Limited Information Estimation of Simultaneous Equations with Indicator Endogenous Variables*

PHOEBUS J. DHRYMES
Columbia University

Original July 2004; this version, late November 2004

Abstract

1 Introduction

The subject of "dummy" endogenous variables has received considerable attention over the last thirty years or so. Invariably the problem was addressed in a simultaneous equations model and its estimation was carried out by (full information) maximum likelihood methods, assuming joint normality of the errors. Generally, this leads to rather cumbersome procedures, both to carry out and to comprehend.

This paper deals with the problem in terms of limited information maximum likelihood or 2SLS-like methods. It is much easier to comprehend and implement, and allows more flexibility in model specification.

Before we examine the problem of indicator endogenous variables we begin with the standard general linear structural econometric model (GLSEM),

$$y_{t1}^* = y_{t2}^* \beta_{21}^* + x_{t1\cdot} \gamma_{\cdot 1}^* + u_{t1}^* \tag{1}$$

$$y_{t2}^* = y_{t1}^* \beta_{12}^* + x_{t2\cdot} \gamma_{\cdot 2}^* + u_{t2}^*, \tag{2}$$

where the y^* are the **endogenous variables**, $x_{t1\cdot}$, $x_{t2\cdot}$ are the **predetermined or exogenous variables**, and the u^* 's are the structural errors, which may or may not be correlated. Identification requires that some variables in the vector $x_{t1\cdot}$ **not appear in** $x_{t2\cdot}$ **and vice versa**.

By substituting the first equation into the second, we can write

$$y_{t2}^* = z_{t\cdot} \gamma + u_{t2}, \tag{3}$$

where now

$$u_{t2} = \frac{1}{1 - \beta_{12}^* \beta_{21}^*} [u_{t2}^* + \beta_{12}^* u_{t1}^*], \tag{4}$$

and $z_{t\cdot}$ contains the union of the variables in $x_{t1\cdot}$ and $x_{t2\cdot}$. Notice that if we consider Eqs. (1) and (3) as a complete system of simultaneous equations, we have a semi-decomposable system; it is not **simply** decomposable because u_{t2} and u_{t1}^* are **necessarily** correlated. This system forms the basis of innumerable empirical papers on topics involving, e.g. wages and education, as well as other labor-related topics. We re-iterate that identification for this model requires that the vector $z_{t\cdot}$ contain variables beyond those contained in x_{t1} .

Remark 1. Notice that if Eqs. (1) and (2) are the equations of interest and if they constitute a complete simultaneous equations system, identification requires that some variables in x_{t1} are not in x_{t2} , and vice-versa, as noted earlier. We have seen that such a system has a modified form as in Eqs. (1) and (3), in which the parameters of Eq. (1) are the parameters of interest and **are identified**, while the parameters of Eq. (2) appear in Eq. (3) as linear combinations involving those in Eq. (1). This is the typical Limited Information context in which only the parameters of Eq. (1) are of primary interest.

Given Eqs. (1) and (2), they can be reduced to Eqs. (1) and (3), **but the latter can, just as well, be specified** *ab initio* **as the appropriate model**.

We propose to use this model in order to solve the problem of indicator endogenous variables in a limited information context. In particular, we show that the standard simultaneous equations model in Eqs. (1) and (3) above, and the indicator endogenous variable model are intimately related, the first being appropriate when y^*_{t2} is **fully observed**, and the second when y^*_{t2} is **censored**.

2 Specification

Consider the model

$$y_{t1} = y_{t2}\beta_{21} + x_{t\cdot}\gamma_{\cdot 1} + u_t \tag{5}$$

$$y_{t2} = z_{t\cdot}\gamma_{\cdot 2} + v_t, \tag{6}$$

where, for identification, $z_{t\cdot}$ (essentially) contains some variables not contained in $x_{t\cdot}$. The variables u and v are defined on the probability space (Ω , $\mathcal{A}$, $\mathcal{P}$) and moreover

$$(u_t,\ v_t)' \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \quad \sigma_{12} \neq 0. \tag{7}$$

Remark 2. As we develop the model we shall determine that σ_{22} cannot be separately identified, and we shall impose the normalization $\sigma_{22} = 1$.

If y_{t2} were **fully observable**, we would simply estimate $\gamma_{\cdot 2}$ by least squares, write $y_{t2} = z_{t\cdot}\tilde{\gamma}_{\cdot 2} + \tilde{v}_t$, substitute that for y_{t2} in the first equation, and then apply least squares, using $\tilde{y}_{t2} = z_{t\cdot}\tilde{\gamma}_{\cdot 2}$, in lieu of y_{t2} . Thus, we obtain the limited information (2SLS) estimator of the **structural** parameters in the first equation.

We shall now show that, *mutatis mutandis*, the same procedure will yield the limited information estimator of the structural parameters of the first equation, **when** y_{t2} **is censored**.

3 Estimation of Parameters when an Endogenous Variable is Censored

The difficulty with censored models arises from the fact that, in our case, the magnitude of y_{t2} is not directly observed; rather what is recorded (observed) is whether the event $y_{t2} \geq 0$ has occurred or **failed to occur**. To this end, we **modify** the model above, and simplify notation by writing

$$y_t = x_{t\cdot}\gamma_{\cdot 1} + \beta_{21}\delta(t) + u_t \tag{8}$$

$$y_{t2} = z_{t\cdot}\gamma_{\cdot 2} + v_t, \tag{9}$$

and note that, due to the censoring of y_{t2}, what is actually observed in the second equation is

$$\delta(t) = 1, \quad \textbf{iff} \ \ y_{t2} \geq 0\,, \quad \delta(t) = 0, \quad \textbf{iff} \ \ y_{t2} < 0\,. \tag{10}$$

Without loss of generality, make the change in variable

$$(\epsilon_t, v_t)' = \begin{bmatrix} 1 & -\frac{\sigma_{12}}{\sigma_{22}} \\ 0 & 1 \end{bmatrix} (u_t, v_t)'. \tag{11}$$

Notice that by construction ϵ_t and v_t are **uncorrelated**, and if the u's and v's are **jointly normal**, then ϵ_t and v_t are also jointly normal and thus **mutually independent**. In fact, we determine that

$$\mathrm{Cov}[(\epsilon_t, v_t)'] = \begin{bmatrix} 1 & -\frac{\sigma_{12}}{\sigma_{22}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & -\frac{\sigma_{12}}{\sigma_{22}} \\ 0 & 1 \end{bmatrix}' = \begin{bmatrix} \sigma_{11} - (\sigma_{12}^2/\sigma_{22}) & 0 \\ 0 & \sigma_{22} \end{bmatrix}. \tag{12}$$

Because the Jacobean of the transformation above is unity we may, without loss of generality, write the model of Eqs. (8) and (9) as

$$y_t = x_{t\cdot}\gamma_{\cdot 1} + \beta_{21}\delta(t) + \frac{\sigma_{12}}{\sigma_{22}} v_t + \epsilon_t \tag{13}$$

$$y_{t2} = z_{t\cdot}\gamma_{\cdot 2} + v_t; \tag{14}$$

moreover, y_t can be observed only under two regimes: for $y_{t2} \geq 0$, corresponding to $\delta(t) = 1$, or for $y_{t2} < 0$, corresponding to $\delta(t) = 0$. Hence, under the two regimes the observations obey, respectively,

$$E(y_t|x_{t\cdot}, v_t \geq -z_{t\cdot}\gamma_{\cdot 2}) = \beta_{21}E(\delta(t)|v_{t\cdot} \geq -\gamma_{\cdot 2}) + x_{t\cdot}\gamma_{\cdot 1} + \frac{\sigma_{12}}{\sigma_{22}} E(v_t|v(t) \geq -z_{t\cdot}\gamma_{\cdot 2}))$$

$$E(y_t|x_{t\cdot}, v_t < -z_{t\cdot}\gamma_{\cdot 2}) = \beta_{21}E(\delta(t)|v_{t\cdot} < -\gamma_{\cdot 2}) + x_{t\cdot}\gamma_{\cdot 1} + \frac{\sigma_{12}}{\sigma_{22}} E(v_t|v(t) < -z_{t\cdot}\gamma_{\cdot 2}) \tag{15}$$

The novel feature here is to obtain the **mean of the truncated** distribution of v_t, i.e. to obtain

$$I_{t1} = E(v_t|v_t \geq -z_{t\cdot}\gamma_{\cdot 2}), \quad \text{and} \quad I_{t2} = E(v_t|v_t < -z_{t\cdot}\gamma_{\cdot 2}). \tag{16}$$

Carrying out the integrations we find

$$I_{t1} = \frac{f(z_{t\cdot}\gamma_{\cdot 2})}{F(z_{t\cdot}\gamma_{\cdot 2})}, \quad I_{t2} = -\frac{f(z_{t\cdot}\gamma_{\cdot 2})}{1 - F(z_{t\cdot}\gamma_{\cdot 2})}, \tag{17}$$

where f, F are, respectively, the pdf and cdf of a normal variable with mean zero and variance σ_{22}. We notice, however, that

$$f(z_{t\cdot}\gamma_{\cdot 2}) = \phi[z_{t\cdot}(\gamma_{\cdot 2}/\sigma_{22}^{1/2})], \quad F(z_{t\cdot}\gamma_{\cdot 2}) = \Phi[z_{t\cdot}(\gamma_{\cdot 2}/\sigma_{22}^{1/2})], \tag{18}$$

which shows that $\gamma_{\cdot 2}$ and σ_{22} are not separately identifiable, where ϕ and Φ are, respectively, the pdf and cdf of the **unit** normal random variable. This is so because $\gamma_{\cdot 2}$ can only be estimated by probit methods. Thus, **we need to impose the normalization**

$$\sigma_{22} = 1.$$

4 Estimation

Without loss of generality, the observations y_t **conditional on** $v_t \geq -z_{t\cdot}\gamma_{\cdot 2}$ behave like

$$y_{t(1)} = w_{t(1)\cdot}\theta_{\cdot 1} + \epsilon_{t(1)}, \;\; w_{t(1)\cdot} = (x_{t(1)\cdot}, 1, s_t), s_t = \frac{\phi(z_{t\cdot}\gamma_{\cdot 2})}{\Phi(z_{t\cdot}\gamma_{\cdot 2})}, \theta_{\cdot 1} = (\gamma_{\cdot 1}, \beta_{21}, \sigma_{12})', \tag{19}$$

while observations **conditional on** $v_t < -z_{t\cdot}\gamma_{\cdot 2}$ behave like

$$y_{t(2)} = w_{t(2)\cdot}\theta_{\cdot 1} + \epsilon_{t(2)}, \quad w_{t(2)\cdot} = (x_{t\cdot}, 0, s_t^*), s_t^* = -\frac{\phi(z_{t\cdot}\gamma_{\cdot 2})}{1 - \Phi(z_{t\cdot}\gamma_{\cdot 2})}. \tag{20}$$

It follows, therefore, that the likelihood function may be written as

$$L^*(\theta_{\cdot 1}, \sigma_{11.2}; y, X, Z) = \prod_{t=1}^{T} [f(y_t - w_{t1}.\theta_{\cdot 1})]^{\delta(t)} [f(y_t - w_{t2}.\theta_{\cdot 1})]^{1-\delta(t)}, \tag{21}$$

where f is the pdf of a random variable with mean zero and variance

$$\sigma_{11.2} = \sigma_{11} - \sigma_{12}^2, \tag{22}$$

which is non-negative by the Cauchy inequality for integrals.

In developing the estimating procedure we shall assume that f is the density of a **normal** variable with mean zero and variance $\sigma_{11.2}$. The logarithm of the entity in Eq. (21) is given by

$$L(\theta_{\cdot 1}, \sigma_{11.2}; y, X, Z) = -\frac{T}{2}\ln(2\pi\sigma_{11.2}) - \frac{1}{2\sigma_{11.2}}[\sum_{t=1}^{T_1}\delta(t)(y_t - w_{t1\cdot}\theta_{\cdot 1})^2$$

$$+ \sum_{t=T_1+1}^{T}(1-\delta(t))(y_t - w_{t2\cdot}\theta_{\cdot 1})^2] \tag{23}$$

$$= -\frac{T}{2}\ln(2\pi\sigma_{11.2}) - \frac{1}{2\sigma_{11.2}}[(y_{(1)} - W_1\theta_{\cdot 1})'(y_{(1)} - W_1\theta_{\cdot 1})$$

$$+(y_{(2)} - W_2\theta_{\cdot 1})(y_{(2)} - W_2\theta_{\cdot 1})], \tag{24}$$

where, for ease of exposition and without loss of generality, we have arranged to have the first T_1 observations correspond to the event $v_t \geq -z_{t\cdot}\gamma_{\cdot 2}$ and the last $T_2 = T - T_1$ observations correspond to the event $v_t < -z_{t\cdot}\gamma_{\cdot 2}$. The subscripts (1) and (2) denote, respectively, the observations under the two regimes. Later, when we deal with the properties of the resulting estimator we shall dispense with this convention which, in fact, cannot be generalized if we have more than one dependent indicator variable.

The first order conditions are

$$\frac{\partial L}{\partial \theta_{\cdot 1}} = \frac{1}{\sigma_{11.2}}(y_{(1)} - W_1\theta_{\cdot 1})'W_1$$

$$+\frac{1}{\sigma_{11.2}}(y_{(2)} - W_2\theta_{\cdot 1})'W_2 = 0 \tag{25}$$

$$\frac{\partial L}{\partial \sigma_{11.2}} = -\frac{T}{2}\frac{1}{\sigma_{11.2}} + \frac{1}{\sigma_{11.2}^2}(y_{(1)} - W_1\theta_{\cdot 1})'(y_{(1)} - W_1\theta_{\cdot 1})$$

$$\frac{1}{\sigma_{11.2}^2}(y_{(2)} - W_2\theta_{\cdot 1})'(y_{(2)} - W_2\theta_{\cdot 1})), \tag{26}$$

from which we obtain

$$\hat{\theta}_{\cdot 1} = [W_1'W_1 + W_2'W_2]^{-1}[W_1'y_{(1)} + W_2'y_{(2)}] \tag{27}$$

$$\hat{\sigma}_{11.2} = \left(\frac{T_1}{T}\frac{1}{T_1}(y_{(1)} - W_1\hat{\theta}_{\cdot 1})'y_{(1)} - W_1\hat{\theta}_{\cdot 1}) + \frac{T_2}{T}\frac{1}{T_2}(y_{(2)} - W_2\theta_{\cdot 1})'(y_{(2)} - W_2\theta_{\cdot})\right).$$

The representation above prompts the following assumption.

Assumption: The number of observations under the regime $v_t \geq -z.\gamma_{\cdot 2}$, T_1 , and the number of observations under the regime $v_t < -z_{t\cdot}\gamma_{\cdot 2}$, $T_2 = T - T_1$, obey

$$\lim_{T\to\infty} T_i = \infty, \quad \lim_{T\to\infty} \frac{T_i}{T} = \alpha_i, \quad \alpha_i \neq 0, \quad i = 1, 2.$$

4.1 Consistency

While the notation employed in the previous section is useful for pedagogical and computational purposes, it is not particularly suited for discussion of the probabilistic properties of the estimators in Eq. (27), or for generalization of the results above, because it conceals the essential randomness inherent in the definitions of the matrices W_1 and W_2 . Thus, consider now the alternative notational scheme, where

$$y = (y_t), W_1 = (x_{t\cdot}, 1, s_t), W_2 = (x_{t\cdot}, 0, s_t^*), \ t = 1, 2, \ldots, T, \tag{28}$$

and the second equation remains as before, i.e.

$$y_{t2} = z_{t\cdot}\gamma_{\cdot 2} + v_t, \ \ \delta(t) = 1 \ \text{ if and only if } \ v_t \geq -z_{t\cdot}\gamma_{\cdot 2}, \quad \text{and 0 otherwise.} \tag{29}$$

In this notation the log-likelihood fucntion of Eq. (23) is given by

$$L = \frac{T}{2}\ln 2\pi - \frac{\mathrm{T}}{2}\ln\sigma_{11.2} - \frac{1}{2\sigma_{11.2}}\sum_{1}^{\mathrm{T}}\left[\delta(\mathrm{t})(\mathrm{y_t} - \mathrm{w_{t1}}.\theta_{\cdot 1})^2 + (1-\delta(\mathrm{t}))(\mathrm{y_t} - \mathrm{w_{t2}}.\theta_{\cdot 1})^2\right], \tag{30}$$

which may be written more compactly as

$$L = \frac{T}{2}\ln 2\pi - \frac{\mathrm{T}}{2}\ln\sigma_{11.2} - \frac{1}{2\sigma_{11.2}}[(\mathrm{y} - \mathrm{W_1}\theta_{\cdot 1})'\mathrm{D}(\mathrm{y} - \mathrm{W_1}\theta_{\cdot 1})$$

$$+(y - W_2\theta_{\cdot 1})'(I - D)(y - W_1\theta_{\cdot 1})], \tag{31}$$

where

$$D = \mathrm{diag}(\delta(1), \delta(2), \ldots, \delta(T)). \tag{32}$$

One may easily verify that the notation above correctly renders the representation of Eqs. (23) and (24), without pre-committing as to which observations correspond to the regime $v_t \geq -z_{t\cdot}\gamma_{\cdot 2}$, and which to the regime $v_t < -z_{t\cdot}\gamma_{\cdot 2}$.

We may now maximize Eq. (31) with respect to $\theta_{\cdot 1}$ and $\sigma_{11.2}$. The first order conditions are easily obtained as

$$\frac{\partial L}{\partial \theta_{\cdot 1}} = \frac{1}{\sigma_{11.2}}\left[(y - W_{\cdot 1}\theta_{\cdot 1})'DW_1 + (y - W_{\cdot 2}\theta_{\cdot 1})'(I - D)W_2\right] = 0 \tag{33}$$

$$\frac{\partial L}{\partial \sigma_{11.2}} = -\frac{T}{2}\frac{1}{\sigma_{11.2}} + \frac{1}{2}\frac{1}{\sigma_{11.2}}[(y - W_1\theta_{\cdot 1})'D(y - W_1\theta_{\cdot 1})$$

$$| \quad +(y - W_2\theta_{\cdot 1})'(I - D)(y - W_2\theta_{\cdot 1})] = 0. \tag{34}$$

Solving the two equations above yields

$$\hat{\theta}_{\cdot 1} = \theta_{\cdot 1} + [W_1'DW_1 + W_2'(I - D)W_2]^{-1}[W_1'D + W_2'(I - D)W_2]\epsilon \tag{35}$$

$$\hat{\sigma}_{11.2} = \frac{1}{T}\left[(y - W_1\theta_{\cdot 1})'D(y - W_1\theta_{\cdot 1}) + (y - W_2\theta_{\cdot 1})'(I - D)(y - W_2\theta_{\cdot 1})\right] \tag{36}$$

The last equation is fully operational if we substitute therein $\hat{\theta}_{\cdot 1}$, for $\theta_{\cdot 1}$, and $\hat{\gamma}_{\cdot 2}$ for γ_2 , whenever the latter is required in the definition of s_t and s_t^* . Evidently, $\gamma_{\cdot 2}$ can be consistently estimated by probit methods applied to the second equation of the model.

To prove consistency we need to prove first, that

$$s_t(\hat{\gamma}_{\cdot 2}) \xrightarrow{\mathrm{P}} s_t(\gamma_{\cdot 2}), \qquad s_t^*(\hat{\gamma}_{\cdot 2}) \xrightarrow{\mathrm{P}} s_t^*(\gamma_{\cdot 2});$$

second, that

$$\tilde{Q} = \frac{W_1'DW_1 + W_2'(I - D)W_2}{T} = \frac{1}{T}\begin{bmatrix} X'X & X'\delta & X'[Ds + (I - D)s^*] \\ \delta'X & \delta'\delta & \delta'Ds \\ [s'D + s^{*\prime}(I - D)]X & s'D\delta & s'Ds + s^{*\prime}(I - D)s^* \end{bmatrix} \xrightarrow{\mathrm{P}} M, \tag{37}$$

M being non-singular and third,

$$\frac{1}{T}[W_1'D + W_2'(I - D)]\epsilon \xrightarrow{\mathrm{P}} 0.$$

Lemma 1. Let $\hat{\gamma}_{\cdot 2}$ be the estimator of $\gamma_{\cdot 2}$ obtained by probit analysis of the equation

$$y_{t2} = z_{t\cdot}\gamma_{\cdot 2} + v_t,$$

where y_{t2} is censored as above. It is well known that, at least, $\hat{\gamma}_{\cdot 2} \xrightarrow{\mathrm{P}} \gamma_{\cdot 2}$. Then

$$\phi(z_{t\cdot}\hat{\gamma}_{\cdot 2}),\ \Phi(z_{t\cdot}\hat{\gamma}_{\cdot 2}) \xrightarrow{\mathrm{P}} \phi(z_{t\cdot}\gamma_{\cdot 2}),\quad \Phi(z_{t\cdot}\gamma_{\cdot 2}),\quad \text{respectively.} \tag{38}$$

Proof: We shall do the proof for Φ ; the proof for ϕ is completely analogous. Expanding by the mean value theorem around the true value $\gamma_{\cdot 2}$, we have

$$\Phi(z_{t\cdot}\hat{\gamma}_{\cdot 2}) - \Phi(z_{t\cdot}\gamma_{\cdot 2}) = \phi(z_{t\cdot}\gamma_{\cdot 2}^*)z_{t\cdot}[\hat{\gamma}_{\cdot 2} - \gamma_{\cdot 2}], \tag{39}$$

where $|\gamma^*_{\cdot 2} - \gamma_{\cdot 2}| \leq |\hat{\gamma}_{\cdot 2} - \gamma_{\cdot 2}|$. Since

$$|\phi(z_{t\cdot}\gamma_{\cdot 2})z_{t\cdot}| < k,$$

it follows that

$$\phi(z_{t\cdot}\hat{\gamma}_{\cdot 2}),\ \Phi(z_{t\cdot}\hat{\gamma}_{\cdot 2}) \xrightarrow{\mathrm{P}} \phi(z_{t\cdot}\gamma_{\cdot 2}),\quad \Phi(z_{t\cdot}\gamma_{\cdot 2}),\quad \text{respectively.} \tag{40}$$

q.e.d.

Lemma 2. The matrix

$$\tilde{Q} = \frac{W_1' D W_1 + W_2'(I-D)W_2}{T}$$

is nonsingular, with probability 1 and

$$\operatorname*{plim}_{T\to\infty} \tilde{Q} = M > 0, \tag{41}$$

where

$$M = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} x_{t\cdot}' x_{t\cdot} & \Phi_t x_{t\cdot}' & 0 \\ \Phi_t x_{t\cdot} & \Phi_t & \phi_t \\ 0 & \phi_t & \frac{\phi_t^2}{\Phi_t(1-\Phi_t)} \end{bmatrix}. \tag{42}$$

Proof: Suppose not, i.e. suppose that $\tilde{Q}$ is singular; then there exists a non-null vector (c, c_1, c_2) such that, at least,

$$0 = X'Xc + c_1 X'\delta + c_2 X'[Ds + (I-D)s^*], \quad \text{or}$$

$$X'Xc = -(c_1 X'\delta + c_2 X'[Ds + (I-D)s^*]). \tag{43}$$

But this is a contradiction with probability one, since X is either a non-stochastic matrix or, if stochastic, it is independent of v and hence of δ ; thus, the event of Eq. (42) can only occur with probability zero. To complete the proof of the Lemma we need only obtain an expression for the probability limit of $\tilde{Q}$. First we note that from Eq. (36)

$$\tilde{Q} = \frac{1}{T} \begin{bmatrix} X'X & X'\delta & X'[Ds + (I-D)s^*] \\ \delta' X & \delta'\delta & \delta' D s \\ [s'D + s^{*\prime}(I-D)]X & s'D\delta & s'Ds + s^{*\prime}(I-D)s^* \end{bmatrix}. \tag{44}$$

But,

$$\frac{1}{T}\delta'\delta = \frac{1}{T}\sum_{t=1}^{T} (\delta(t))^2 \tag{45}$$

$$\frac{1}{T}\delta' X = \frac{1}{T}\sum_{t=1}^{T} \delta(t) x_{t\cdot} \tag{46}$$

$$\frac{1}{T}\delta' s = \frac{1}{T}\sum_{t=1}^{T} \frac{\phi_t}{\Phi_t}\delta(t) \tag{47}$$

$$\frac{1}{T}X' X = \frac{1}{T}\sum_{t=1}^{T} x_{t\cdot}' x_{t\cdot} \xrightarrow{\mathrm{P}} M_{xx} > 0. \tag{48}$$

Moreover,

$$\frac{1}{T}X'[Ds + (I-D)s^*] = \frac{1}{T}\left(\sum_{t=1}^{T} \frac{\phi_t}{\Phi_t}\delta(t) x_{t\cdot}' - \sum_{t=1}^{T} \frac{\phi_t}{1-\Phi_t}(1-\delta(t) x_{t\cdot}'\right) \tag{49}$$

$$\frac{1}{T}[s' Ds + s^{*\prime}(I-D)s^*] = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\phi_t^2}{\Phi_t^2}\delta(t)^2 + \frac{\phi_t^2}{(1-\Phi_t)^2}(1-\delta(t))^2\right). \tag{50}$$

In Eq. (45) we see that we are dealing with a sequence of independent random variables with

$$E\delta(t)^2 = \Phi_t, \quad \mathrm{Var}(\delta(t)^2) = \Phi_t(1-\Phi_t). \tag{51}$$

Since their variance is **uniformly** bounded (by 1), we conclude from Kolmogorov's criterion, see Proposition 22, Dhrymes (1989), p. 186,

$$\frac{1}{T}\delta'\delta \stackrel{\mathrm{a.c}}{\rightarrow} \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Phi_t. \tag{52}$$

In Eq. (46) we are dealing with a sequence of independent non-identically distributed random vectors with

$$E\delta(t)x_{t\cdot} = \Phi_t x_{t\cdot}, \quad \mathrm{Cov}(\delta(t)x_{t\cdot}') = (1-\Phi_t)\Phi_t x_{t\cdot}' x_{t\cdot}. \tag{53}$$

Since $\Phi_t(1-\Phi_t) < 1$, by the assumptions [1] made on X we conclude

$$\operatorname*{plim}_{T\to\infty}\frac{1}{T}\delta' X = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Phi_t x_{t\cdot}. \tag{54}$$

The limits of the remaining entities follow from the derivations above. Thus,

$$\operatorname*{plim}_{T\to\infty}\frac{1}{T}X'[Ds + (I-D)s^*] \;=\; \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}[\phi_t - \phi_t]x_{t\cdot}' = 0 \tag{55}$$

[1]If we were to assume that the x's are **bounded** the sequence would obey the Kolmogorov criterion referred to above and, hence, converge to its limit with probability one.

$$\frac{s'Ds + s^{*\prime}(I-D)s^*}{T} \stackrel{\text{a.c.}}{\rightarrow} \lim_{T\rightarrow\infty} \frac{1}{T}\sum_{t=1}^{T} \frac{\phi_t^2}{\Phi_t(1-\Phi_t)}. \tag{56}$$

q.e.d.

Lemma 3. Given W_1, W_2, D , as defined above,

$$\zeta_T = \frac{1}{T}[W_1'D + W_2'(I-D)]\epsilon = \frac{1}{T}\sum_{t=1}^{T}\xi_t \stackrel{\text{P}}{\rightarrow} 0. \tag{57}$$

Proof: The sequence above is one of independent non-identically distributed random vectors with

$$E\xi_t = E[w_{t1}'.\delta(t) + w_{t2}'.(1-\delta_(t))]\epsilon_t = 0, \quad \mathrm{Cov}(\xi_t) = \sigma_{11.2}m_t, \tag{58}$$

where m_t is the t th summand of M , as defined in Eq. (40). Thus,

$$E\zeta_T = 0, \quad \lim_{T\rightarrow\infty}\mathrm{Cov}(\zeta_T) = \lim_{T\rightarrow\infty}\sigma_{11.2}\frac{1}{T}M = 0. \tag{59}$$

But this implies that $\zeta_T \stackrel{\text{q.m}}{\rightarrow} 0$ and hence $\zeta_T \stackrel{\text{P}}{\rightarrow} 0$.

q.e.d.

Proposition 1: The estimator of $\theta_{\cdot 1}$ given in Eq. (27) converges (to the true parameter) at least in probability.

Proof: Lemmata 1, 2 and 3.

4.2 Limiting Distribution

From Eq. (35) we can write

$$\sqrt{T}(\hat{\theta}_{\cdot 1} - \theta_{\cdot 1}) = \tilde{Q}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix} x_{t\cdot}' \\ \delta_t \\ s_t\delta(t) + s_t^*(1-\delta(t)) \end{pmatrix}\epsilon_t = \tilde{Q}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_t. \tag{60}$$

Since

$$\tilde{Q} \stackrel{\text{P}}{\rightarrow} M, \tag{61}$$

where M is as defined in Eq. (40) and ξ_t , $t = 1, 2, \ldots$ is a sequence of independent non-identically distributed random vectors with

$$E\xi_t = 0, \quad \mathrm{Cov}(\xi_t) = \sigma_{11.2} m_t, \tag{62}$$

and since

$$\tilde{Q} \quad \stackrel{\mathrm{P}}{\to} \quad M > 0 \tag{63}$$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} m_t = M, \tag{64}$$

it follows from the discussion in Dhrymes (1994) pp. 64-68, that the sequence obeys the Lindeberg condition and hence, by the Lindeberg central limit theorem, see Dhrymes (1989) pp. 266-271 we conclude that

$$\zeta_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} w'_{t\cdot}\epsilon_t \stackrel{\mathrm{d}}{\to} N(0,\ \sigma_{11.2}M). \tag{65}$$

Moreover, since

$$\sqrt{T}(\hat{\theta}_{\cdot 1} - \theta_{\cdot 1}) \sim M^{-1}\zeta_T \tag{66}$$

it follows that

$$\sqrt{T}(\hat{\theta}_{\cdot 1} - \theta_{\cdot 1}) \stackrel{\mathrm{d}}{\to} N(0,\ \sigma_{11.2}M^{-1}). \tag{67}$$

5 Comparison with Alternative Estimators

This problem has been examined *inter alia* by Heckman (1974), (1978) in a much broader context, using the equivalent of full information maximum likelihood methods. However, by far the most common use of this model, see e.g. Angrist, (give also other references) is one in which one writes the second equation as a "linear probability" model, i.e. one writes the model as

$$y_t = x_{t\cdot}\gamma_{\cdot 1} + u_t, \quad \delta(t) = z_{t\cdot}\gamma_{\cdot 2} + v_t, \tag{68}$$

and "estimates"

$$\hat{\delta}(t) = Z(Z'Z)^{-1}Z'\delta(t), \tag{69}$$

and obtains the "2SLS" estimator

$$\tilde{\eta}_{\cdot 1} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y, \quad \tilde{X}_* = (X, \hat{\delta}). \tag{70}$$

This estimator may be shown to be consistent; bearing in mind that $X = ZS_k$, where S_k is a suitable **selection** matrix, it has the limiting distribution

$$\sqrt{T}(\tilde{\eta}_{\cdot 1} - \eta_{\cdot 1}) \xrightarrow{\mathrm{d}} N(0, \sigma_{11} C_2^{-1}), \tag{71}$$

where

$$C_2 = \operatorname*{plim}_{T\to\infty} \frac{1}{T} \begin{bmatrix} X'X & X'\delta \\ \delta'X & \hat{\delta}'\hat{\delta} \end{bmatrix} = \begin{bmatrix} M_{xx} & M_{x\Phi} \\ M_{\Phi x} & M_{\Phi z} M_{zz}^{-1} M_{z\Phi} \end{bmatrix} \tag{72}$$

where

$$M_{x\Phi} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_{t\cdot}' \Phi_t \tag{73}$$

$$M_{z\Phi} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_{t\cdot}' \Phi_t \tag{74}$$

$$M_{zz} = \operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_{t\cdot}' z_{t\cdot}. \tag{75}$$

Remark 2. The use of a "linear probability" model in this context is intellectually indefensible. However, it is not necessary to use a linear probability model rationalization or its equivalent at all in this context. Instead one could use the method given in Dhrymes (1969). Since by assumption Z is of full column rank, $Z'Z$ is a positive definite matrix. By Corollary 2.15 in Dhrymes (2000), p. 86, there exists a **nonsingular** matrix R such that

$$Z'Z = RR'. \tag{76}$$

Consider

$$R^{-1}Z'y = R^{-1}Z'X\gamma_{\cdot 1} + (R^{-1}Z'\delta)\beta_{21} + R^{-1}Z'v. \tag{77}$$

and note that it may be readily verified that the OLS estimator of the parameters of the transformed model above is the "2SLS" estimator examined earlier.

An aspect of this proceduree, which may well indicate that this is not an efficient estimator even in a limited information context, is that exactly the same approach will work when y_2 is not censored, but it is completely observable. Thus, such a procedure does not fully exploit what is special in this case.

6 Relative Efficiency

Since both estimators converge at the same rate to a normal distribution centered on the true parameter, the question of relative efficiency involves only the comparison of the covariance matrices of the two limiting distributions. However, because $\theta_{\cdot 1}$ involves one more parameter(σ_{12}) than does $\eta_{\cdot 1}$, we need to produce the **marginal** distribution of $\hat{\eta}_{\cdot 1}$ as obtained from the distribution of $\hat{\theta}_{\cdot 1}$. By the standard properties of the normal distribution this yields

$$\sqrt{T}(\hat{\eta}_{\cdot 1}-\theta) \xrightarrow{\mathrm{d}} N(0, \sigma_{11.2}C_1^{-1}), \quad C_1 = \begin{bmatrix} M_{xx} & M_{x\Phi} \\ M_{\Phi x} & \bar{\Phi}-\nu \end{bmatrix}, \tag{78}$$

$$\nu = \frac{\bar{\phi}^2}{\nu_1}, \quad \nu_1 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \frac{\phi_t^2}{\Phi_t(1-\Phi_t)}, \tag{79}$$

where

$$\bar{\Phi} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \Phi_t, \quad \bar{\phi} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \phi_t. \tag{80}$$

By Proposition 2.66 in Dhrymes (2000, p. 89, $\hat{\eta}_{\cdot 1}$ is efficient relative to $\tilde{\eta}_{\cdot 1}$ if and only if

$$\omega = \frac{1}{\sigma_{11.2}}C_1 - \frac{1}{\sigma_{11}}C_2 \geq 0. \tag{81}$$

But this can be written as

$$\omega = \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{11.2}} \begin{bmatrix} M_{xx} & M_{xPhi} \\ M_{\Phi x} & \bar{\Phi} \end{bmatrix} \tag{82}$$

$$+\frac{1}{\sigma_{11}\sigma_{11.2}} \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{11.2}\nu_2 - \sigma_{11}\frac{\bar{\phi}^2}{\nu_1} \end{bmatrix}, \tag{83}$$

where $\nu_2 = \mathrm{plim}_{T\to\infty} \frac{1}{T}\delta'(I-P_z)\delta$ and P_z is the **projection** matrix $Z(Z'Z)^{-1}Z'$. Since the first matrix in the right member of Eq. (83) is unambiguously positive definite, a sufficient, but not a necessary, condition for $\hat{\eta}_{\cdot 1}$ to be efficient relative to $\tilde{\eta}_{\cdot 1}$ is that

$$\sigma_{11.2} \operatorname*{plim}_{T\to\infty} \frac{1}{T}\delta'(I-P_z)\delta - \sigma_{11}\frac{\bar{\phi}^2}{\nu_1} \geq 0, \tag{84}$$

provided $\sigma_{12} \neq 0$. If $\sigma_{12} = 0$, the necessary and sufficient condition for efficiency becomes

$$\nu_2 - \frac{\bar{\phi}^2}{\nu_1} \geq 0. \tag{85}$$

A few observations may be made regarding the behavior of entities like

$$\frac{\phi_t^2}{\Phi_t(1-\Phi_t)},$$

i.e. that both numerator and denominator vanish at $\pm\infty$, so the fraction may be defined to be **zero** at those values; both numerator and denominator assume their (global) maxima at zero. At zero, the fraction becomes

$$\frac{(2\pi)^{-1}}{.25} \approx .63.$$

In addition the maximum of ϕ_t^2 occurs at zero and its value is approximately .16. If the numerator and the denominator of ν assume a value mid-way between their maxima and minima we should have that

$$\nu \approx \frac{.08}{.32} = .25.$$

Finally, about the only thing we can say about the other term is

$$\nu_2 \leq \bar{\Phi}.$$

The latter can assume a value in (0, 1), and a value in the mid range is .5. With these numerical assumptions the proposed estimator is **efficient**. However, a Monte Carlo study is needed to clarify this issue.

This proposed estimator has the advantage that it allows a routine test of the hypothesis of endogeneity, since it yields an estimator of the parameter σ_{12}. By contrast the standard estimator does not, so that the hypothesis of endogeneity cannot be tested, except through the dubious Hausman "test", which is not really a test in the strict meaning of the term.

REFERENCES

Angrist, J. D. (2001) “Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressor: Simple Strategies for Empirical Practice, *Journal of Business and Economic Statistics*, vol. 19, pp.2-16.

Dhrymes, P. J. (1969) “Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS and 3SLS Estimated Parameters”, *Review of Economic Studies*, vol. 36, pp. 213-226.

Dhrymes, P. J. (1989), *Topics in Advanced Econometrics: Vol. I Probability Foundations*, Springer Verlag, New York.

Dhrymes, P. J. (1994), *Topics in Advanced Econometrics: Vol. II, Linear and Nonlinear Simultaneous Equations*, Springer-Verlag, New York.

Dhrymes, P. J. (2000), *Mathematics for Econometrics*, (third edition), Springer-Verlag, New York.

Heckman J. (1974), “Shadow Prices, Market Wages and Labor Supply *Econometrica*, vol. 42, pp.679-674.

Heckman, J. (1978), “Dummy Endogenous Variables in a Simultaneous Equations System, *Econometrica*, vol. 46, pp. 931-955.