Dummy Endogenous Variables in a Simultaneous Equation System

Econometrica, Vol. 46, No. 4 (Jul., 1978), 931-959.

James J. Heckman

THE UNIVERSITY OF CHICAGO

Econ 312, Spring 2019

1. A GENERAL MODEL FOR THE TWO EQUATION CASE

- Pair of simultaneous equations for continuous latent random variables y_{1i}^* and $\mathrm{y}_{2\mathrm{i}}^*$,

(1a) $y_{1i}^* = X_{1i}\alpha_1 + d_i\beta_1 + y_{2i}^*\lambda_1 + U_{1i}$,

(2a) $y_{2i}^* = X_{2i}\alpha_2 + d_i\beta_2 + y_{1i}^*\lambda_2 + U_{2i}$,

where dummy variable d_i is defined by

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

$d_i = 0$ otherwise,

and

$$E\left(U_{ji}\right) = 0, \qquad E\left(U_{ji}^2\right) = \sigma_{jj}, \qquad E(U_{1i}U_{2i}) = \sigma_{12}, \qquad j = 1,2; i = 1, \dots, I.$$

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

"X_{1i}" and "X_{2i}" are, respectively, $1 \times K_1$ and $1 \times K_2$ row vectors of bounded exogenous variables.

- Equations (1a) and (1b) are identified under standard conditions if $\beta_1 = \beta_2 = 0$ and both y^*_{1i} and y^*_{2i} are observed for each of the I observations.

- 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.

- 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 in inequality (1c), but this can be readily amended t 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 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.

- 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.

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

  - **CASE l:** *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 ($\beta_1 = \beta_2 = 0$).

  - **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 $y_{1i}^*$and $y_{2i}^*$are observed (i.e., one knows whether or not the latent variables have crossed a threshold). The notation of equations (1a)-(1b) 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 [3], Amemiya [2], and Zellner and Lee [30].

- **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 ($\beta_1 = \beta_2 = 0$).

- **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 ($\beta_1 = \beta_2 = 0$).

- **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.

2. 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.

- Consider identification only; Heckman (1978) for additional discussion.

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

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

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

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

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

where

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

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

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

- 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:

$$E(V_{1i}) = 0, \qquad E(V_{2i}) = 0,$$
$$E\left(V_{1i}^2\right) = \omega_{1i}, \qquad E(V_{1i}V_{2i}) = \omega_{12}, \qquad E\left(V_{2i}^2\right) = \omega_{22}.$$

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*

- 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 principal 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 $y_{2i}^* \geqslant 0$ given $d_i = 1$ must be unity, so that one may write

$$\Pr(V_{2i} > l_i | d_i = 1) = 1$$

where the symbols l_i and l_i' are defined by $l_i = -(X_{1i}\pi_{21} + X_{22} + \pi_{23})$ and $l_i' = l_i + \pi_{23}$.

- Alternatively, one may write this condition as

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

and obviously

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

- Using similar reasoning, one can conclude that

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

and

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

- 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 $d_i = 1$, 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

$$
\begin{aligned}
y_{1i}^* &= X_{1i}a_1 + y_{2i}^*\gamma_1 + U_{1i}, \\
y_{2i}^* &= X_{2i}a_2 + y_{1i}^*\gamma_2 + U_{2i}, \\
d_i &= 1 \text{ iff } \mathrm{y}_{2\mathrm{i}}^* > 0, \\
d_i &= 0 \text{ otherwise.}
\end{aligned}
$$

- 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} and y_{1i}^* in the system given above, one concludes that

$$\begin{aligned} y_{1i} &= X_{1i}a_1 + d_1\beta_1 + y_{2i}^*\gamma_1 + U_{1i}, \\ y_{2i}^* &= X_{2i}a_2 + (y_{1i} - d_i\beta_1)\gamma_2 + U_{21}. \end{aligned}$$

- Invoking the principal assumption, one reaches equations (1a)-(1c) including 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}^*.

- The principal assumption essentially requires that the latent variable y_{1i}^* and not the measured variable y_{1i}^* appears in the second structural equation.

- It is possible to use the latent variable in the second equation because β_1 can be estimated as will be shown.

- It is important to note that the principal assumption does not rule out structural shift in equations (1a) and (1b).

- It simply restricts the nature of the shift. However, the principal assumption does exclude any structural shift in the reduced form equation that determines the probability of shift (equation (3b)).