Simultaneous equations models with truncated dependent variables: a simultaneous tobit model

Robin C. Sickles and Peter Schmidt

In a recent paper Amemiya has discussed a simultaneous equations model in which all of the dependent variables are truncated [2]. This model is an extension to the simultaneous equations case of the Tobit model (Tobin [15]; Amemiya [1]). While it is certainly useful to embed the Tobit model into a simultaneous equations setting, it should be recognized that, in empirical applications, truncated variables are the exception rather than the rule. It is possible to have a simultaneous equations model with all of the dependent variables truncated, but it is apt to be far more common to have a model in which only one or two of the dependent variables are truncated with the other dependent variables being of the usual kind. The purpose of this work is to carry out an analysis of simultaneous equations models when some, but *not all*, of the dependent variables are truncated. As it might be expected, many of our results are similar to those of Amemiya [2]; however, this similarity is not always the case. Also developed here are some new results for Amemiya's case—all dependent variables truncated—as well as our results.

In both the models of Amemiya [2] and in the models of this work, it is the truncated (observed) dependent variable which appears on the right hand side of the structural equations. An alternative specification is to have the untruncated (hypothetical) version of the dependent variable on the right hand side of the structural model. This alternative specification has been analyzed by Nelson and Olson [13] and Amemiya [3] and is related closely to work of Heckman [9]. The rather substantial differences between these alternative specifications are explored in some detail.

To simplify the exposition, the second part will consider in detail the case of a two-equation model with one dependent variable truncated. The third section considers the general case. Our conclusions are presented in the fourth section.

The two-equation case

The two-equation system considered by Amemiya [2] can be written as follows:

$$(1)\quad y_{1t}^* = \gamma_1 y_{2t} + \delta_1' x_t + \epsilon_{1t}$$
$$y_{2t}^* = \gamma_2 y_{1t} + \delta_2' x_t + \epsilon_{2t}$$

where

$$(2)\quad y_{1t} = \begin{cases} y_{1t}^* \text{ when } y_{1t}^* > 0 \\ 0 \text{ otherwise} \end{cases}$$

$$y_{2t} = \begin{cases} y_{2t}^* \text{ when } y_{2t}^* > 0. \\ 0 \text{ otherwise} \end{cases}$$

Here x_t is a $K \times 1$ vector of exogenous variables; δ_1 and δ_2 are $K \times 1$ vectors of coefficients; and γ_1 and γ_2 are scalars. One observes y_{1t} and y_{2t} but not y_{1t}^* and y_{2t}^*.

The system to be dealt with in this article is the same except that y_2 is not truncated. Thus, one has

$$(3)\quad y_{1t}^* = \gamma_1 y_{2t} + \delta_1' x_t + \epsilon_{1t}$$
$$y_{2t}^* = \gamma_2 y_{1t} + \delta_2' x_t + \epsilon_{2t}$$

where

$$(4)\quad y_{1t} = \begin{cases} y_{1t}^* \text{ when } y_{1t}^* > 0. \\ 0 \text{ otherwise} \end{cases}$$

Here one observes y_{1t} and y_{2t} but not y_{1t}^*.

It should be stressed at this point that it is the *observed* variable y_{1t} which appears on the right hand side of the second structural equation in (3), not the unobserved (hypothetical) variable y_{1t}^*. This distinction is important. For example, Nelson and Olson [13] and

Dr. Sickles is an assistant professor in the department of economics at George Washington University, Washington, D.C.; Dr. Schmidt is a professor in the department of economics at Michigan State University, East Lansing, MI. We would like to thank James J. Heckman for his helpful comments. Remaining errors are our responsibility.

Amemiya [3] have considered the model

$$(5)\quad y_{1t}^* = \gamma_1 y_{2t} + \delta_1' x_t + \epsilon_{1t}$$
$$y_{2t} = \gamma_2 y_{1t}^* + \delta_2' x_t + \epsilon_{2t}$$

plus equation (4). As one will see, this model differs from that in (3) in several important ways. First, model (5) has a single reduced form which easily is estimated; while model (3) has two reduced forms which are not estimated easily. Second, the conditions for identification of the two models are quite different. Third, in model (3) the internal consistency of the model requires certain restrictions on the parameters which is not true in model (5).

Which model is more reasonable really depends on the economic application. In some cases there actually may be a variable y_{1t}^*, with an obvious interpretation, which is truncated only by limits in observation. Then model (5) may be preferred. On the other hand, in other cases, y_{1t}^* may be an entirely hypothetical variable with no obvious interpretation in which case model (3) may be preferred. With these preliminaries out of the way, one now returns to consideration of our model as given in (3) and (4). Let one partition this set of observations into those sets for which the truncation is effective and those for which it is not

$$(6)\quad t_1 \equiv \{t \mid y_{t1} > 0\} = \{t \mid y_{t1}^* > 0\} = \{t \mid \text{truncation is effective}\};$$
$$t_2 \equiv \{t \mid y_{t1} = 0\} = \{t \mid y_{t1}^* \leq 0\} = \{t \mid \text{truncation is effective}\}.$$

Then, one notes that for $t \epsilon t_1$ the reduced form of the model is

$$(7)\quad y_{1t}^* = \frac{1}{1-\gamma_1\gamma_2}(\delta_1' + \gamma_1\delta_2')x_t + \frac{1}{1-\gamma_1\gamma_2}(\epsilon_{1t} + \gamma_1\epsilon_{2t})$$
$$y_{2t} = \frac{1}{1-\gamma_1\gamma_2}(\delta_2' + \gamma_2\delta_1')x_t + \frac{1}{1-\gamma_1\gamma_2}(\gamma_2\epsilon_{1t} + \epsilon_{2t});$$

while for $t \epsilon t_2$ the reduced form is

$$(8)\quad y_{1t}^* = (\delta_1' + \gamma_1\delta_2')x_t + (\epsilon_{1t} + \gamma_1\epsilon_{2t})$$
$$y_{2t} = \delta_2' x_t + \epsilon_{2t}.$$

The reduced form expression for y_{1t}^* in (8) obtains when $y_{1t}^* \leq 0$; that is, when $(\delta_1 + \gamma_1\delta_2)'x_t + (\epsilon_{1t} + \gamma_1\epsilon_{2t}) \leq 0$. One sees that $(\delta_1 + \gamma_1\delta_2)'x_t + (\epsilon_{1t} + \gamma_1\epsilon_{2t})$ determines the sign of y_{1t}^*. When it is negative, y_{1t}^* is negative; and this reduced form expression is consistent with (8). Conversely, when $(\delta_1 + \gamma_1\delta_2)'x_t + (\epsilon_{1t} + \gamma_1\epsilon_{2t})$ is positive, y_{1t}^* must be positive. However, this is consistent with the expression for y_{1t}^* in (7) if, and only if, $1 - \gamma_1\gamma_2 > 0$.

One has found that a necessary condition for the existence of our model, as given in (3) and (4), is $1 - \gamma_1\gamma_2 > 0$. This finding is identical to the condition given by Amemiya [2, p. 1008] for the existence of his model, as given by (1) and (2). It will be seen later that this equivalence of Amemiya's condition and ours does not always hold for larger models—as long as not all variables are truncated.

It is easy to contrast this condition with the reduced form of the model given by equations (5) and (4). For this model the reduced form in (7) holds for all t. No restrictions are needed for the internal consistency of the model. On the other hand, one can solve for the structural parameters in (5) from the reduced form parameters in (7) only if the usual rank and order conditions for identification are satisfied.

It also should be pointed out that, when the reduced form (7) is derived from the model (5), it is essentially identical to the model of Heckman [9, p. 476], though with a slightly different truncation rule. The first equation in (7) can be estimated consistently by Tobit analysis, while the second can be estimated consistently by least squares. Consistent estimates of the structural parameters then can be found by the analogue of indirect least squares (if they are identified). Alternatively, consistent and asymptotically efficient estimates can be substained via the method of maximum likelihood.

Now the authors return to their two-equation model given in (3) and (4) with its two reduced forms (7) and (8). Since each reduced form holds only for a (nonrandom) subset of the observations, estimation of the reduced forms is not trivial. The authors demonstrate that one can obtain consistent estimates of the reduced form

(7). If the usual rank and order conditions hold, these can be transformed into consistent estimates of the structural parameters; therefore, these may be of interest as a computationally simple way of obtaining consistent estimates of the structural parameters. Also it will be shown that one can estimate enough additional parameters of the reduced form (8) to guarantee the identification of the *second* structural equation in (3); one exclusion restriction on the first structural equation suffices to identify the model.

The reduced form (7) is considered now in which the authors wish to estimate using the observations for which $y_{1t} > 0$ ($t \epsilon t_1$, as defined in (6)). Essentially the same problem is treated by Amemiya [2] for his model in which both dependent variables are truncated and in which one needs only to make the necessary modifications in his estimator.

To do so, rewrite equation (7) as

$$(9)\quad \begin{aligned} y_{1t} &= \pi_1' x_t + w_{1t} \\ y_{2t} &= \pi_2' x_t + w_{2t} \end{aligned}$$

where

$$(10)\quad \pi_1 = \frac{1}{1-\gamma_1\gamma_2}(\delta_1 + \gamma_1\delta_2)$$

$$\pi_2 = \frac{1}{1-\gamma_1\gamma_2}(\delta_2 + \gamma_2\delta_1)$$

and

$$(11)\quad w_{1t} = \frac{1}{1-\gamma_1\gamma_2}(\epsilon_{1t} + \gamma_1\epsilon_{2t})$$

$$w_{2t} = \frac{1}{1-\gamma_1\gamma_2}(\gamma_2\epsilon_{1t} + \epsilon_{2t}).$$

Let the distribution of $(w_{1t}, w_{2t})'$, *not* conditional upon $t\epsilon t_1$, be $N(0, \Sigma)$ where

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

The basic equations which are derived are as follows:

$$(12)\quad y_{1t}^2 = \begin{bmatrix} \dfrac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}} \\[2ex] \pi_1 - \dfrac{\sigma_{12}\pi_2}{\sigma_{22}} \\[2ex] \dfrac{\sigma_{12}}{\sigma_{22}} \end{bmatrix}' \begin{bmatrix} 1 \\ y_{1t}x_t \\ y_{1t}y_{2t} \end{bmatrix} + \eta_{1t}$$

$$(13)\quad y_{2t}^2 = \begin{bmatrix} \dfrac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{11}} \\[2ex] \pi_2 - \dfrac{\sigma_{12}\pi_1}{\sigma_{11}} \\[2ex] \dfrac{\sigma_{12}}{\sigma_{11}} \end{bmatrix}' \begin{bmatrix} 1 \\ y_{2t}x_t \\ y_{1t}y_{2t} \end{bmatrix} + \eta_{2t}$$

where $E(\eta_{1t}) = E(\eta_{2t}) = 0$.

The set of equations (12) and (13) is similar to equation (2.15) of Amemiya [2, p. 1003]. As in Amemiya's case, ordinary least squares is inappropriate because of the correlation between the η's and the y's. However, one can use the instrumental variables method to get consistent estimates. The required instruments can be formed by regressing y_{1t} and y_{2t} on x_t and higher powers of x_t to form $\hat{y}_{1t}$ and $\hat{y}_{2t}$ and then by replacing y_{1t} and y_{2t} on the right hand sides of (12) and (13) by these predicted values. The resulting instrumental variables estimators are consistent estimators of the parameters in (12) and (13). Finally, from these estimates one can solve for estimates of the reduced form parameters ($\pi_1, \pi_2, \sigma_{11}, \sigma_{12}, \sigma_{22}$).[1]

One has arrived at consistent estimates of the reduced form parameters corresponding to the set of observations for which $y_{1t} > 0$. The relationship of these parameters to the original structural parameters is evident from equation (7) which gives the reduced form for this set of observations. As in the usual simultaneous

1. In some cases, variables not in the model could be used as instruments; however, it is difficult to make any general statement of where these variables would come from.

equations case, one can solve for the structural parameters from the reduced form parameters if the usual rank condition is satisfied. (In the present context, this solution essentially requires that at least one element of both δ_1 and δ_2 be known a priori to equal zero.) It should be stressed that in the present model, unlike the usual simultaneous equations model, this condition is *not* necessary for identification of the structural parameters; it is only necessary for the identification of the structural parameters from only the set of observations for which $y_{1t} > 0$. But, in any case, if the usual rank condition holds, one has arrived at a sort of indirect least squares method of getting consistent estimates of the structural parameters.

These indirect least squares estimates are consistent but inefficient. (Clearly, one cannot hope to get efficient estimates when one restricts his attention to only a subset of the observations.) Furthermore, their asymptotic distribution is not known. Their main usefulness is as starting values for an iterative scheme to calculate the maximum likelihood estimates to be discussed shortly. Alternatively, one can rely on the well-known result that if one starts with any consistent estimate of the parameters, the result of a single Newton-Raphson iteration on the log-likelihood function is consistent and asymptotically efficient. (This result is just the so-called "method of scoring" or "linearized maximum likelihood method"—see Rothenberg and Leenders [14] or Dhrymes [4, 5]. Use of these initial consistent estimates can help to cut the computational expense of finding asymptotically efficient estimates.

One can, if he wishes, use these initial consistent estimates in one other way. Proceeding along the lines of Amemiya [2, p. 1011] and Heckman [9, pp. 479–80], one can write

$$(14)\quad y_{1t} = \pi_1'x_t + \sigma_{11}\frac{f(\pi_1'x_t)}{F(-\pi_1'x_t)} + e_{1t}$$

$$y_{2t} = \pi_2'x_t + \sigma_{12}\frac{f(\pi_1'x_t)}{F(-\pi_1'x_t)} + e_{2t}$$

where $E(e_{1t}) = E(e_{2t}) = 0$ (conditional on $t \epsilon t_1$) and where f and F are, respectively, the density and cdf of $N(0, \sigma_{11})$. One then can use our initial consistent estimates $\hat{\pi}_1$ and $\hat{\sigma}_{11}$ in place of π_1 and σ_{11} to form the regressor $f(\pi_1'x_t)/F(-\pi_1'x_t)$ and to apply ordinary least squares (or generalized least squares since e_{1t} and e_{2t} are heteroskedastic). Amemiya [2, p. 1005] claims that the substitution of $\hat{\pi}_1$ and $\hat{\sigma}_{11}$ for π_1 and σ_{11} does not affect the asymptotic distribution of the estimates. However, this claim is not correct, as shown by Heckman [10]. The correct asymptotic distribution unfortunately depends on the asymptotic distribution of the initial consistent estimate which is unknown in this case. As a result there is no good reason to prefer these estimates to the initial consistent estimates themselves.

It also may be worth pointing out that in some cases one may have available only the observations for which $y_{1t} > 0$. (That is, one may not have collected the observations for $t \epsilon t_2$.) For one such example, see Hausman and Wise [7] who deal with a sample truncation arising in the New Jersey negative income tax experiment. The simple consistent estimates described here still can be used and indeed can seem more reasonable in this case since there are no observations in t_2 being ignored. Of course, they still will not be as efficient as the maximum likelihood estimates.

Now the discussion of the two equation case is continued by returning to the question of the identification of the structural parameters. It has been seen that there are two different reduced forms for the model corresponding to the sets of observations for which the truncation on y_1 is and is not effective; these are given in equations (7) and (8). The estimation of the reduced form has been discussed in (7) from the set of observations for which the truncation is not effective; and it has been noted that the reduced form parameters in (7) imply unique values for the structural parameters if, and only if, the usual rank condition holds. However, this condition is *not* necessary for identification of the structural parameters, at least for the second structural equations. The reason is that it ignores the information available in the reduced form given in equation (8)—that is, the information in the set of observations for which the truncation is effective.

To be more precise, one notes that one can estimate from the observations for which the first reduced form, equation (7), is effective, the following parameters:

$$(15)\quad \pi_1 = \frac{1}{1-\gamma_1\gamma_2}(\delta_1 + \gamma_1\delta_2)$$

$$\pi_2 = \frac{1}{1-\gamma_1\gamma_2}(\delta_2 + \gamma_2\delta_1)$$

$$\sigma_{11} = \frac{1}{(1-\gamma_1\gamma_2)^2}(\Omega_{11} + 2\gamma_1\Omega_{12} + \gamma_1^2\Omega_{22})$$

$$\sigma_{12} = \frac{1}{(1-\gamma_1\gamma_2)^2}[\gamma_2\Omega_{11} + (1+\gamma_1\gamma_2)\Omega_{12} + \gamma_1\Omega_{22}]$$

$$\sigma_{22} = \frac{1}{(1-\gamma_1\gamma_2)^2}(\gamma_2^2\Omega_{11} + 2\gamma_2\Omega_{12} + \Omega_{22}).$$

In Appendix B it is shown that from the observations for which the second reduced form is effective, one also can estimate the following parameters:

$$(16)\quad \Psi_2 = \delta_2$$

$$\beta = \frac{-(\Omega_{12} + \gamma_1\Omega_{22})}{1-\gamma_1\gamma_2}.$$

The first equation of (16) clearly shows that δ_2 is identified. Also, the second equation of (16) can be rewritten by observing that

$$\Omega_{12} = -\gamma_2\sigma_{11} - \gamma_1\sigma_{22} + (1+\gamma_1\gamma_2)\sigma_{12}$$

$$\Omega_{22} = \gamma_2^2\sigma_{11} + \sigma_{22} - 2\gamma_2\sigma_{12}.$$

With these substitutions one finds

$$\beta = \gamma_2\sigma_{11} - \sigma_{12}.$$

Since β is estimable and so are σ_{11} and σ_{12}, one can solve for γ_2 as follows:

$$(17)\quad \gamma_2 = (\beta + \sigma_{12})/\sigma_{11}.$$

It is seen that one can identify δ_2 and γ_2, the parameters of the second structural equation in (3), *without any restrictions on that equation.* However, the first structural equation in (3) is identified only if it satisfies one exclusion restriction—that is, if the usual order and rank conditions for that equation hold. That this condition is sufficient for identification of the first structural equation is clear from the usual results for linear systems. To see that it is necessary, one can try to solve (15) and (16) for the structural parameters (γ_1, γ_2, δ_1, δ_2, Ω_{11}, Ω_{12}, Ω_{22}) in terms of estimable quantities (π_1, π_2, σ_{11}, σ_{12}, σ_{22}, Ψ_2, β. This solving is done in Appendix B for the simple case of $K = 1$.

It may seem strange to some readers that the second structural equation is identified without any restrictions. To see why this is the case, consider a linear combination of the two structural equations in (3): a · first equation + b · second equation implies

$$ay_1^* + by_2 = a\gamma_1 y_2 + a\delta_1'x + a\epsilon_1 + b\gamma_2 y_1 + b\delta_2'x + b\epsilon_2.$$

Solving for y_2 gives

$$y_2 = \frac{1}{b - a\gamma_1}\{b\gamma_2 y_1 + (a\delta_1' + b\delta_2')x + (a\epsilon_1 + b\epsilon_2) - ay_1^*\}.$$

This equation is distinguishable from the second structural equation of (3) *because it contains y_1^* as a right-hand-side variable.* Another way of making the same point is to observe that, for the set of observations for which the truncation is effective, y_{1t} does not appear in the second equation (because it is zero). Intuitively, this is equivalent to one exclusion from the second equation, for at least one set of observations.

The authors will return to this point later when they get to the general case. However, they, at this point, mention the relevance of this argument to the model of Amemiya [2], as given in equation (1), in which *both* dependent variables are truncated. In this case both structural equations are identified without any exclusion restrictions. One way to see this identification is to note that, if one takes linear combinations of the equations in (1) and solves for y_1^*, this equation will include y_2^* as a right-hand-side variable, and conversely. Alternatively, one can note that corresponding to model (1) are four reduced forms, corresponding to the possible permutations of the two truncations being or not being effective. For each of the two reduced forms corresponding to one truncation effective, one can estimate parameters precisely analogous to those in (16) above, from which one can solve for the structural parameters of one of the two structural equations.

Finally, the discussion of the two-equation case is finished with a brief discussion of maximum likelihood estimation. If one assumes that the vectors $(\epsilon_{1t}, \epsilon_{2t})'$, $t = 1,2,\ldots,T$, are independently and identically distributed as $N(0,\Omega)$ and if one lets $f(\ )$ denote the density function of $N(0,\Omega)$, then the likelihood function is of the form

$$\text{(18)} \quad L = \underset{t \epsilon t_1}{\pi} (1 - \gamma_1\gamma_2) f(y_{1t} - \gamma_1 y_{2t} - \delta_1' x_t, y_{2t} - \gamma_2 y_{1t} - \delta_2' x_t)$$

$$\underset{t \epsilon t_2}{\pi} \int_{-\infty}^{[-\gamma_1 y_{2t} - \delta_1' x_t]} f(\epsilon_1, y_{2t} - \delta_2' \chi_t) d\epsilon_1$$

This likelihood function can be maximized by using a numerical maximization program to obtain the maximum likelihood estimates. Asymptotic standard errors can be obtained from the information matrix which in this case is best obtained by numerical differentiation, due to the complexity of the expressions for first and second derivatives.

The maximum likelihood estimates are consistent and asymptotically efficient. In particular, they are more efficient than the simple consistent estimators discussed earlier. However, they are also more expensive to compute. Even if the maximum likelihood estimates are to be computed, the simple consistent estimators make good starting values for the iterative scheme. Also, as mentioned earlier, one Newton-Raphson iteration is sufficient to obtain asymptotically efficient estimates if consistent starting values are used. Finally, the consideration of the simple consistent estimators is a step in verifying identifiability of the structural parameters which is clearly necessary before maximum likelihood is used.

The general case

Now the general case of a system of G equations in which the first S endogenous variables are subject to truncation is discussed. As a matter of notation, let the observable dependent variables be denoted as

$$\text{(19)} \quad Y = [Y_S, Y_{G-S}]$$

where

$$Y_S = [y_1, y_2, \ldots, y_S], \; Y_{G-S} = [y_{S+1}, \ldots, y_G].$$

The variables in Y_S are those subject to truncation. Define

$$\text{(20)} \quad Y_S^* = [y_1^*, y_2^*, \ldots, y_S^*]$$

and define

$$\text{(21)} \quad Y^* = [Y_S^*, Y_{G-S}].$$

Then,

$$\text{(22)} \quad y_i = \begin{cases} y_i^* \text{ if } y_i^* > 0 \\ 0 \text{ if } y_i^* \leq 0 \end{cases}, \; i = 1, 2, \ldots, S;$$

and one can write the model as

$$\text{(23)} \quad Y^* = Y\Gamma + X\Delta + \epsilon.$$

Here Y and Y^* are (1xG) row vectors of endogenous variables, as described above; X is a (lx$K$) row vector of exogenous variables; ϵ is a (lx$G$) row vector of observations; and Γ and Δ are GxG and KxG matrices of coefficients. The normalization rule is that the diagonal elements of Γ are equal to zero.

As in the two-equation case, certain conditions are necessary for the existence (i.e., internal consistency) of the model. Explicitly, these are as follows.

Conditions for existence of the model. The determinant of $(I - \Gamma)$ must have the same sign as any principal minor of $(I - \Gamma)$ involving at least the last G–S rows and columns. These conditions will be illustrated now for the three-equation case. One has

$$\text{(24)} \quad \Gamma = \begin{bmatrix} 0 & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & 0 & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & 0 \end{bmatrix}$$

$$\text{and } I - \Gamma = \begin{bmatrix} 1 & -\gamma_{12} & -\gamma_{13} \\ -\gamma_{21} & 1 & -\gamma_{23} \\ -\gamma_{31} & -\gamma_{32} & 1 \end{bmatrix}.$$

One notes that if all three endogenous variables are subject to truncation—the case of Amemiya [2]—then the conditions above are that all principal minors of $I - \Gamma$ be positive which agrees with Amemiya's condition [2, p. 1006], as of course it must.

The case in which only the first two endogenous variables are truncated now is considered in some detail. The model is

$$(y_1^*, y_2^*, y_3) = (y_1, y_2, y_3)\Gamma + X(\delta_1, \delta_2, \delta_3) + (\epsilon_1, \epsilon_2, \epsilon_3)$$

with Γ as given in (24) above. Note that there are four subsamples:

$$t_1 = \{t \mid y_{1t} > 0, y_{2t} > 0\}$$
$$t_2 = \{t \mid y_{1t} > 0, y_{2t} = 0\}$$
$$t_3 = \{t \mid y_{1t} = 0, y_{2t} > 0\}$$
$$t_4 = \{t \mid y_{1t} = 0, y_{2t} = 0\}.$$

For $t\epsilon t_1$, the reduced form equations (analogous to those in equation (7) of the second section) for y_1^* and y_2^* are

$$(25)\quad y_1^* = \frac{1}{D}\{X[\delta_1(1-\gamma_{23}\gamma_{32}) + \delta_2(\gamma_{21}+\gamma_{31}\gamma_{23}) + \delta_3(\gamma_{31}+\gamma_{21}\gamma_{32})] + [\epsilon_1(1-\gamma_{23}\gamma_{32}) + \epsilon_2(\gamma_{21}+\gamma_{31}\gamma_{23}) + \epsilon_3(\gamma_{31}+\gamma_{21}\gamma_{32})]\}$$

$$y_2^* = \frac{1}{D}\{X[\delta_1(\gamma_{12}+\gamma_{32}\gamma_{13}) + \delta_2(1-\gamma_{31}\gamma_{13}) + \delta_3(\gamma_{32}+\gamma_{31}\gamma_{12})] + [\epsilon_1(\gamma_{12}+\gamma_{32}\gamma_{13}) + \epsilon_2(1-\gamma_{31}\gamma_{13}) + \epsilon_3(\gamma_{32}+\gamma_{31}\gamma_{12})]\}.$$

where $D = |I - \Gamma|$. For $t\epsilon t_2$, the reduced form equations for y_1^* and y_2^* are

$$(26)\quad y_1^* = \frac{1}{D_2}\{X[\delta_1 + \delta_3\gamma_{31}] + [\epsilon_1 + \epsilon_3\gamma_{31}]\}$$

$$y_2^* = \frac{1}{D_2}\{X[\delta_1(\gamma_{12}+\gamma_{32}\gamma_{13}) + \delta_2(1-\gamma_{31}\gamma_{13}) + \delta_3(\gamma_{32}+\gamma_{31}\gamma_{12})] + [\epsilon_1(\gamma_{12}+\gamma_{32}\gamma_{13}) + \epsilon_2(1-\gamma_{31}\gamma_{13}) + \epsilon_3(\gamma_{32}+\gamma_{31}\gamma_{12})]\},$$

where

$$D_2 = 1 - \gamma_{13}\gamma_{31}.$$

Now, the expression for y_2^* in (26) obtains when $y_2^* \leq 0$ (and $y_1^* > 0$). When it is positive, it must be the case that $y_2^* > 0$. But this is consistent with the expression for y_2^* in (25) if, and only if, the sign of D is the same as the sign of D_2.

Similarly, for $t\epsilon t_3$ the reduced form equations for y_1^* and y_2^* are

$$(27)\quad y_1^* = \frac{1}{D_1}\{X[\delta_1(1-\gamma_{23}\gamma_{32}) + \delta_2(\gamma_{21}+\gamma_{31}\gamma_{23}) + \delta_3(\gamma_{31}+\gamma_{21}\gamma_{32})] + [\epsilon_1(1-\gamma_{23}\gamma_{32}) + \epsilon_2(\gamma_{21}+\gamma_{31}\gamma_{23}) + \epsilon_3(\gamma_{31}+\gamma_{21}\gamma_{32})]\}$$

$$y_2^* = \frac{1}{D_1}\{X[\delta_2 + \delta_3\gamma_{32}] + [\epsilon_2 + \epsilon_3\gamma_{32}]\}$$

where $D_1 = 1 - \gamma_{23}\gamma_{32}$. Comparing the expressions for y_1^* in (25) and (27) leads to the conclusion that D and D_1 must have the same signs.

Finally, the $t\epsilon t_4$, the reduced form equations for y_1^* and y_2^*, are

$$(28)\quad y_1^* = \frac{1}{D_{12}}\{X[\delta_1 + \delta_3\gamma_{31}] + [\epsilon_1 + \epsilon_3\gamma_{31}]\}$$

$$y_2^* = \frac{1}{D_{12}}\{X[\delta_2 + \delta_3\gamma_{32} + [\epsilon_2 + \epsilon_3\gamma_{32}]\}$$

where $D_{12} = 1$. Comparing the expressions for y_1^* in (26) and (28) leads to the conclusion that D_2 and D_{12} must have the same sign; comparing the expressions for y_2 in (27) and (28) leads to the conclusion that D_1 and D_{12} must have the same sign. Since $D_{12} = 1$ is clearly positive, it is concluded that D, D_1 and D_2 must be positive. However, one can draw no conclusion about one of the principal minors of $I - \Gamma$; $1 - \gamma_{21}\gamma_{12}$ can be of either sign.

If one looks at the case in which only the first endogenous variable is subject to truncation, then one can conclude only that D and D_1 have the same sign. One cannot even say whether this sign is positive or negative which is seen easily by the same line of reasoning as in the case in which two variables are truncated.

These examples are revealing because they make clear the pattern for the general case of G equations with S variables subject to truncation. In this case there will be 2^S reduced forms, corresponding to all possible permutations of the events $y_i > 0$ or $y_i = 0$, $i = 1, 2,$

. . ., S. Each of these 2^S reduced forms will have in the denominator of each equation a different principal minor of $(I - \Gamma)$; specifically, it will be the principal minor involving the last G–S rows and columns, plus all rows and columns for which the corresponding y_i is positive in that reduced form. Now, if one compares the expressions for y_i^* in any two reduced forms which agree on the classification (positive versus zero) of all other $y_j\,(j \neq i)$, one finds them to be identical except for the different principal minors in the denominator. By the same line of reasoning used above, one can conclude that these principal minors must have the same sign. Finally, by making all such possible comparisons[2], one arrives at the conditions for existence of the model as given above.

Now the maximum likelihood estimation of the model is turned to. Again, it is easier to understand the general case if one first considers a special case, so the special case of the three-equation model with two variables subject to truncation is considered. Then one has

$$
\begin{aligned}
(29)\quad L = {} & \pi_{t \epsilon t_1} D\, f(y_{1t} - \gamma_{21} y_{2t} - \gamma_{31} y_{3t} \\
& - X_t\delta_1, y_{2t} - \gamma_{12} y_{1t} - \gamma_{32} y_{3t} \\
& - X_t\delta_2, y_{3t} - \gamma_{13} y_{1t} \\
& - \gamma_{23} y_{2t} - X_t\delta_3) \\
& \cdot \pi_{t \epsilon t_2} D_2 \int_{-\infty}^{(-\gamma_{12}y_{1t} - \gamma_{32}y_{3t} - X_t\delta_2)} f(y_{1t} - \gamma_{31}y_{3t} - X_t\delta_1, \epsilon_2, y_{3t} \\
& - \gamma_{13} y_{1t} - X_t\delta_3)\, d\epsilon_2 \\
& \cdot \pi_{t \epsilon t_3} D_1 \int_{-\infty}^{(-\gamma_{21}y_{3t} - \gamma_{31}y_{3t} - X_t\delta_1)} f(\epsilon_1, y_{2t} - \gamma_{32}y_{3t} - X_t\delta_2, y_{3t} \\
& - \gamma_{23} y_{2t} - X_t\delta_3)\, d\epsilon_1 \\
& \cdot \pi_{t \epsilon t_4} D_{12} \int_{-\infty}^{(-\gamma_{31}y_{3t} - X_t\delta_1)} \int_{-\infty}^{(-\delta_{32}y_{3t} - X_t\delta_2)} f(\epsilon_1, \epsilon_2, y_{3t} - X_t\delta_3)\, d\epsilon_2\, d\epsilon_1 .
\end{aligned}
$$

D, D_1, D_2, and D_{12} are as defined above; and f is the density function of $N(0,\Omega)$, the distribution of the (structural) error terms. Note the obvious similarity between this expression and the expression given in (17) for the case of two equations and one variable subject to truncation.

In the general case of G equations with S variables subject to truncation, one has 2^S subsamples to consider; and the likelihood function is of the form

$$
(30)\quad L = \pi_{i=1}^{2^S} \pi_{t \epsilon t_i} g_i (y_{1t}, \ldots, y_{St}, \ldots, y_{Gt}).
$$

The function g_i is the appropriate density or probability function for observations in the i^{th} subsample. If the i^{th} subsample has E_i effective truncations (i.e., E_i y's equal zero, and S–E_i y's are positive), then g_i will be the appropriate E_i – dimensional integral of the density f of the structural error terms. It will be multiplied by a Jacobian which will equal the absolute value of the principal minor of $(I - \Gamma)$ which includes all rows and columns except those corresponding to the E_i effective truncations.[3]

The analysis of identification is fairly obvious from the discussion in the second section. Recall that, in the case of the two equations, with only the first dependent variable truncated, the second structural equation is identified without any exclusion restrictions. The reason is that *linear combinations of the two equations* which give nonzero weight to the first equation will contain y_1^* (the *untruncated* value of y_1) as a right-hand-side variable. Thus, the second equation is identified with respect to the first.[4] Since this is just a two-equation model, identification of the second equation with respect to the first equation implies identification of the second equation.

In the general case in which the first S variables are truncated (and the last G–S are not), *one has the analogous result that* each of the equations is identified with respect to each of the first S equations. (This identification is apparent since a linear combination of equations giving nonzero weight to the i^{th} equation, for any $i \leq S$, will contain y_i^* which is not contained in any of the structural equations except the i^{th}.) The further conditions necessary for

2. Actually, all possible comparisons don't have to be made; 2^S comparisons will be sufficient if they are distinct comparisons.

3. This principal minor is the same one that appears in the denominator of the reduced form expressions for the corresponding subsample of observations.

4. For a discussion of identification of one equation with respect to another, see Fisher [6, pp. 166 ff].

identification (with respect to *all* equations) then follow from the usual theory of identification; see Fisher [6]. Of course, in the case of Amemiya [3] where $S = G$ (all variables truncated), each equation is identified automatically without any exclusions whatever.

Given the identification of the structural parameters, there is no conceptual difficulty in obtaining maximum likelihood estimators or in claiming the usual desirable asymptotic properties for them. Of course, as the size of the system (and particularly the number of truncated variables) increases, the necessary computations become increasingly difficult.

In principle, the simpler estimation procedure given in the second section could be generalized also which would involve finding relationships between the moments of a G-variate normal distribution with S truncations, which is a very difficult undertaking except when G is rather small or when S equals zero or G. This procedure remains an interesting topic for further research.[5]

Finally, the reader is reminded that in our model it is the observed (truncated) Y which appears on the right hand side of the structural equations—see equation (23). As it was done in the two-equation case, the alternative specification having the untruncated Y^* on the right hand side could be considered. That is, one could replace (23) by

$$(31)\quad Y^* = Y^*\Gamma + X\Delta + \epsilon$$

to get the model of the type analyzed by Nelson and Olson [13] and Amemiya [3]. In such a model there would be a simple reduced form,

$$Y^* = X(I - \Gamma)^{-1} + \epsilon(I - \Gamma)^{-1},$$

rather than the 2^S different reduced forms which (23) implies. These reduced form equations could be estimated easily by Tobit and least squares, so that for this model simple consistent estimates are available in the general case. Also, no restrictions on the parameters are necessary for the existence of this alternative model; though the usual order and rank conditions are necessary for its identification. It must be admitted, then, that in many respects this alternative model is more convenient to work with than our model that has been presented here. However, it is clearly the case that the choice between these models should be made not on grounds of convenience but on the basis of which better describes the reality being modeled. The potential user will have to decide this on a case by case basis.

5. Such simple consistent estimators can be found in Lee [12].

Conclusions

This article has considered simultaneous equation models in which some of the dependent variables are subject to a Tobit-type truncation. The analysis is similar to that of Amemiya [2] for the case in which *all* variables are subject to such a truncation, though the authors do derive some new results on identification. They have found conditions which are necessary for the logical consistency of the model and have derived the conditions for the model to be identified. The model can be estimated by maximum likelihood.

The special case of a two-equation system with one truncated variable has been considered in some detail. In this case, a computationally simple alternative estimation procedure is derived as an alternative to maximum likelihood. These estimates are consistent and can be used as the starting point for an efficient two-step, Newton-Raphson procedure.

REFERENCES

[1] T. Amemiya, "Regression Analysis When the Dependent Variable Is Truncated Normal," *Econometrica*, 41:997–1016 (November 1973).

[2] ———, "Multiple Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal," *Econometrica*, 42:999–1012 (November 1974).

[3] ———, "The Estimation of a Simultaneous Equation Tobit Model," Technical Report No. 236, Institute for Mathematical Studies in the Social Sciences, Palo Alto, CA., Stanford University, 1977.

[4] P. J. Dhrymes, *Econometrices: Statistical Foundations and Applications*, New York, Harper & Row Publishers, Inc., 1970.

[5] ———, "A Note on an Efficient Two-Step Estimator," *Journal of Econometrics*, 2:301–04 (September 1974).

[6] F. J. Fisher, *The Identification Problem in Econometrics*, New York, McGraw-Hill Book Company, 1966.

[7] J. A. Hausman and D. A. Wise, "Social Experimentation, Truncated Distributions, and Efficient Estimation," *Econometrica*, 45:919–38 (November 1977).

[8] J. Heckman, "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42:679–94 (July 1974).

[9] ———, "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables, and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5:475–92 (Fall 1976).

[10] ———, "Sample Selection Bias as a Specification Error," unpublished paper, Chicago, University of Chicago, 1977.

[11] N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, New York, John Wiley & Sons, 1972.

[12] L. F. Lee, "Multivariate Regression and Simultaneous Equations Models with Some Dependent Variables Truncated," mimeograph, 1976.

[13] F. Nelson and L. Olson, "Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables," Social Science Working Paper No. 149, Pasadena, CA., California Institute of Technology, 1977.

[14] T. J. Rothenberg and C. T. Leenders, "Efficient Estimation of Simultaneous Equations Systems," *Econometrica*, 32:57–76 (January 1964).

[15] J. Tobin, "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26:24–36 (January 1958).

APPENDIX A

In this Appendix it is shown how one can estimate the parameters given in (16). For $t \epsilon t_2$, the reduced form is (omitting subscripts for simplicity)

$$\text{(A1)} \quad y_1^* = \Psi_1'x + v_1$$
$$y_2 = \Psi_2'x + v_2$$

where

$$\text{(A2)} \quad \Psi_1 = \delta_1 + \gamma_1\delta_2 = (1 - \gamma_1\gamma_2)\pi_1$$
$$\Psi_2 = \delta_2$$
$$v_1 = \epsilon_1 + \gamma_1\epsilon_2 = (1 - \gamma_1\gamma_2)w_1$$
$$v_2 = \epsilon_2$$

and where π_1 and w_1 as defined in (12) and (13). Also let ϕ be the covariance matrix of $v = (v_1, v_2)'$, so that

$$\text{(A3)} \quad \phi_{11} = \Omega_{11} + 2\gamma_1\Omega_{12} + \gamma_1^2\Omega_{22}$$
$$\phi_{12} = \gamma_1\Omega_{22} + \Omega_{12}$$
$$\phi_{22} = \Omega_{22}.$$

The problem is that one does not observe y_1^*, and $y_1^* \leqslant 0$ for all observations here.

One considers the singly truncated bivariate normal distribution with density

$$\frac{1}{P^*} g^*(v_1, v_2) \quad \text{if } v_1 < b_1$$

$$0 \quad \text{if } v_2 \geqslant b_1$$

where g^* is the density of $N(0,\phi)$, b_1 is an arbitrary scalar, and

$$P^* = \int_{-\infty}^{b} \int_{-\infty}^{\infty} g^*(v_1, v_2)\, dv_2\, dv_1.$$

It is easy to show that

$$\text{(A4)} \quad E(v_2) = -\frac{\phi_{12} f^*(-b_1)}{F^*(-b_1)}$$

where

$$f^*(c) = \frac{1}{\sqrt{2\pi\,\phi_{11}}} \exp\left[\frac{-1}{2\phi_{11}} c^2\right]$$

$$F^*(c) = \int_{c}^{\infty} f^*(-v_1)\, d(-v_1) = P^*.$$

The density f^* and *cdf* F^* are similar to f and F in equation (14) but are for the $N(0, \phi_{11})$ distribution rather than for $N(0, \sigma_{11})$.

With $b_1 = -\Psi_1'x$ and $v_2 = y_2 - \Psi_2'x$, one has

$$\text{(A5)} \quad E(y_2) = \Psi_2'x - \frac{\phi_{12} f^*(\Psi_1'x)}{F^*(\Psi_1'x)}.$$

One cannot use this directly as one has not yet identified Ψ_1 or ϕ. But note that

$$f^*(\Psi_1'x)$$
$$= \frac{1}{\sqrt{2\pi\,\phi_{11}}} \exp\left[\frac{-1}{2\phi_{11}}(\Psi_1'x)^2\right]$$
$$= \frac{1}{1-\gamma_1\gamma_2} \frac{1}{\sqrt{2\pi\,\sigma_{11}}} \exp\left[\frac{-1}{2\sigma_{11}}(\pi_1'x)^2\right]$$
$$= \frac{1}{1-\gamma_1\gamma_2} f(\pi_1'x)$$

and similarly $F^*(\Psi_1'x) = F(\pi_1'x)$. Therefore,

$$E(y_2) = \Psi_2'x - \frac{\phi_{12}}{1-\gamma_1\gamma_2} \frac{f(\pi_1'x)}{F(\pi_1'x)}$$

and

$$\text{(A6)} \quad y_2 = \Psi_2'x + \beta \frac{f(\pi_1'x)}{F(\pi_1'x)} + \theta$$

where

$$\beta = -\phi_{12}/(1-\gamma_1\gamma_2) \qquad \text{and } E(\theta) = 0.$$

One does not observe the regressor $f(\pi_1'x)/F(\pi_1'x)$ because it involves the parameters π_1 and σ_{11}. However, one can estimate these parameters consistently from the other reduced form, as previously seen, so as to form a consistent estimate of $f(\pi_1'x)/F(\pi_1'x)$. Ordinary least squares applied to (A6), with this substitution, then yields consistent estimates of Ψ_2 and β.

Appendix B

In this Appendix a solution is attempted for the structural parameters in terms of the estimable reduced form parameters for the special case of a single explanatory variable ($K = 1$).

The basic equations are as given in equations (15) and (16) of the text which for your convenience are recopied here:

$$\text{(B1)} \quad \pi_1 = \frac{1}{1-\gamma_1\gamma_2}(\delta_1 + \gamma_1\delta_2)$$

$$\text{(B2)} \quad \pi_2 = \frac{1}{1-\gamma_1\gamma_2}(\delta_2 + \gamma_2\delta_1)$$

$$\text{(B3)} \quad \sigma_{11} = \frac{1}{(1-\gamma_1\gamma_2)^2}[\Omega_{11} + 2\gamma_1\Omega_{12} + \gamma_1^2\Omega_{22}]$$

$$\text{(B4)} \quad \sigma_{12} = \frac{1}{(1-\gamma_1\gamma_2)^2}[\gamma_2\Omega_{11} + (1+\gamma_1\gamma_2)\Omega_{12} + \gamma_1\Omega_{22}]$$

$$\text{(B5)} \quad \sigma_{22} = \frac{1}{(1-\gamma_1\gamma_2)^2}[\gamma_2^2\Omega_{11} + 2\gamma_2\Omega_{12} + \Omega_{22}]$$

$$\text{(B6)} \quad \Psi_2 = \delta_2$$

$$\text{(B7)} \quad \beta = \frac{1}{1-\gamma_1\gamma_2}(\Omega_{12} - \gamma_1\Omega_{22}).$$

The problem is to solve for δ_1, δ_2, γ_1, γ_2, σ_{11}, σ_{12} and σ_{22}.

Clearly, it is easy to solve for δ_2 from (B6):

$$\text{(B8)} \quad \delta_2 = \Psi_2.$$

Also, (B3), (B4) and (B5) are just a reflection of the equation (in matrix form) $\Sigma = [(I-\Gamma)^{-1}]'\Omega(I-\Gamma)^{-1}$ which implies $\Omega = (I-\Gamma)'\Sigma(I-\Gamma)$, or

$$\text{(B9)} \quad \Omega_{11} = \sigma_{11} - 2\gamma_1\sigma_{12} + \gamma_1^2\sigma_{22}$$

$$\text{(B10)} \quad \Omega_{12} = \gamma_2\sigma_{11} - \gamma_1\sigma_{22} + (1+\gamma_1\gamma_2)\sigma_{12}$$

$$\text{(B11)} \quad \Omega_{22} = \gamma_2^2\sigma_{11} - 2\gamma_2\sigma_{12} + \sigma_{22}.$$

This equation demonstrates the solution for Ω if one can solve for γ_1 and γ_2.

Also, (B1) and (B2) can be solved for δ_1 and δ_2 to get

$$\text{(B12)} \quad \delta_1 = \pi_1 - \gamma_1\pi_2$$
$$\delta_2 = \pi_2 - \gamma_2\pi_1 .$$

Since δ_2 has been solved for in (B8), one can solve now for γ_2:

$$\text{(B13)} \quad \gamma_2 = \frac{1}{\pi_1}(\pi_2 - \delta_2).$$

Also (B12) will give the solution for δ_1 if one can solve for γ_1.

This solution leaves γ_1 as the only parameter still needed. One now substitutes (B10) and (B11) into (B7) to get

$$\begin{aligned}(1-\gamma_1\gamma_2)\beta &= \gamma_2\sigma_{11} + \gamma_1\sigma_{22} \\ &\quad - (1+\gamma_1\gamma_2)\sigma_{12} - \gamma_1\gamma_2^2\sigma_{11} \\ &\quad + 2\gamma_1\gamma_2\sigma_{12} - \gamma_1\sigma_{22} \\ &= \gamma_2(1-\gamma_1\gamma_2)\sigma_{11} - (1-\gamma_1\gamma_2)\sigma_{12},\end{aligned}$$

or

$$\beta = \gamma_2\sigma_{11} - \sigma_{12}$$

which does not depend on γ_1. So one cannot solve for γ_1; therefore, neither can one solve for δ_1, Ω_{11} or Ω_{12}.

A similar perversity occurs if, as in the text, one uses (B6) and (B7) to solve for δ_2 and γ_2 and then attempts to solve (B1) and (B2) for δ_1 and γ_1. Also, it can be mentioned that, in addition to the parameters listed in (B1)–(B7), one also can estimate the variance of v_2 ($= \epsilon_2$), conditional on $t\epsilon_2$, from the residuals of the (estimable) second question of that reduced form. However, this variance turns out to equal

$$\Omega_{22} + (\sigma_{12} - \gamma_2\sigma_{11})/\sigma_{11}$$

which clearly doesn't help in determining the structural parameters of the first equation. Finally, the reader can verify that taking $K > 1$ still does not help in identifying the first structural equation.