} \\ \Sigma\, X_{1i}\hat{y}_{2i}^{*}\omega_{22}^{-1/2} & \Sigma\,(\hat{y}_{2i}^{*}\omega_{22}^{-1/2})^{2}\end{bmatrix}^{-1}$$

is the appropriate asymptotic variance-covariance matrix for the regression coefficients.

The source of the problem comes in the final term in $\tilde{U}_{1i}$. Utilizing the reduced form for $y_{2i}^{*}\omega_{22}^{-1/2}$, this term becomes

$$y_{2i}^{*}\omega_{22}^{-1/2} - \hat{y}_{2i}^{*}\omega_{22}^{-1/2} = X_{1i}(\pi_{21}^{*} - \hat{\pi}_{21}) + X_{2i}(\pi_{22}^{*} - \hat{\pi}_{22}^{*}) + V_{1i}$$

V_{1i} is an iid random variable. But the first two terms are not iid. Since $\hat{\pi}_{21}^{*}$ and $\hat{\pi}_{22}^{*}$ are maximum likelihood estimators, they possess asymptotic normal limiting distribution, and in large samples the first two terms converge in probability to zero. But their rate of convergence is not fast enough.

In fact, regression estimates of equation (1a') with $\beta_1 = 0$ obey the following relationship.

$$\sqrt{I}\begin{bmatrix}\hat{\alpha}_1 - \alpha_1 \\ \hat{\gamma}_1^{*} - \gamma_1^{*}\end{bmatrix} = \left[\frac{1}{I}\begin{bmatrix}\Sigma\, X_{1i}X_{1i}' & \Sigma\, X_{1i}'(\hat{y}_{2i}^{*}\omega_{22}^{-1/2}) \\ \Sigma\, X_{1i}(y_{2i}^{*}\omega_{22}^{-1/2}) & \Sigma\,(\hat{y}_{2i}^{*}\omega_{22}^{-1/2})^{2}\end{bmatrix}\right]^{-1} \cdot$$

$$\left[\frac{1}{\sqrt{I}}\begin{bmatrix}\Sigma\,(X_{1i}'(U_{1i} + V_{1i}) \\ \Sigma\,(\hat{y}_{2i}^{*}\omega_{22}^{-1/2})(U_{1i} + V_{ii})\end{bmatrix} + \right.$$

$$\left.\frac{1}{\sqrt{I}}\begin{bmatrix}\Sigma\,(X_{1i}'(X_{1i}[\pi_{21}^{*} - \hat{\pi}_{21}^{*}] + X_{2i}[\pi_{22}^{*} - \hat{\pi}_{22}^{*}] \\ \Sigma(\hat{y}_{2i}^{*}\omega_{22}^{-1/2}(X_{1i}[\pi_{21}^{*} - \hat{\pi}_{21}^{*}] + X_{2i}[\pi_{22}^{*} - \hat{\pi}_{22}^{*}]\end{bmatrix}\right]$$

Both terms on the right hand side converge to proper normal random variables.[1] Accordingly, the standard formula for the asymptotic variance-covariance matrix is inapplicable. Although the correct asymptotic variance-covariance matrix can be estimated, the computational burden of doing so is greater than direct optimization of the likelihood function.

Now turn to the second case. Consider the estimation of equation (1a') for a case in which $\gamma_1 = 0$ so that no latent variable is present in equation (1a). The analysis of this case is thoroughly conventional.

The estimated reduced form probability $\hat{P}_i$ may be used as an instrumental variable for d_i. Standard instrumental variable formulas may be used to estimate the appropriate asymptotic variance-covariance matrix of the coefficients.

The procedure to be used is as follows. $\hat{P}_i$ may be employed as an instrument for d_i in equation (1a'), and consistent estimators of α_1 and β_1 may be produced. Using actual values for d_i, and the estimated coefficients, one may estimate the residuals for each observation which when squared, summed and divided by I yield an estimator of σ_{11}. The appropriate asymptotic variance covariance matrix for the regression coefficients α_1, β_1 may be consistently estimated by the standard instrumental variable formula

$$\hat{\sigma}_{11}\begin{bmatrix}\Sigma X_{1i}X_{1i}' & \Sigma X_{1i}d_i \\ \Sigma X_{1i}P_i & \Sigma d_i\end{bmatrix}^{-1} \cdot \begin{bmatrix}\Sigma X_{1i}X_{1i} & \Sigma X_{1i}'\hat{P}_i \\ \Sigma X_{1i}\hat{P}_i & \Sigma \hat{P}_i^2\end{bmatrix} \cdot \begin{bmatrix}\Sigma X_{1i}X_{1i}' & \Sigma X_{1i}'\hat{P}_i \\ \Sigma X_{1i}d_i & \Sigma d_i\end{bmatrix}^{-1}$$

Note that since the residuals from the prediction of d_i $(d_i - \hat{P}_i)$ are not guaranteed to be orthogonal to the X_{1i} regressors, the instrumental variable formula is not equivalent to the standard two stage least squares formula, and

[1] The proof is straightforward. See Heckman (1977a, Appendix A for a more complete discussion).

the instrumental variable estimator is not equivalent to the standard two stage least squares estimator.[1]

Note further that if the sole purpose of the analysis is to estimate equation (1a), it is not necessary to estimate probit functions at all. It is possible to generate an instrumental variable for d_i by estimating a simple linear probability model with d_i as a dependent variable that contains at least all of the variables in X_{1i} and some other exogenous variable as regressors.[2] If these estimators are utilized, the standard two stage least squares procedure applies and predicted values of d_i may be utilized as regressors since in this case the regression residuals from the prediction of d_i are constructed to be orthogonal to the X_{1i} regressors. This result simply restates the well known point that it is unnecessary to obtain consistent estimates of the parameters of reduced form equations in order to consistently estimate structural equations.

Since the linear probability procedure is the simplest one to use, it is recommended. However, it is likely that the use of the probit instrument results in more efficient estimates although no proof of this assertion is offered.

The discussion of these two cases is illuminating. For both cases simply computed consistent parameter estimates are available. In the first case, with an unobserved random variable present, the estimators converge to a normal distribution but the theoretically appropriate asymptotic covariance matrix is cumbersome to compute. In this case it is suggested that analysts utilize the consistent estimators discussed in this section as starting values for at least one Newton step towards the likelihood optimum to produce estimators with desirable large sample properties.[3] The

[1] See Sant (1975).

[2] This follows directly from the analysis of Kelejian (1971).

[3] A copy of a computer program to produce both one step Newton iterates and full information maximum likelihood estimates is available, on request, from the author for the cost of duplication, postage and handling charges.

second case requires only a simple application of conventional instrumental variable estimator theory. For both cases and in the general case that contains both special cases, full system maximum likelihood estimation will produce asymptotically efficient estimates and is certainly recommended for all but the special second case.

The hybrid model can be generalized in several ways. Two extensions are particularly important. First, several dummy indicator variables can be introduced into the model. Two types of multiple dummy shift variables can be introduced. The first type of dummy variable represents a polytomization of a single latent variable and is appropriate for the case of ordered dichotomous variables. The second type is for intrinsically unordered case.[1] These models, and obvious multivariate extensions, are briefly discussed in Appendix B. Second, the random variable y_{1i} may be a truncated variable. This case, which nests Tobin's model (1958) into a simultaneous equation system, follows as a trivial extension of the previous analysis and hence is not discussed here.

III. Multivariate Probit Models with Structural Shift

In this section multivariate probit models are discussed. In these models there are no observed latent variables so that the only information available is that $y^*_{1i} \gtrless 0$ and $y^*_{2i} \gtrless 0$. These models are superficially different from those considered in the analysis of the hybrid model. Appearances are deceiving. Both models are generated from underlying continuous latent variables and the analysis of one model readily applies to the analysis of the other.

Equations (1a)-(1c) apply to this case as well. As before, d_i is defined as the dichotomization of y^*_{2i}

[1] An important reference for such models is Amemiya (1975).

$$d_i = 1 \quad \text{iff} \quad y^*_{2i} \geq 0$$

$$d_i = 0 \quad \text{otherwise}$$

and define dummy variable a_i as the dichotomization of y^*_{1i}

(15) $\quad a_i = 1 \quad \text{iff} \quad y^*_{1i} \geq 0$

$\quad a_i = 0 \quad$ otherwise.

The argument of Section II may be applied to this case.

As in the case of the hybrid model, the Principle Assumption ($\gamma_2\beta_1 + \beta_2 = 0$) is a requirement for a meaningful statistical model to exist. Accordingly, the argument of Section II (i) of this paper applies to the multivariate probit model. The models of Ashford and Sowden (1970), Amemiya (1975), and Zellner and Lee (1965) satisfy this assumption since none of these papers considers structural shift in the equations (i.e., they assume that $\beta_1 = \beta_2 = 0$).

The identification procedure in Section II (ii) must be modified since no observations are available on y_{1i}. The analysis of identification of π^*_{21} and π^*_{22} is as before. But the analysis of equation (5a) must be modified. Two distinct cases are worth considering. First suppose that $\pi_{13} = 0$ so that there is no structural shift in the equations.

In this case, normalized parameters of equation (5a) may be estimated. That is, one may use probit analysis to estimate

$$\pi^*_{11} = \frac{\pi_{11}}{(\omega_{11})^{1/2}} \quad \text{and} \quad \pi^*_{12} = \frac{\pi_{12}}{(\omega_{11})^{1/2}} .$$

The correlation between V_{1i} and V_{2i} may also be estimated, even though both y^*_{1i} and y^*_{2i} are latent variables. This result is well known in the theory of tetrachoric correlation (Kendall and Stuart, Vol. II, 1967). To establish this result is is useful to recall that c_i is defined as

$$c_i = -(X_{1i}\pi^*_{21} + X_{2i}\pi^*_{22})$$

and that b_i can be defined as

$$b_i = -(X_{1i}\pi^*_{11} + X_{2i}\pi^*_{12}).$$

The probability of the events a_i and d_i can be written as

(16a) $P_{11}(i) = \text{Prob}(a_i = 1 \wedge d_i = 1) = F(-b_i, -c_i, \rho)$

(16b) $P_{01}(i) = \text{Prob}(a_i = 0 \wedge d_i = 1) = F(b_i, -c_i, -\rho)$

(16c) $P_{10}(i) = \text{Prob}(a_i = 1 \wedge d_i = 0) = F(-b_i, c_i, -\rho)$

(16d) $P_{00}(i) = \text{Prob}(a_i = 0 \wedge d_i = 0) = F(b_i, c_i, \rho)$

where $F(,)$ is a standarized bivariate normal cumulative distribution.[1] Substituting consistent estimators of b_i and c_i in place of the true values, these probabilities are solely a function of ρ, the correlation coefficient. The sample likelihood function may be maximized with respect to ρ to achieve a consistent estimator of that parameter. The appropriate likelihood function is

$$\mathcal{L}(\rho) = \prod_{i=1}^{I} [P_{11}(i)]^{d_i a_i} [P_{01}(i)]^{d_i(1-a_i)} [P_{10}(i)]^{(1-d_i)a_i} [P_{00}(i)]^{(1-a_i)(1-d_i)}$$

There are alternative minimum chi square estimators and modified minimum chi square estimators for this parameter discussed elsewhere (Amemiya (1975); Heckman (1976)). All of these estimators are consistent but not efficient since the information matrix for the reduced form coefficients is not block diagonal in ρ.

Next suppose that $\pi_{13} \neq 0$ so that there is structural shift in reduced form equation (5a). For this case, initial consistent estimators are also available. The conditional distribution of a_i given d_i may be written as

---

[1] Thus, $F(b_i, c_i, \rho) = \int_{-\infty}^{b_i} \int_{-\infty}^{c_i} h(V_1^*, V_2^*) dV_2^* dV_1^*$

where $V_1^* = V_1/\omega_{11}^{1/2}$.

$$\text{Prob}(a_i\ d_i) = \left[\frac{P_{11}(i)}{P_{1.}(i)}\right]^{a_i d_i} \left[\frac{P_{10}(i)}{P_{0.}(i)}\right]^{(a_i)(1-d_i)} \left[\frac{P_{01}(i)}{P_{1.}(i)}\right]^{(1-a_i)d_i} \left[\frac{P_{00}(i)}{P_{0.}(i)}\right]^{(1-a_i)(1-d_i)}$$

where $P_{0.}(i) = F(\infty, c_i)$ and $P_{1.}(i) = 1 - P_{0.}(i)$, and where b_i is replaced everywhere by b_i' defined by

$$b_i' = b_i - \pi_{13}^* d_i.$$

Since consistent estimators of c_i are available, these may be inserted as parameters in the appropriate conditional (on d_i) likelihood function. If that function is maximized with respect to ρ, π_{13}^*, π_{12}^* and π_{11}^* consistent estimators result. These estimators are not efficient since the full system information matrix is not block diagonal with respect to these parameters.

This analysis establishes that it is possible to estimate all of the normalized reduced form parameters: ρ, π_{11}^*, π_{12}^*, π_{13}^*, π_{21}^* and π_{22}^*. Under the identification hypothesis postulated in Section I, one can utilize equation (2) to solve for normalized structural parameters

$$(17) \qquad \alpha_1^* = \alpha_1/\omega_{11}^{1/2}\ ,\ \beta_1^* = \beta_1/\omega_{11}^{1/2}\ ,\ \gamma_1^{**} = \gamma_1^*/\omega_{11}^{1/2} = \gamma_1\left(\frac{\omega_{22}}{\omega_{11}}\right)^{1/2}$$

$$\alpha_2^* = \alpha_2/\omega_{22}^{1/2}\ ,\ \beta_2^* = \beta_2/\omega_{22}^{1/2}\ ,\ \gamma_2^{**} = \gamma_2^*\omega_{11}^{1/2} = \gamma_2\left(\frac{\omega_{22}}{\omega_{11}}\right)^{1/2}$$

Moreover, the argument presented in Section II(ii) on the estimation of structural covariances may be readily extended to show that it is possible to estimate

$$(18) \qquad \sigma_{11}^* = \sigma_{11}/\omega_{11}\ ,\ \sigma_{12}^{**} = \sigma_{12}^*\omega_{11}^{1/2} = \sigma_{12}\left(\frac{\omega_{11}}{\omega_{22}}\right)^{1/2} \text{ and } \sigma_{22}^{**} = \sigma_{22}^*\omega_{11} = \sigma_{22}\left(\frac{\omega_{11}}{\omega_{22}}\right).$$

This completes the analysis of parameter identification.

The analysis of the full information maximum likelihood estimators is straightforward and need not be belabored. The likelihood function is

$$\mathcal{L} = \prod_{i=1}^{I} [P_{00}(i)]^{(1-d_i)(1-a_i)} [P_{01}(i)]^{(1-d_i)a_i} [P_{10}(i)]^{d_i(1-a_i)} [P_{11}(i)]^{a_i d_i} .$$

The function may be maximized with respect to the parameters listed in equations (17) and (18). As in the hybrid model, in a finite sample there is some probability that maximum likelihood estimators fail to exist but this probability becomes arbitrarily small as sample size becomes arbitrarily large. The maximum likelihood estimators are consistent, asymptotically normal and efficient.[1]

IV. The formulation of Simultaneous Equation Models with Discrete Endogenous Variables[2]

In this section, the models developed in this paper are contrasted with previous work on discrete models with jointly endogenous variables by Goodman (1970) and Nerlove and Press (1976). These models deal with purely discrete random variables. Accordingly, the appropriate comparison is one between that work and the models of Section III although an important topic to be considered is the issue of generalizing purely discrete models to accommodate both discrete and continuous endogenous variables.

It is argued here that the log linear model of Goodman and Nerlove and Press is an inadequate scheme for formulating the simultaneous equation model required in econometrics. This is so for two reasons. First, the log linear model is designed to simplify the estimation of conditional probabilities

---

[1] The same sort of existence conditions are required as those presented in Section II. With probability one, maximum likelihood estimators exist in large sample samples.

[2] I have greatly benefited from discussions with Marc Nerlove on the material in this section.

whereas the simultaneous equation model is intrinsically on unconditional representation of random variables. Estimators of conditional probabilities in the log linear model have the same interpretation as direct least squares estimators in classical simultaneous equation theory which are also conditional probability statements. Both estimators confound true structural parameters with elements of the equation error covariance structure. Second, the log linear model does not readily generalize to accommodate continuous endogenous variables while the multivariate normal structure can easily be modified to do so, as has been shown.

To fix ideas, consider a log linear model for a two equation system comparable to the model of Section III. Nerlove and Press (1976, p. 51) explicitly consider a log linear model for this case. Altering their notation to conform with the notation of Section III and suppressing subscript i, the log linear analogue of equations (16a)-(16d) is

$$\ell n \text{ Prob } (a = 0 \wedge d = 0) = \ell n\, P_{00} = \alpha_0 + \alpha_1 + \beta + \mu$$

$$\ell n \text{ Prob } (a = 0 \wedge d = 1) = \ell n\, P_{01} = \alpha_0 - \alpha_1 - \beta + \mu$$

$$\ell n \text{ Prob } (a = 1 \wedge d = 0) = \ell n\, P_{10} = -\alpha_0 + \alpha_1 - \beta + \mu$$

$$\ell n \text{ Prob } (a = 1 \wedge d = 1) = \ell n\, P_{11} = -\alpha_0 - \alpha_1 + \beta + \mu$$

where $\mu = -\ell n\, [\exp(\alpha_0 + \alpha_1 + \beta) + \exp(\alpha_0 - \alpha_1 - \beta) + \exp(-\alpha_0 + \alpha_1 - \beta) + \exp(-\alpha_0 - \alpha_1 + \beta)]$ and α_0, α_1 and β may be parameterized as functions of exogenous variables.

The marginal probability of the event $a = 0$ in the log linear model is

$$\text{(19)} \qquad \text{Prob } (a = 0) = \frac{\exp(\alpha_0 + \alpha_1 + \beta) + \exp(\alpha_0 - \alpha_1 - \beta)}{\exp(-\mu)}$$

This expression is to be compared with the corresponding probability given in Section III for the normal model with structural shift

(20) $$\text{Prob } (a = 0) = \sum_{d=0,1} [F(b - \pi^*_{13}d, -c)]^d \; [F(b - \pi^*_{13}d, c)]^{1-d}$$

Recall that F is a function of correlation parameter ρ.

When $\beta = 0$, the log linear probability model collapses to a simple logit model

$$\text{Prob } (a = 0) = \frac{1}{1 + e^{-2\alpha_0}} .$$

With $\rho = 0$, the normal model becomes a probit model with structural shift

$$\text{Prob } (a = 0) = \sum_{d=0,1} [\Phi(b - \pi^*_{13}d)\Phi(-c)]^d \; [\Phi(b)\Phi(c)]^{1-d}$$

where $\Phi(t)$ is the standard univariate cumulative density ($= F(\infty, t)$). Finally, note that if there is no structural shift ($\pi^*_{13} = 0$), as well as no covariance ($\rho=0$)

$$\text{Prob } (a = 0) = \Phi(b)$$

so that a simple probit model arises.

Further note that in the log linear model, the conditional probability that $a = 0$ given d may be written as

(21) $$\text{Prob } (a = 0|d) = \frac{1}{1 + e^{-2\alpha_0} \, e^{(2\beta d - 2\beta(1-d))}}$$

The simplicity of this representation is the basis for the popularity of the log linear model. The comparable expression for the normal model is

$$\text{Prob } (a = 0|d) = \left[\frac{F(b - \pi^*_{13}d, -c)}{F(\infty, -c)}\right]^d \cdot \left[\frac{F(b - \pi^*_{13}d, c)}{F(\infty, c)}\right]^{1-d}$$

Note that α_0 and b, α_1 and c play similar roles in the model in which they appear. The important point to note, however, is that β and ρ and π^*_{13} play similar roles. In the normal model the probability that a = 0 given d depends on d for two conceptually distinct reasons: one related to the true structure of the model ($\pi^*_{13} \neq 0$) and the other due to covariance in latent errors ($\rho \neq 0$). In the log linear model, these effects are indistinguishable. Thus the log linear parameter of association, β, corresponds to two distinct parameters in the normal model ρ and π^*_{13}.

As long as one only seeks to estimate empirical relationships among endogenous variables, this issue may be ignored. Suppose, however, that one seeks to utilize fitted econometric relationships to answer policy questions. Then, as Haavelmo (1944) has stressed, it is important to identify structural parameters. A simple example will fix ideas.

Let a = 1 if a family has a child and let a = 0 if it does not. Let d = 1 if the family uses birth control and d = 0 otherwise. It may happen that because of unmeasured taste and knowledge factors, families more likely to contracept are more likely not to have a child. This effect would be captured in a normal model by setting values of the parameter $\rho < 0$. There is, however, a second effect. For obvious structural reasons families on birth control will have fewer children ($\pi^*_{13} < 0$ in the normal model). For either reason $\beta < 0$ in the log linear model.

Suppose that the government forces all families to contracept, say through a sterilization program. The normal model would permit identification of the effect of this policy shift through use of π^*_{13}. Thus the predicted change in the probability of a couple having no children as a result of the policy would be derived from equation (20) as

$$\Delta P(a = 0) = \sum_{d=0,1} [F(b - \pi^*_{13}, -c)]^d \, [F(b - \pi^*_{13}, c)]^{1-d}$$

$$- \sum_{d=0,1} [F(b - d\pi^*_{13}, -c)]^d \, [F(b - d\pi^*_{13}, c)]^{1-d} \; \underline{1/}$$

Notice that if $\pi^*_{13} = 0$, there would be no effect predicted for the policy, whether or not $\rho = 0$.

The estimate of the policy effect from the log linear model would be given by subtracting the conditional probability that $a = 0$ given $d = 1$ (given in equation (21)) from the marginal probability (given in equation (19)),

$$\Delta P = \text{Prob}\,(a = 0 | d = 1) - \text{Prob}\,(a = 0).$$

Since it is not possible to disentangle purely statistical association from purely causal association in the log linear model, it is not possible to identify meaningful structural parameters interpretable within the classical simultaneous equation framework. If one were to follow Amemiya's (1975) suggestion and use the log linear model to approximate a multivariate normal model, misleading predictions of policy effects might occur. Indeed, if $\pi^*_{13} = 0$, but $\rho < 0$ in the normal model, the log linear model would predict an effect of the program ($\beta < 0$) even when none in fact would occur. Conversely, if $\pi^*_{13} > 0$ and $\rho < 0$, estimated independence of events in the log linear model ($\beta = 0$) would lead to incorrect forecasts of policy effects. Note, however, that if dummy variables are defined as indicators of latent variables that cross thresholds, and not as structural shift parameters ($\pi^*_{13} = 0$), Amemiya's suggestion is appropriate, and the log linear model may be used to approximate a multivariate normal model. The parameters β and ρ then play

---

[1] The difference between the first term and the second term is that d is set to one <u>inside</u> the brackets signifying that everyone in the population is forced to use birth control.

the same role in their respective models.

A second, more minor point, concerns the computational intractability of the multivariate logistic model when both continuous and discrete endogenous variables are present. The reader is invited to differentiate the cumulative distribution function of the multivariate distribution to confirm this point.[1] Further, as Amemiya (1975) has noted the multivariate logistic distribution arbitrarily fixes the correlation structure among the random variables, a highly unattractive feature.[2] For both reasons, the log linear structure does not generalize to accommodate continuous and discrete endogenous variables in a simultaneous equation system.

---

[1] That distribution is

$$f(Y_1, \ldots, Y_n) = \frac{1}{1 + \sum_{i=1}^{n} \exp(-Y_i)}$$

[2] Note, however, that it is misleading to think that the cumulative logistic distribution introduced in the previous footnote plays the same role in the log linear model as the multivariate normal plays in the models of Sections II and III of this paper. In the multivariate case, the log linear probabilities are not dichotomizations of underlying continuous latent logistic random variables. Thus, the transition from the log linear model to the multivariate normal model involves more than a convenient choice of a joint distribution for the latent variables introduced in Section I of this paper. Moreover, if one were to alter the log linear model to incorporate a structural shift term in α_0 that would play the same role as parameter π^*_{13} in the normal model, that term would not be identified. This result is simply a restatement of the argument in the text.

Summary

This paper develops a class of econometric models for simultaneous equation systems with dummy endogenous variables. These models are based on the pioneering work of Pearson (1900) on dichotomized variables. The general model presented here includes simultaneous probit and ordinary simultaneous equation models as special cases. Dummy endogenous variables are introduced in two conceptually distinct roles: (1) as proxy variables for unmeasured latent variables crossing thresholds, as in the classical quantal response model (Amemiya, 1975) and (2) as direct shifters of structural behavioral equations formulated in terms of latent variables. This distinction is shown to be quite important in the formulation and interpretation of the econometric models developed here. Maximum likelihood and alternative estimators are discussed. Conditions for the existence of a meaningful statistical model are derived.

The models presented here have already been put to practical use (see Edwards (1975) and Heckman (1975, 1977)). They are computationally tractable and in the applications cited have led to new interpretations of old evidence.

The models formulated here are compared with alternative models by Goodman (1970) and Nerlove and Press (1976). It is shown that the log linear model does not provide a natural framework for formulating the simultaneous equation model of econometrics whereas the models presented in this paper do.

APPENDIX A

DERIVATION OF THE DENSITY $f(y_{1i}, d_i)$

In this appendix, there is a brief discussion of the derivation of density $f(y_{1i}, d_i)$ that is used in the text. This discussion is useful because random variables that are the sum of underlying continuous and discrete random variables are unfamiliar in econometrics. The joint density of V_{1i}, V^*_{2i} $(= V_{2i}/\omega_{22}^{1/2})$ is given by $h(V_{1i}, V^*_{2i})$, a bivariate normal density. The joint density of V_{1i} and d_i is

$$e(V_{1i}, d_i) = \left[\int_{c_i}^{\infty} h(V_{1i}, V^*_{2i})dV^*_{2i}\right]^{d_i} \left[\int_{-\infty}^{c_i} h(V_{1i}, V^*_{2i})dV^*_{2i}\right]^{1-d_i}$$

where $c_i = -(X_{1i}\,\pi^*_{21} + X_{2i}\,\pi^*_{22})$.

Define a random variable $Z_i = V_{1i} + \pi_{13}\,d_i$. The joint density of Z_i, d_i is simply

$$e(Z_i - \pi_{13}d_i, d_i) = \left[\int_{c_i}^{\infty} h(Z_i - \pi_{13}d_i, V^*_{2i})dV^*_{2i}\right]^{d_i} \cdot \left[\int_{-\infty}^{c_i} h(Z_i - \pi_{13}d_i, V^*_{2i})dV^*_{2i}\right]^{1-d_i}.$$

Substitute $y_{1i} - X_{1i}\pi_{11} - X_{2i}\pi_{12}$ for Z_i (noting that the Jacobian is unity) to reach the density $f(y_{1i}, d_i)$ used in the text.

The measured density for Z_i is

$$e_1(Z_i) = \int_{c_i}^{\infty} h(Z_i - \pi_{13},\ V^*_{2i})dV^*_{2i} + \int_{-\infty}^{c_i} h(Z_i,\ V^*_{2i})dV^*_{2i} .$$

Using conditional normal theory one can write

$$h(Z_i - \pi_{13},\ V^*_{2i}) = \phi(Z_i - \pi_{13}|V^*_{2i})h_2(V^*_{2i})$$

and also, clearly

$$h(Z_i,\ V^*_{2i}) = \phi(Z_i|V^*_{2i})h_2(V^*_{2i})$$

where $\phi(|)$ is a conditional normal density. Then obviously $e_1(Z_i)$ is a proper density since

$$\int_{-\infty}^{\infty} e_1(Z_i)dZ_i = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(Z_i|V^*_{2i})h_2(V^*_{2i})dZ_i dV^*_{2i}$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(Z_i - \pi_{13}|V^*_{2i})h_2(V^*_{2i})dZ_i dV^*_{2i}$$

$$= 1$$

since $\int_{-\infty}^{\infty} \phi(Z_i|V^*_{2i})dZ_i = 1.$

APPENDIX B

MULTIVARIATE EXTENSIONS

Multivariate extensions of the models of Sections II and III in the text are presented in this appendix. Let Y_i^* be a row vector of G jointly endogenous latent random variables, some of which may be observed. Let Γ be a G x G nonsingular matrix with unit diagonal elements. X_i is a 1 x K row vector of bounded exogenous variables. A is a K x G coefficient matrix for the exogenous variables. "d_i" is a 1 x C vector of dummy variables ($C \leq G$). Only $C_1 (\leq C)$ of these dummy variables act as shifters of the structural equations. Associated with the C_1 shift dummy variables is a C_1 x G coefficient matrix B. U_i is a 1 x G vector of disturbances for observation i, i = 1, . . . , I.

The structural equation system may be written as

$$Y_i^* \Gamma + X_i A + d_i B = U_i \tag{B-1}$$

where

$$E(U_i) = 0 \quad E(U_i' U_i) = \Sigma$$

and

$$E(U_i' U_j) = 0 \quad i \neq j.$$

U_i is assumed to be a multivariate normal random vector. Σ is positive definite. For notational convenience, suppose that the first C latent variables activate the C dummy variables

$$d_{ij} = 1 \quad \text{iff} \quad Y^*_{ij} \geq 0$$

$$d_{ij} = 0 \quad \text{otherwise, } j = 1, \ldots, C$$

and that the first C continuous variables in Y^*_i are unobserved while the remaining G-C variables are observed.[1] Finally, note that $C \geq C_1$, and assume that the first C_1 latent variables generate the C_1 shift dummies.

The semi-reduced form for the system is

(B-2) $$Y^*_i = X_i\pi + d_i\tilde{\pi} + V_i$$

where

$$\pi = -A\Gamma^{-1}, \tilde{\pi} = -B\Gamma^{-1} \text{ and } V_i = U_i\Gamma^{-1}.$$

It is convenient to work with reduced form variance normalized versions of these coefficients. Define Ω as

$$\Omega = E(V_i'V_i) = (\Gamma^{-1})' \Sigma \Gamma^{-1}.$$

Diagonal matrix D is defined by

$$D^2 = \text{diag } \Omega.$$

D displays the population standard deviation for each element of V_i. Partition D, and define D^* as

$$D^* = \begin{pmatrix} D_C & 0 \\ 0 & I_{G-C} \end{pmatrix}$$

---

[1] Note that case 2 in Part I of the text is excluded by the assumption that $C_1 < C$ and that only unobserved latent variables generate structural shift. The model can readily be generalized to include this case.

where D_C is a C xC submatrix of the first C diagonal elements of D, and I_{G-C} is a (G-C) x (G-C) identity matrix.

Postmultiply equation (B-2) by $(D^*)^{-1}$ to reach

$$Y_i^{**} = X_i\pi^* + d_i\tilde{\pi}^* + V_i^* \tag{B-3}$$

$$E\left((V_i^*)' \; V_i^*\right) = \Omega^* = (D^*)^{-1}\,\Omega(D^*)^{-1}.$$

This operation normalizes the first C semi-reduced form equations to have a unit variance disturbance, on obvious generalization of the procedure utilized in Section II in the text. Note that

$$\pi^* = \pi(D^*)^{-1}, \;\; \tilde{\pi}^* = \tilde{\pi}(D^*)^{-1}.$$

In the notation of this appendix, the Principle Assumption in the text requires that the first C_1 rows of the first C_1 columns of $\tilde{\pi}^*$ should vanish. Thus the Principle Assumption in this model becomes

$$\tilde{\pi}_{ij}^* = 0 \quad i, j = 1, \ldots, C_1.$$

Assuming that the reduced form model is of full rank, the coefficients in the first C_1 columns of π^* may be estimated by applying probit analysis to each equation. If $C > C_1$, the coefficients in the next $C-C_1$ columns of π^* and $\tilde{\pi}^*$ may be estimated by applying the methods of Section III to each equation. Assuming G > C, the coefficients of the final G-C columns of π^* and $\tilde{\pi}^*$ may be consistently estimated by applying the methods of Section II to each equation. Precisely the same type of argument offered in Section II establishes that all elements of Ω^* are estimable.

Now consider the estimation of structural coefficients. Assume

that through exclusion restrictions all nonzero coefficients in equation (B-1) could be identified if Y_i^* were observable.[1] Clearly, in the transition to equation system (B-1) with some continuous variables unobserved, the same regression coefficients that can be identified in the previous case can be identified in the current model if they are suitably re-normalized. To obtain the required normalization, rewrite equation (B-1) in terms of normalized latent variables

$$Y_i^*(D^*)^{-1} D^*\Gamma + X_i A + d_i B = U_i. \tag{B-1}$$

For the normalization implicit in the choice of Γ, it is natural to post-multiply this equation by $(D^*)^{-1}$ to reach

$$(Y_i^*(D^*)^{-1})(D^*\Gamma(D^*)^{-1}) + X_i A(D^*)^{-1} + d_i B(D^*)^{-1} = U_i(D^*)^{-1}. \tag{B-1'}$$

Clearly, then, one can identify the following parameters:

$$\Gamma^* = D^*\Gamma(D^*)^{-1}, A^* = A(D^*)^{-1}, B^* = B(D^*)^{-1}. \tag{B-4}$$

Finally, it is clear that one can identify the following parameters of the structural covariance terms

$$\Sigma^* = (D^*)^{-1} \Sigma (D^*)^{-1}. \tag{B-5}$$

This completes the analysis of parameter identification.

The likelihood function for the model may be generated from the density for random variables d_i and Y_i^{ℓ} where Y_i^{ℓ} is the 1 X (G-C) subvector of Y_i^* corresponding to the observed continuous variables. That density is defined next. Let ψ_i be defined as

---

[1] The restriction to exclusion restrictions is overly stringent. Identification through use of covariance restrictions is also permitted so long as such restrictions can be imposed on Ω^*.

$$\psi_i = X_i\pi^* + d_i\pi^*.$$

Partition ψ_i into $\psi_{i,C}$ and $\psi_{i,G-C}$, i.e., $\psi_i = (\psi_{i,C}, \psi_{i,G-C})$. Then the density for d_i, Y_i^{ℓ}, $g(d_i, Y_i^{\ell})$ is given by

$$g(d_i, Y_i^{\ell}) = F_2(\psi_{i,C} * [2d_i - \iota], Y_i^{\ell} - \psi_{i,G-C}; \Sigma^* * [2d_i - \iota][2d_i - \iota]')$$

where * denotes a Hademard product, ι is a $1 \times C$ vector of "1's," and 2 is the number "2". F_2 is the derivative of the cumulative distribution for the multivariate normal with respect to the final G-C elements of $U_i^* = U_i(D^*)^{-1}$, i.e.,

$$F_2(U_i^*; \Sigma^*) \propto \int_{-\infty}^{u_{i1}} \cdots \int_{-\infty}^{u_{iC}} |\Sigma^*|^{-1/2} \exp -1/2(U_i^* (\Sigma^*)^{-1} U_i^{*\prime}) dU_{iC}^*$$

where $U_i^* = (U_{iC}^*, U_{iG-C}^*)$.

The sample likelihood function is

$$\mathcal{L} = \prod_{i=1}^{I} g(d_i, Y_i^{\ell})$$

which is to be maximized with respect to the terms in equations (B-4) and (B-5). As in the text, the identification analysis produces initial consistent estimators to use as starting values. In large samples, maximum likelihood estimators exist, and are consistent and asymptotically efficient.

One final multivariate extension is worth noting. The models developed thus far are for unordered dichotomous variables. In some cases, dummy endogenous variables may be naturally ordered. For example, in an analysis of the effects of legislation on the income of blacks, one might distinguish existing laws by their "strength" and a natural ordering would

exist. One simple way to generate such ordered dichotomies is to polytomize a single latent continuous random variable. Thus, each element of d_i, say $d_{ic'}$, might be replaced by a vector of dummy variables, with a typical element $d_{ic'}^{(j)}$ defined as

$$d_{ic'}^{(j)} = 1 \quad \text{iff} \quad \phi_j^{c'} \leq Y_{ic'}^* \leq \phi_j^{c'}$$

$$= 0 \quad \text{otherwise, } j = 1, \ldots, J_{c'}$$

where the categories are mutually exclusive, and the $\phi_j^{c'}$, $j = 1, I_{c'}$ are a set of estimable constants (fixing $\phi_J = \infty$ and $\phi_o = -\infty$).

Each of the C dummy variables may be polytomized in this fashion.[1]

---

[1]This procedure for generating ordered dichotomous variables is discussed in more specialized cases by Johnson (1972) and Amemiya (1975).

REFERENCES

Amemiya, T. "Estimation of the Parameters of a Discontinuous Density Function," Working Paper No. 50, September, 1974, The Economic Series, Stanford University.

_____. "Qualitative Response Models," Annals of Economic and Social Measurement 4, No. 3 (Summer 1975): 363-372.

Ashford, J. R. and R. R. Sowden. "Multivariate Probit Analysis," Biometrics 26 (1970): 535-546.

Edwards, L. "Measuring the Effectiveness of a Law When It Is Endogenous," Unpublished NBER Paper, 1975, Forthcoming in The Journal of Law and Economics.

Goodman, L. "The Multivariate Analysis of Qualitative Data: Interactions among Multiple Classifications," Journal of the American Statistical Association 65, pp. 226-256.

Haavelmo, T. "The Probability Approach in Econometrics," Supplement to Econometrica 12 (1944).

Haberman, S. The Analysis of Frequency Data, University of Chicago Press, 1974.

Heckman, James. "Sample Selection Bias as a Specification Error," Rand Corporation, R-1984 (April 1976).

_____. "Simultaneous Equation Models with Both Continuous and Discrete Endogenous Variables with and without Structural Shift in the Equations," University of Chicago (July 1975).

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

Johnson, T. "Qualitative and Limited Dependent Variables in Economic Relationships," Econometrica 40, pp. 455-462.

Kendall, M. and J. Stewart. The Advanced Theory of Statistics, Vol. 2. New York: Hafner Publishing Company, 1967.

LeCam, L. "On Some Asymptotic Properties of Maximum Likelihood Estimators and Related Bayes' Estimates," University of California Publications in Statistics 1, No. 11, University of California Press, Berkeley, 1953.

Lord, F. and M. Novick. Statistical Theories of Mental Test Scores, Addison-Wesley Publishing Company, 1968.

McFadden, D. "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), Frontiers in Econometrics, Academic Press, 1974.

_____. "Quantal Choice Analysis: A Survey," Annals of Economic and Social Measurement 5, No. 4 (December 1976): 363-390.

Nerlove, M. and J. Press. "Multivariate Log Linear Probability Models for the Analysis of Qualitative Data," Discussion Paper #1, Center for Statistics and Probability, June 1976.

Olsen, R. "An Econometric Model of Family Labor Supply," Unpublished Ph.D. Dissertation, University of Chicago, 1977.

Pearson, K. "On the Correlation of Characters not Quantitatively Measurable," Royal Society Philosophical Transactions, Series A, Vol. 195, 1900, pp, 1-47.

Sant, D. "A Comment on Two Stage Least Squares and Instrumental Variable Estimators," Unpublished Paper, Princeton, July 1975.

Tate, R. F. "The Theory of Correlation between Two Continuous Variables when One is Dichotomized," Biometrika 42, (1955): 205-216.

Telser, L. "Iterative Estimation of a Set of Linear Regression Equations," Journal of the American Statistical Association 59 (September 1964): 845-862.

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

Zellner, A. and T. Lee. "Joint Estimation of Relationships Involving Discrete Random Variables," Econometrica 33, pp. 382-394.